Lior Gideon *Editor*

Handbook of Survey Methodology for the Social Sciences

Springer

Handbook of Survey Methodology for the Social Sciences

Lior Gideon
Editor

Handbook of Survey Methodology for the Social Sciences

Springer

Editor
Lior Gideon
John Jay College of Criminal Justice
City University of New York
899 Tenth Ave.
New York, NY 10019
USA

ISBN 978-1-4614-3875-5 ISBN 978-1-4614-3876-2 (eBook)
DOI 10.1007/978-1-4614-3876-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012940458

© Springer Science+Business Media New York 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Acknowledgments

This project was a truly a unique one. The process of recruiting contributors who specialize in methodology and care about the topic of survey methodology was not an easy one. Time and again, the panel of experts experienced attrition and subject mortality. (I still wonder where some "committed" contributors disappeared to without leaving a trace). This resulted in delays and the recruitment of new experts. As can be gleaned from the biographies of the contributors to this project, this handbook is a result of an international collaboration, one that brings together the devotion of people from different countries, and from different academic disciplines. I was fortunate enough to connect with some wonderful colleagues across the Atlantic who found it important to assist and bring this project to maturity. In particular, I would like to thank Ineke Stoop (Netherlands Institute for Social Research), Eric Harrison (Centre for Comparative Social Surveys at City University London), Vera Toepoel (Department of Leisure Studies at Tilburg University, the Netherlands), and Burke Johnson (College of Education at the University of South Alabama). Not only did they contribute to the handbook, but they were available for advice and also referred other potential contributors. The success of this handbook is theirs as well.

In addition to recruiting and following up with each contributor, there was also the challenge of bringing all the chapters together to use the same voice. This was not an easy task by any means, and in fact, in the end, I chose not to fully integrate the varying writing approaches. In many aspects, the handbook is even better in its current format, presented in different voices, as the topics covered reflect the many and complex facets of social inquiry. No other format could so clearly make the point that surveys are used by a very wide variety of scholars, for different purposes in different formats, in different modes of administration, and in different places across the globe.

I am glad to see this important and unique work being published, and I hope it will be used as a guide by the many researchers around the world who aim to use survey methodology in their own work. The goal was to lay out the theories of survey methodology in a simple way, so that those who are interested will not shy away from it simply because it involves complex and at times unfriendly formulas. Reviews of the chapters confirmed this to be the case, remarking that the chapters were

written in a user-friendly style and conveyed the ideas in a simple and elegant manner. For this I am grateful to all those who contributed, reviewed, and advised during the many months, long hours, and rocky roads of making this handbook the way it is.

I would also like to thank Welmoed Sphar, senior editor at Springer, for her support. Welmoed took the time to listen to the idea and fell in love with it and did not let go until the project was on its way. She has high hopes for this handbook, and I hope her hopes will mature. I know that I am very pleased with it. Katie Chabalko, who assumed editorship on this handbook, was also supportive and willing. She provided me with a lot of support, contacted contributors, and helped me meet the deadline. I would also like to thank Morgan Ryan, from Springer, who assisted in the final stages of preparing the manuscript for print. Zora O'Neill, my personal editor for the past year, was instrumental in reviewing my thoughts and writings. I think it is because of her that my ideas can be clear to others, and for that I am grateful to her.

On a different level, I am filled with gratitude to my family, Hila, Jonathan, and Eithan, for their never-ending support and understanding, and for giving me the space when I needed it the most.

Contents

Part I Introduction to Issues of Survey Methodology

1 Introduction 3
Lior Gideon

2 Classification of Surveys 7
Ineke Stoop and Eric Harrison

3 Survey Research Ethics 23
Robert W. Oldendick

4 An Overlooked Approach in Survey Research: Total Survey Error 37
René Bautista

Part II Designing The Survey

5 Common Survey Sampling Techniques 53
Mary Hibberts, R. Burke Johnson and Kenneth Hudson

6 Frames, Framing Effects, and Survey Responses 75
Loretta J. Stalans

7 The Art of Question Phrasing 91
Lior Gideon

8 Interviewing 109
Lior Gideon and Peter Moskos

Part III Response and Non-response Errors

9 Unit Non-Response Due to Refusal 121
Ineke Stoop

10 Non-Response and Measurement Error 149
Jaak Billiet and Hideko Matsuo

11 Why People Agree to Participate in Surveys 179
Gerald Albaum and Scott M. Smith

12 Respondents Cooperation: Demographic Profile of Survey Respondents and Its Implication 195
Patrick Glaser

13 Effects of Incentives in Surveys 209
Vera Toepoel

Part IV Survey Designs, Modes and Applications

14 Designing the Face-to-Face Survey 227
W. Lawrence Neuman

15 Repeated Cross-Sectional Surveys Using FTF 249
Ineke Stoop and Eric Harrison

16 Costs and Errors in Fixed and Mobile Phone Surveys 277
Vasja Vehovar, Ana Slavec and Nejc Berzelak

17 Mail Survey in Social Research 297
Alana Henninger and Hung-En Sung

18 E-Mail Surveys 313
Gustavo Mesch

19 Increasing Response Rate in Web-Based/Internet Surveys 327
Amber N. Manzo and Jennifer M. Burke

20 Building Your Own Online Panel Via E-Mail and Other Digital Media 345
Vera Toepoel

21 Does Paying More Mean Getting a Better Product: Comparison of Modes of Survey Administration 361
Beau Shine and Brandon Dulisse

Part V Sensitive and Difficult Survey Topics

22 Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items 377
Susan McNeeley

23 Researching Difficult Populations: Interviewing Techniques and Methodological Issues in Face-to-Face Interviews in the Study of Organized Crime 397
Jana Arsovska

24 What Survey Modes are Most Effective in Eliciting Self-Reports of Criminal or Delinquent Behavior? 417
Gary Kleck and Kelly Roberts

Part VI Survey Designs and Construction

25 Issues in Survey Design: Using Surveys of Victimization and Fear of Crime as Examples 443
Sue-Ming Yang and Joshua C. Hinkle

26 What Would You Do? Conducting Web-Based Factorial Vignette Surveys 463
Hadar Aviram

Part VII Special Issues in Survey Methodology

27 Comparability of Survey Measurements 477
Daniel L. Oberski

28 Employee Surveys as Catalysts for Change: Turning Data into Action 499
Patrick Hyland and Orly Dotan-Eliaz

Index 517

Contributors

Gerald Albaum is Research Professor at the Robert O. Anderson Schools of Management at the University of New Mexico. He also is a senior Research Fellow at the IC2 Institute, University of Texas at Austin, and Professor Emeritus of Marketing at the University of Oregon. He received his Ph.D. in 1962 from the University of Wisconsin-Madison and his MBA (1958) and BA (1954) from the University of Washington. He is the author or co-author of numerous books and articles in refereed journals and conference proceedings. His writings deal with issues in research methods, international marketing activities, and direct selling.

Jana Arsovska is an Assistant Professor at the Sociology Department at John Jay College of Criminal Justice. She holds a Ph.D. degree in Criminology from the Catholic University of Leuven where she studied the role of cultural codes in the evolution of ethnic Albanian organized crime groups. Dr. Arsovska has published extensively on organized crime, radical Islam, and human trafficking in the Balkan region. Prior to her current post, she worked for the European Forum for Restorative Justice and underwent training at INTERPOL. Dr. Arsovska has acted as an anti-organized crime consultant for several organizations, including the World Bank and is a co-editor of the Standing Group Organized Crime Newsletter. She is also the co-editor of the book *Restoring Justice After Large-Scale Conflict: Kosovo, Congo and the Israeli-Palestinian Case*. Currently, Dr. Arsovska is working on her new book provisionally titled *Decoding Ethnic Albanian Organized Crime*.

Hadar Aviram is Professor of Law at UC Hastings College of Law. Her work lies at the crossroads of law, sociology, and criminology, and examines the criminal process and the correctional apparatus through social science perspectives. She often combines quantitative, qualitative, and experimental methods. She has studied public willingness to turn to the police, plea bargaining, courtroom practices, and the behavior of legal actors. Her current work examines the impact of the financial crisis on the American correctional landscape.

Rene Bautista is a survey methodologist in the statistics and methodology department at NORC at the University of Chicago. Previously, he

worked at the Gallup Research Center at the University of Nebraska. He holds a master degree in survey methodology, and he is currently a Ph.D. candidate at the University of Nebraska Survey Research and Methodology program. His research focuses on nonresponse, measurement error, interviewer effects, mixed modes, data collection methods, and "election day" surveys. His work has been published in scholarly journals and books. He serves as a reviewer for leading journals in survey methodology, and is a frequent speaker in major national and international conferences on survey methodology.

Nejc Berzelak is a Research Assistant at the Faculty of Social Sciences, University of Ljubljana, Slovenia. His research work mainly focuses on analysis of survey response processes, measurement errors in surveys, web survey methodology, and development of software tools for web surveys.

Jaak Billiet is Emeritus Professor in Social Methodology at the University of Leuven (KU Leuven), Belgium. He was from the very beginning of the European Social Survey involved in the central coordination team. His main research interest in methodology deals with validity assessment and the modeling of measurement error in social surveys. His substantial research covers longitudinal and comparative research in the domains of ethnocentrism, political attitudes, and religious orientations.

Jennifer M. Burke is currently a doctoral student in Criminology at the University of Cincinnati where she has taught undergraduate classes in Criminal Justice. She also teaches as an adjunct in the Political Science Department of John Carroll University. She received her J.D. from Boston College Law school and has practiced law since 2001. Ms. Burke's research interests include regulatory crime prevention, survey techniques, and professional ethics.

Orly Dotan-Eliaz has worked in not-for-profit as well as corporate settings doing program evaluation and organizational research and development, including survey work and consulting. Orly earned her doctorate in Industrial-Organizational Psychology from CUNY's Graduate Center and is also a certified Coach. She has taught undergraduate and graduate level courses about leadership, organizational research and consulting, and organizational behavior. Her current research interests include examining the impact of linguistic diversity within organizations as well as effective survey utilization.

Brandon Dulisse is a Ph.D. student in the School of Criminal Justice at the University of Cincinnati. He has attended the University of Cincinnati throughout his academic career, where he has earned his B.S. and M.S. in Criminal Justice. He currently serves as a facilitator for the Distance Learning program (an online Master's program through the University of Cincinnati's School of Criminal Justice). His areas of interest include biosocial criminology and corrections.

Lior Gideon is an Associate Professor at the John Jay College of Criminal Justice in New York, New York, is a devout methodologist,

and has over 15 years of international experience in teaching methodology courses and training future cohorts of researchers in the field of criminology and criminal justice research. He also specializes in corrections-based program evaluation and focuses his research on rehabilitation, reentry, and reintegration issues and in particular by examining offenders' perceptions of their needs. To that extent, Dr. Gideon developed many survey-based measurements to examine level of punitiveness, attitudes supportive of rehabilitation, and recently measures of social support. His research interests also involve international and comparative corrections-related public opinion surveys and their affect on policy. To that extent, Dr. Gideon published several manuscripts on these topics, including two previously published books on offenders' needs in the reintegration process: *Substance Abusing Inmates*: *Experiences of Recovering Drug Addicts on Their Way Back Home* (2010, Springer), and *Rethinking Corrections*: *Rehabilitation, Reentry, and Reintegration* (with Hung-En Sung, 2011, Sage). Apart from the above, Dr. Gideon has published a methodology book titled *Theories of Research Methodology*: *readings in methods*, which is now available in its second addition. His other works were recently published in *The Prison Journal*, the *International Journal of Offender Therapy and Comparative Criminology*, and *the Asian Journal of Criminology*. Dr. Gideon earned his Ph.D. from the Faculty of Law, Institute of Criminology at the Hebrew University in Jerusalem, Israel, and completed a postdoctoral fellowship at the University of Maryland's Bureau of Governmental Research.

Eric Harrison is Senior Research Fellow in the Centre for Comparative Social Surveys at City University London. Since 2006 he has been a member of the Core Scientific Team of the European Social Survey, with responsibilities for the quality of socio-economic variables, the development of attitudinal indicators of societal progress, and the implementation of new methods of coding media reporting during ESS fieldwork periods. He studied Social and Political Sciences at the University of Cambridge, completed a doctorate at Nuffield College, Oxford, and has held teaching and research positions at the Universities of Manchester, Plymouth, Oxford and Essex. His principal interests lie in the analysis of social stratification, social inequality and the methodological and substantive challenges of comparative social research.

Mary Hibberts is a Ph.D. student in Instructional Design and Development at the University of South Alabama. She works in the Center for Evaluation, Measurement, and Statistics and assists in quantitative methods courses in the college of education. She plans on becoming a professor in Instructional Design with an emphasis on research, statistics, and program evaluation.

Joshua C. Hinkle is an Assistant Professor in the Department of Criminal Justice and Criminology at Georgia State University. He received his Ph.D. from the Department of Criminology and Criminal Justice at the University of Maryland in 2009. His research interests

include evidence-based policing, the impacts of police tactics on residents of targeted areas, the role of disorder in communities and theory testing. Recent publications have appeared in the Journal of Experimental Criminology, Criminology and Public Policy and the Journal of Criminal Justice.

Kenneth Hudson is Associate Professor of Sociology and Social Work atthe University of South Alabama. His areas of research are social stratification, labor markets, and demography. He currently teaches graduate and undergraduate courses in applied statistics and research methods.

Patrick Hyland is the Director of Research and Development at Sirota Survey Intelligence. He has over a decade of experience in organizational research and consulting. At Sirota, he is engaged in all aspects of survey research and actioning with various clients, including various Fortune 100 companies. Patrick has a Ph.D. in Social-Organizational Psychology from Teachers College at Columbia University, where he also serves as an Adjunct Professor. His dissertation research focused on dispositional sources of resistance to change and leadership practices that can be used to help change-resistant employees cope with change. He received his Bachelor of Arts in English from the University of Pennsylvania.

R. Burke Johnson is Professor of Professional Studies at the University of South Alabama. His current interests are in mixed research and the philosophy of social science. He is coauthor of "Research Methods, Design, and Analysis" (11th edition, Pearson, 2011); "Dictionary of Statistics and Methodology" (4th edition, Sage, 2011); "Educational Research: Quantitative, Qualitative, and Mixed Approaches" (4th edition, Sage, 2011). He served as one of four editors of "The Sage Glossary of the Social and Behavioral Sciences" (Sage, 2009), and is co-editor of a four volume set of books on "Correlation and Regression" (Sage London, in press).

Gary Kleck is Professor of Criminology and Criminal Justice at Florida State University. He received his doctorate in Sociology from the University of Illinois in 1979, where he received the University of Illinois Foundation Fellowship in Sociology. He has been at Florida State since 1978. His research has focused on the topics of the impact of firearms and gun control on violence, deterrence, crime control, and violence. He is the author of *Point Blank*: *Guns and Violence in America*, which won the 1993 Michael J. Hindelang Award of the American Society of Criminology, awarded to the book of the previous several years which "made the most outstanding contribution to criminology". More recently, he is the author of *Targeting Guns* (1997) and, with Don B. Kates, Jr., *The Great American Gun Debate* (1997) and *Armed* (2001). His articles have been published in the *American Sociological Review, American Journal of Sociology, Social Forces, Social Problems, Criminology, Journal of Criminal Law and Criminology, Law & Society Review, Journal of Research in Crime and Delinquency, Journal of Quantitative Criminology, Crime and Delinquency, UCLA Law Review, the Journal of the American Medical Association*, and many

other journals. Kleck has testified before Congress and state legislatures on gun control issues, and worked as a consultant to the National Research Council, National Academy of Sciences Panel on the Understanding and Prevention of Violence and as a member of the U.S. Sentencing Commission's Drugs-Violence Task Force. He is a referee for over a dozen professional journals, and serves as a grants consultant to the National Science Foundation.

Amber N. Manzo is currently a doctoral student in Criminology at the University of Cincinnati where she also teaches classes in the University's online distance learning program in Criminal Justice. She received her M.S. in Criminal Justice from the University of Cincinnati and her B.S. in Criminal Justice from the University of Nevada, Reno. Ms. Manzo's research interests include biosocial criminology, survey techniques, and federal sentencing.

Hideko Matsuo (Ph.D. Demography 2003, Groningen University, Netherlands) is a researcher at the Centre of Sociological Research at KU Leuven since 2006. As a central coordination team member, she has been involved in the study response-based quality assessment of ESS contact files focusing on non-response issues for the past three rounds.

Susan McNeeley is a doctoral student in Criminal Justice at the University of Cincinnati. Her research interests include courts and sentencing, environmental criminology, and public opinion. Current projects include work on prosecutorial decision-making, public support for punitive crime control measures, and the effect of community characteristics on fear of crime.

Gustavo Mesch is Professor of Sociology at the University of Haifa. His research focus on Internet effects on society, online social networks, and social capital. He was the Chair of the Information and Communication Technologies Section of the American Sociological Association and is a member elected of the Board of Directors of the Israel Internet Society.

Peter Moskos is an Associate Professor in the Department of Law and Police Science at John Jay College of Criminal Justice and the City University of New York's Doctoral Program in Sociology. He also teaches at LaGuardia Community College in Queens. Moskos is a Harvard and Princeton trained sociologist and former police officer. He specializes in police culture, crime prevention, qualitative methods, and ending the war on drugs. His first book, *Cop in the Hood*, won the 2008 American Publishers Award for Professional and Scholarly Excellence, Best Book in Sociology. His second book, *In Defense of Flogging*, is a subversive attack on the prison-industrial complex. Moskos was born in Chicago, went to public schools in Evanston, Illinois, and lives in Astoria, Queens.

W. Lawrence Neuman is Professor of Sociology, Chair of the Department of Sociology, Anthropology, and Criminal Justice, and

Coordinator of Asian Studies at the University of Wisconsin-Whitewater. He has authored several textbooks in social research methodology. He has published in *Sociological Inquiry, Social Problems, Social Science Quarterly, Sociological Quarterly, Critical Asian Studies, Journal of Contemporary Asia*, and other academic journals. In addition to research methodology, his research interests include political sociology, East Asian societies, and social diversity-inequality.

Daniel Oberski obtained his Ph.D. from the department of methods and statistics in Tilburg University, The Netherlands, through the Interuniversity Graduate School of Psychometrics and Sociometrics. Between 2009 and 2011 he was representative of Pompeu Fabra University, Spain, in the Central Coordinating Team of the European Social Survey (ESS), where he worked on evaluation of the ESS questionnaire and design and analysis of survey experiments. He was a founding board member of the European Survey Research Association (ESRA). His main research interests are measurement error in survey questions and latent variable modeling.

Robert W. Oldendick is a Professor of Political Science and the Director of the Institute for Public Service and Policy Research at the University of South Carolina. He received his Ph.D. from the University of Cincinnati, where he worked for 15 years in the University's Institute for Policy Research before moving to South Carolina. He has more than 35 years experience in the field of survey research and public opinion and has served as principal investigator or project manager on more than 250 survey-based projects. He is the author of numerous articles on survey research methodology, including procedures for sample selection in telephone surveys, methods of respondent selection within households, and the effects of question wording and format on responses to survey questions. He is also co-author (with Barbara A. Bardes) of *Public Opinion*: *Measuring the American Mind*, the 4th edition of which was published in 2012.

Kelly Roberts is a doctoral student in the Criminology and Criminal Justice program at Florida State University. She received her M.S. in Criminology and Criminal Justice from the Florida State University in 2009. She is a member of the American Society of Criminology, and has presented research on sentencing, racial threat, punitive attitudes, and neighborhood influences on antisocial behavior.

Beau Shine is a Ph.D. student in the School of Criminal Justice at the University of Cincinnati. He attended Western Michigan University where he earned his B.A. in Communication in 2008. He went on to earn his M.S. in Criminal Justice at the University of Cincinnati in 2010, and continues to work towards pursuing his Ph.D. He has facilitated for the Distance Learning program (an online Master's program through the University of Cincinnati's School of Criminal Justice), and currently serves at the Field Placement Coordinator for the School of Criminal Justice. His areas of interest include institutional corrections and reentry.

Ana Slavec is a Research Assistant at Faculty of Social Sciences, University of Ljubljana, Slovenia. Her current research areas are post-survey adjustments, survey quality, dual frame surveys, non-response, online surveys, mixed modes, cross-national research.

Scott M. Smith is President of Qualtrics Labs, Inc. and Professor Emeritus at Brigham Young University. He received his Ph.D. from the Pennsylvania State University, his MBA from Michigan State University, and BS from Brigham Young University. He is the author or co-author of numerous books and articles in referred journals and conference proceedings. His writings deal with issues in Internet marketing, online survey research methods, and cross-cultural research.

Loretta J. Stalans is a Professor in the Departments of Criminal Justice and Criminology and Psychology, and an Affiliated Professor of the Women's Studies and Gender Studies Program at Loyola University Chicago. She is associate editor of Victims and Offenders. She has published two books and numerous refereed academic articles on public opinion about justice and legal authorities. Her Oxford University book, titled *Penal Populism and Public Opinion*: *Lessons from Five Countries*, was written with colleagues Julian V. Roberts, Mike Hough, and David Indermaur. Her other research interests include legal authorities' decisionmaking, jury nullification, improving risk assessment of violent or sex offenders, women offenders, gender comparisons and sexual and intimate partner violence.

Ineke Stoop is head of the Department of Data Services and IT, The Netherlands Institute for Social Research/SCP. She has been a member of the Core Scientific Team of the European Social Survey since the project began in 2001, and is a Laureate of the 2005 Descartes Prize for Excellence in Scientific Collaborative Research. She obtained her Ph.D. from Utrecht University for a thesis on survey nonresponse, and has lectured on the topic both in the Netherlands, and also more widely on ESS training courses and at the ECPR Summer School. She is a member of the European Statistical Advisory Committee (ESAC), and scientific secretary of the International Association of Survey Statisticians (IASS). Her principal research interests are comparative social surveys and nonresponse. She is also the co-author of a recent book titled: Improving Survey Response: Lessons Learned from the European Social Survey (2010).

Hung-En Sung is professor of criminal justice at John Jay College of Criminal Justice. He specializes in offender rehabilitation, correctional health, and comparative analysis of crime and justice. His current work focuses on the diversion and treatment of chronic offenders with co-occurring disorders and the therapeutic mechanisms of faith-based recovery interventions. Dr. Sung is also examining the impact of morbidity and healthcare needs on criminal recidivism among offenders under institutional or community supervision. In 2010, the National Institute of Justice awarded him the W.E.B Du Bois Fellowship to

research on the safety and health consequences of the legal exclusion of undocumented migrants.

Vera Toepoel is Assistant Professor in Leisure Studies. She has a Masters degree in leisure studies and a Ph.D. in survey methodology. Vera previously worked for Centerdata, a research institute specializing in (online) panel surveys. She built and maintained several survey panels (probability and non-probability-based) and conducted hundreds of surveys. She is currently working on a book on conducting surveys online and building a panel for leisure studies (www.leisurepanel.eu).

Vasja Vehovar is a full professor of Statistics at Faculty of Social Sciences, University of Ljubljana, Slovenia. He teaches courses on Sampling, Survey methodology, and Information Society. Member of the European Social Survey methods group; Developer of the WebSM portal devoted to web survey methodology for which he was awarded the AAPOR Waren J. Mitofski Innovators Award 2009. Principal investigator of the national project Internet in Slovenia (RIS) and also involved in various EU information society research projects. Current research areas: survey methodology, especially web surveys, mixed modes, non-response, sampling, complex designs, post-survey adjustments, cross-national research, substantive Internet research.

Sue-Ming Yang is an Assistant Professor in the Department of Criminology at National Chung Cheng University in Taiwan. She received her Ph.D. from the Department of Criminology and Criminal Justice at the University of Maryland in 2007. Her research interests include place-based criminology, criminological theory testing, experimental research methods, analysis of longitudinal terrorism patterns, and understanding the relationship between disorder and crime over time. Recent publications have appeared in Criminology & Public Policy, the Journal of Quantitative Criminology and the Journal of Experimental Criminology.

Part I

Introduction to Issues of Survey Methodology

Introduction

1

Lior Gideon

1.1 Introduction

Surveys have become a major part of our lives. In an era in which a wealth of information is highly accessible and rapidly changing, many researchers use surveys to inform knowledge, challenge existing assumptions, and shape policies. Surveys are used by so many people for so many different purposes that it gives the impression that conducting a survey is as easy as a walk in the park. Many beginning researchers think surveys are simply a way of collecting information by asking questions—nothing sophisticated or difficult, just "ask and you will know."

Unfortunately, such an attitude pervades the foundations of social research, leading some people in the field to contribute knowledge that may be unreliable at best, and outright damaging at the worst. The dangers become even clearer when researchers design and execute a full survey-based study under the name of a respectable academic institution, while knowing very little about method. In the end, they deliver only low-quality results that, due to the institution's prestige, are nonetheless used to inform public policy.

L. Gideon (✉)
Department of Law, Police Science and CJA
John Jay College of Criminal Justice, 899 Tenth Ave, New York, NY 10019, USA
e-mail: lgideon@jjay.cuny.edu

This is all mainly due to the fact that in the course of their studies, not many social scientists have received adequate training in survey methodology. I have seen this time and again when graduate students have approached me to advise on their doctoral work, and just as often when looking at research papers presented in professional conferences by those who have already completed their dissertations and are now conducting independent research. While their topic of research is interesting, often their data collection tool is badly designed, so their results show low reliability and validity. All of them nonetheless proudly declare that their results are valid and can be generalized to the population, as they have used probability as the sampling technique. In fact, it seems that more emphasis is typically given to sampling techniques than to data collection methods and proper data collection protocols.

It is within this context that the current handbook has been written to provide social scientists with a simple point of reference and to educate on the nuts and bolts of this important method. The aim of this book is to examine the various issues surrounding survey methodology, and there is no better way to jump in than to begin with the concept of total survey error (TSE), the theoretical heart of survey methodology, as well as the chapters that follow. While there are many available books and guides on this topic, many of them are either too difficult for students or appear to be somewhat unfriendly to non-statisticians.

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_1, © Springer Science+Business Media New York 2012

1.2 Total Survey Error

Many who use surveys as their primary data collection method fail to think of the process as a scientific method. "What's the big deal about asking questions?" people may say with a shrug. Instead, the focus of research is usually on sampling and the number of questions to be asked. Much less attention is paid to the lurking sources of bias that are not sample-related. Weisberg (2005) warns that this single-minded focus on sampling error is only the tip of the iceberg in surveys: The total survey error is much greater than that. Unfortunately, the emphasis has been—and for many young researchers, continues to be—on sampling errors simply because they are easy to deal with mathematically and can be relatively estimated and resolved by increasing the sample size. On the other hand, errors not related to sampling—what we will call non-sampling errors—have typically been seen as too difficult to estimate, and it has been assumed that their effect on the results would be minimized if samples were big enough and properly representative. In Chap. 4, Bautista discusses the silent bias in survey research while focusing on the concept of total survey error. But for the purpose of paving the way to the other chapters in this book, we will make a brief introduction of this important theoretical framework here.

TSE, as the combined total of both sampling and non-sampling errors, should be the dominant paradigm for developing, analyzing, and understanding surveys and their results. Among researchers using surveys as the main method for data collection, many have assumed people will respond honestly to questions presented to them. There is also a basic assumption that people are generally willing to share their views, opinions, experiences, and attitudes with researchers and thus, all researchers have to do is ask the questions. Such assumptions, however, have been revealed to be untrue. As a result, survey researchers have recently shown an increased interest in what other factors that cause bias in surveys. Returning to the iceberg metaphor, survey researchers have since been able to identify and focus on the submerged part of the iceberg: the core of error not related to sampling, which was previously hidden from view. This effort, along with an accumulated wealth of survey experience in recent decades, has resulted in a better understanding of the multiple sources of survey bias. Figure 1.1 illustrates the concept of TSE using the well-known Pythagorean Theorem as a metaphor: The sum of the squares of both the sampling error and the non-sampling error is equal to that of the squared total survey error—in short, the TSE becomes much bigger than each of its components. Differently put, when both sampling and non-sampling errors occur in a survey, the TSE is exponentially higher. Of course you cannot actually place actual error values and calculate this theorem for the TSE, but it should give readers a good idea of what the actual problem is.

Sampling errors stem from the sampling method used. So researchers must initially identify their population of interest and then clearly define their unit of analysis and what elements will best serve the aim of their study. Once these issues are addressed, researchers progress to the sampling method—either probability or non-probability. It is understood that by using non-probability sampling, bias will naturally be introduced into the research, and no generalization will be possible. This is not to say that one should never use such sampling techniques, but merely to indicate their salient weakness and the corresponding criticism of them. On the other hand, using probability sampling that relies on the principle of randomness will provide a more representative sample, one that better reflects the target population and thus enables generalizations from the sample to the larger population. However, depending on the type of probability sampling used (e.g., simple random, systematic random, stratified proportional, stratified non-proportional, cluster, etc.), the level of sampling error in the model may increase or decrease. Using such methods, a researcher can estimate the sampling error and warn the potential audience

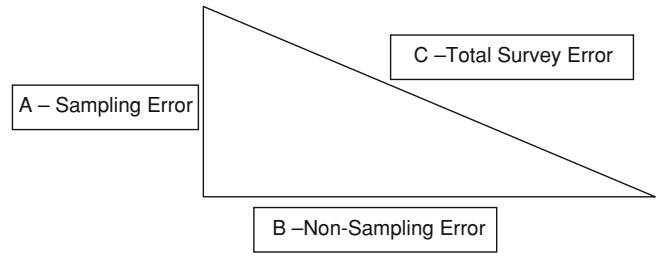


Fig. 1.1 Total survey error

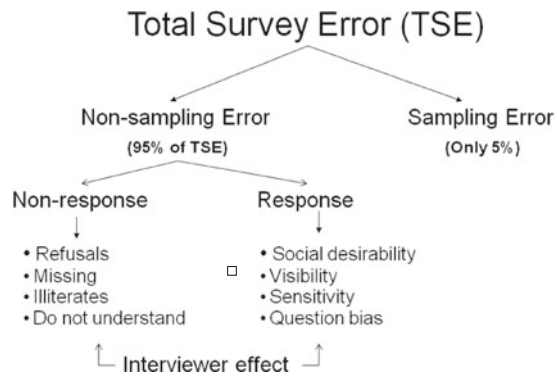


Fig. 1.2 Sources of total survey error

of the source and magnitude of the error. (In Chap. 5, Hibberts, Johnson, and Hudson discuss sampling in more detail, while focusing on the advantages and disadvantages of each sampling method in relation to the researcher's ability to generalize.)

Non-sampling errors, on the other hand, tend to be more complex, and they require researchers' detailed attention, as they may creep into each and every stage of the data collection. As illustrated in Fig. 1.1, its effect on results may be exponentially more damaging to the results of the study. Non-sampling errors can come from a multitude of sources—it is safe to say they can comprise about 95% of the TSE. As illustrated in Fig. 1.2, non-sampling errors are further divided into response and non-response errors, and each of these categories then hosts multiple and additional sources of error. For example, response error can stem from, among other things, social desirability, visibility, the degree of sensitivity of a specific item, the order of the questions, the way in which a specific item is constructed (see Chap. 7 by Gideon)—or even the entire survey topic. (Part VI of this handbook deals with sensitive topics and populations that are difficult to locate populations.) Problems can also stem from the mode used for the questionnaire: In a face-to-face survey, an interviewer can unwittingly increase social desirability, for example. (In Chap. 10, Billiet and Matsuo further discuss the variety of response errors and how they can be controlled.)

On the other hand, non-response errors can stem from simple refusals to answer questions (see Stoop and Harrison, Chap. 9; Albaum and Smith, Chap. 11; and Glaser, Chap. 12). Or they can come as a result of a failure to locate participants who were originally sampled when it comes time to complete the study. Non-response can be for the entire survey, but it can also be for specific questionnaire items. These are important to monitor and examine, as their effect may be detrimental to the results of the study. Non-response to specific items may later affect scaling and can also reduce survey reliability.

It is important to note that interviewer effect may also have a detrimental effect on both response and non-response errors, as can be seen in Fig. 1.2. Accordingly, ample time and emphasis must be devoted for interviewer training and monitoring. In that regard it may be wise to analyze data by interviewer to monitor variations and potential biases prior to the integration and merging of the entire data set.

Non-sampling errors also vary according to the mode of questionnaire administration, as will be discussed further in Part V of this handbook. Each of the above components of the TSE will be addressed at length through the chapters of this handbook.

Yet it is important to understand what course of action we as researchers can take to minimize the TSE, and in particular those errors of response and non-response. Accordingly, the chapters of the handbook will focus on methods for increasing response rate and converting non-response, and the ethical issues that revolve around such practices. Further, methods designed to increase accuracy and quality of response will be discussed. In that regard, Part II of the handbook focuses on the stages of survey design, commencing with common survey sampling techniques, starting with the role of the introduction and questionnaire phrasing on through to interviewing. These chapters should be used by readers as guiding tools in the process of designing a survey.

Reference

Weisberg, H. F. (2005). *The Total survey error approach: A guide to the new science of survey research*. Chicago: University of Chicago Press.

Classification of Surveys

2

Ineke Stoop and Eric Harrison

2.1 Introduction

In 'The Analytical Language of John Wilkins', Borges describes 'a certain Chinese Encyclopaedia, the Celestial Emporium of Benevolent Knowledge', in which it is written that animals are divided into:

- those that belong to the Emperor,
- embalmed ones,
- those that are trained,
- suckling pigs,
- mermaids,
- fabulous ones,
- stray dogs,
- those included in the present classification,
- those that tremble as if they were mad,
- innumerable ones,
- those drawn with a very fine camel hair brush,
- others,
- those that have just broken a flower vase,
- those that from a long way off look like flies.

To modern readers this classification may seem somewhat haphazard, hardly systematic and certainly not exhaustive (although the category 'others' makes up for quite a lot of gaps). Actually, Borges did not find this classification in a Chinese encyclopaedia: he made it up. Making up a classification of surveys at times seems as challenging as making up a classification of animals. A short enquiry into types of surveys yields random samples, telephone surveys, exit polls, multi-actor surveys, business surveys, longitudinal surveys, opinion polls (although some would argue that opinion polls are not surveys), omnibus surveys and so forth. It will be clear that the types of surveys mentioned in this list are neither exhaustive nor mutually exclusive. The 'type' of survey can refer to the survey mode, the target population, the kind of information to be collected and a number of other characteristics. Sometimes these different characteristics interact, but some combinations are rarely found together. Surveys of older persons are rarely web surveys, for instance, and exit polls are never longitudinal surveys.

This chapter presents a brief overview of the different ways in which surveys can be classified. First, however, we need to consider what a survey is. Below is given an abridged version of the section 'What is a survey' from the booklet drafted by Fritz Scheuren from NORC.[1]

I. Stoop (✉)
The Netherlands Institute for Social Research/SCP,
P.O. Box 16164, 2500 BD, The Hague,
The Netherlands
e-mail: i.stoop@scp.nl

E. Harrison
Centre for Comparative Social Surveys,
City University, Northampton Square,
London, EC1V 0HB, UK
e-mail: eric.harrison.1@city.ac.uk

[1] www.whatisasurvey.info/overview.htm

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_2, © Springer Science+Business Media New York 2012

> Today the word 'survey' is used most often to describe a method of gathering information from a sample of individuals. This 'sample' is usually just a fraction of the population being studied…. Not only do surveys have a wide variety of purposes, they also can be conducted in many ways-including over the telephone, by mail, or in person. Nonetheless, all surveys do have certain characteristics in common. Unlike a census, where all members of the population are studied, surveys gather information from only a portion of a population of interest-the size of the sample depending on the purpose of the study. In a bona fide survey, the sample is not selected haphazardly or only from persons who volunteer to participate… Information is collected by means of standardized procedures so that every individual is asked the same questions in more or less the same way. The survey's intent is not to describe the particular individuals who, by chance, are part of the sample but to obtain a composite profile of the population.

In a good survey, the sample that has been studied represents the target population, and the information that has been collected represents the concepts of interest. The standardised procedures with which data are collected are mostly, but not always, questionnaires which are either presented to the sample persons by an interviewer or completed by the sample persons themselves.

In the next section, surveys are classified according to a number of criteria. Underlying this classification is the following poem by Rudyard Kipling:

> I keep six honest serving-men
> (They taught me all I knew);
> Their names are What and Why and When
> And How and Where and Who.

2.2 Classification Criteria

2.2.1 Who: The Target Population

Groves (1989, Chap. 3) starts his theoretical overview of populations (of persons) with the *population of inference*, for instance American citizens in 2011. The *target population* is the finite set of the elements (usually persons) that will be studied in a survey. Generally excluded from the target population are those persons who cannot be contacted or will not be able to participate, such as persons living abroad and those living in institutions (residential care and nursing homes, prisons). The *frame population* is the set of persons for whom some enumeration can be made prior to the selection of the survey sample, i.e. who can be listed in the sampling frame. After the sample has been drawn, ineligible units have to be removed, such as incorrect addresses or persons who are not American citizens. Those who then respond to the survey are the *survey population*, the set of people who, if they have been selected for the survey, could be respondents. *Unit nonresponse* is the failure to collect data from units belonging to the frame population and selected to be in a sample.*The response rate* is the percentage of selected units who participate in the survey.

The population of inference may comprise businesses, households, individuals, days, journeys, etc. In a *business survey*, information is collected on establishments or branches. An informant, or several informants (see Box 2.1), provide(s) information on behalf of a business establishment. A survey among business owners can also be seen as a survey among individuals.

Box 2.1: Examples of business surveys

In two well-known surveys of workplaces, multiple instruments are fielded to different, specifically targeted interest groups.

The 2009 European Companies Survey was conducted using computer assisted telephone interviews (CATI). The companies to be interviewed were selected at random among those with ten or more employees in each country. A management representative and, where possible, an employee representative was interviewed in each company.

The UK's Workplace Employee Relations Survey (WERS) is one of the longest running of its type (since 1980). The most recent wave comprised five separate instruments—some face-to-face and others by self-completion—and the overall design was organised thus:

- An overall sample of 2,500 workplaces, combining 1,700 workplaces that are new

to the study and repeat interviews at 800 workplaces which were first surveyed in 2004.
- At each workplace, an interview with the most senior manager responsible for employment relations and personnel issues was conducted. A self-completion survey on financial performance was distributed to all trading sector workplaces.
- An interview with one trade union employee representative and one non-trade union representative where present (approximately 900 interviews).
- A self-completion survey with a representative group of up to 25 employees, randomly selected from each workplace participating in the study (approximately 23,000 completed surveys).

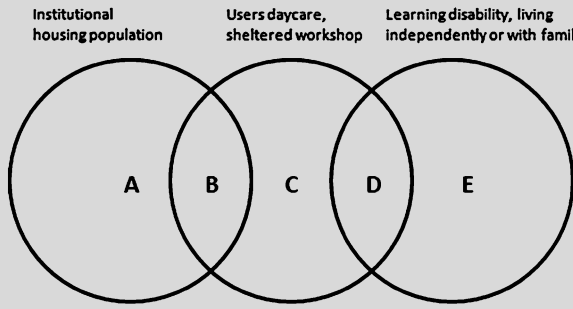
In a *household survey* a responsible adult can function as a household informant. In a *survey among individuals* the respondents usually provide information about themselves, but often also about their households. A respondent can also provide information about other household members, e.g. when providing information on the occupations and education of family members. In some cases the use of *proxies* is allowed, which means that the target respondent has someone else answer the questions for them. A special case of this would be a survey that includes (small) children. In such a case parents can answer questions instead of their children. It is also possible that *all members of the household* have to answer a questionnaire, as for instance in the European Labour Force Survey. In these cases proxies are often allowed. Finally, in *multi-actor* surveys several members of the same family are interviewed, but they will not necessarily be members of the same household. The UK's WERS (see Box 2.1) is also an example of a multi-actor survey. Another example is a Dutch survey among persons with learning disabilities (Stoop et al. 2002, see Box 2.2). A final example of a multi-actor survey is the multi-country survey described in Box 2.7 in Sect. 2.2.6.

Box 2.2: A survey among persons with learning disabilities (see Stoop et al. 2002)

Multiple sampling frames

The frame population consisted entirely of adults aged 18 years and older who had learning disabilities and who were living in an institution or institutionally supported housing arrangement (long-term care home, socio-home, surrogate family unit, supported independent living arrangement) and/or made use of a day-care facility or sheltered workshop. Preceding the fieldwork the frame population was constructed by listing the relevant institutions by completing and joining available address lists. A complication when using the available sampling frames was that the instability of the field: institutions change character, new residential arrangements appear, different residential facilities are hard to distinguish from each other. Additionally, institutions sometimes consist of main locations and satellite sites, which further complicates the sampling procedure.

The selected sampling frames showed overlap and also contained persons who did not belong to the target population (see also figure shown below). Two-thirds of the clients of sheltered workshops, for instance, had physical rather than learning disabilities (C in figure shown below) and were not included in the frame population. Secondly, an unknown number of persons used more than one facility, for instance daycare facilities and residential facilities or services (B in figure shown below). To overcome over coverage, the sampling frame of daycare centres and sheltered workshops was purged of those persons who also used some kind of institutional residential arrangement.



Sampling frame and target population

The sampling procedure was complicated by the fact that different types of institutions were selected and that the final sample would have to be representative according to type of institution and the extent of the learning disability. Firstly, institutions were selected (acknowledging type, size and geographical region) and subsequently clients within institutions, taking into account size and possible overlap between frame populations. The interviewer had to select potential sample persons from a list provided by the local management of the institution, in accordance with a strictly random procedure. In reality, however, this selection was often performed by the local management.

Multiple sources and instruments

Some persons with a learning disability can be interviewed in a survey, whereas others cannot. If possible, the selected sample persons were interviewed personally. They provided information on their daily activities and preferences, autonomy, social networks and leisure activities. Parents or legal representatives were asked about the family background and also, and in greater detail, about the issues in the sample person questionnaire. Support workers or supervisors answered questions on the type and duration of care received, coping abilities and daily activities. Finally, questions on services and facilities provided had to be answered by the local management of institutions providing residential facilities or support, daycare centres and sheltered workshops. The combination of sources was deemed necessary to obtain a complete picture of the quality of life and use of facilities of the sample person. It made the survey particularly complicated, however, because seven different questionnaires had to be used and everybody involved had to cooperate in order to obtain a complete picture.

The population of inference may be the general population of a country (citizens, or residents, which is by no means the same thing). A survey may also aim at representing a special group, such as older persons, members of a minority ethnic group, students, users of a particular product or public service, persons with a learning disability, drug users, inhabitants of a particular neighbourhood, gays and lesbians. In some cases a sampling frame is easy to construct (inhabitants of a particular neighbourhood), and in other cases the survey will have to be preceded by a screening phase to identify the frame population (lesbian and gay people).

Sometimes, sampling is complicated still further when the 'population' under investigation is not a set of individuals but a set of activities or events. In a time use survey, for example, a sample is drawn of households/persons and days (Eurostat 2009), and in passenger surveys the units are journeys (see Box 2.3).

Box 2.3: Passenger surveys

Passenger surveys attempt to establish the perceived quality of a journey. In the UK, this is complicated by the existence of

train operating companies with regionally based but overlapping franchises.

The UK's National (Rail) Passenger Survey (NPS) uses a two-stage cluster sample design for each Train Operating Company (TOC). The first-stage sampling unit is a train station and questionnaires are then distributed to passengers using that station on a particular day during a specified time period. Stations are selected for each TOC with a probability proportionate to size, using the estimated number of passengers as the size measure. A large station may be selected several times. Days of the week and times of day are then assigned to each selected station, using profiles for different types of station. Finally, the sampling points are assigned to weeks at random during the survey period. A completely new sampling plan is generated every two years, utilising data on passenger volumes provided by the Office for Rail Regulation (Passenger Focus 2010).

As mentioned in Sect. 2.1, good survey practises prescribe a survey sample to be selected at random from the frame population. Sampling frames can comprise individuals (a population register, list of students or census records), households, addresses, businesses or institutions. In many cases a two-stage sampling procedure is required, for instance first households, then individuals, or first institutions, then individuals.

There are many ways to draw a probability sample, and according to Kish (1997, see also Häder and Lynn 2007) they all suffice as long as the probability mechanism is clear, which means that every member of the target population has to have a known probability (larger than zero) of being selected for the sample. There are even more ways of selecting a non-probability sample. We will only give some examples here. In many countries, *quota sampling* is quite popular. In this case, a population is first segmented into mutually exclusive sub-groups. Interviewers then have to interview a specified number of people within each subgroup (for further and more in-depth discussion on survey sampling techniques and non-probability samples in surveys, see Hibbert Johnson and Hudson, Chap. 5). How these people are selected is untraceable.

Nowadays online panels, discussed at greater length by Toepoel in Chap. 20, are becoming quite popular (see also Sect. 2.2.4 and Box 2.5). In rare cases these are based on probability samples, as is the Dutch LISS panel (www.lissdata.nl), but the vast majority are not constructed using probability-based recruitment (The American Association for Public Opinion Research 2011). Online access panels offer prospective panel members the opportunity to earn money, make their opinion heard or take part in surveys for fun. In *river sampling* '… respondents are recruited directly to specific surveys using methods similar to the way in which non-probability panels are built. Once a respondent agrees to do a survey, he or she answers a few qualification questions and then is routed to a waiting survey. Sometimes, but not always, these respondents are offered the opportunity to join an online panel' (The American Association for Public Opinion Research 2011).

Rare populations are hard to identify, approach and survey. *Snowball sampling* relies on referrals from initial subjects to generate additional subjects. *Respondent-driven sampling* (RDS) combines 'snowball sampling' with a mathematical model that weights the sample to compensate for the fact that the sample was collected in a non-random way.

2.2.2 What: The Topic

In addition to representing the target population, a survey should represent the concepts of interest. Or, on a more practical note, the second main distinguishing feature of a survey is the topic. Survey topics can be anything, from victimisation to health, from bird-watching to shopping, from political interest to life-long learning and from alcohol and tobacco use to belief in God. There is ample evidence that the topic of a survey is a determinant of the response rate (see Chap. 9 by Stoop).

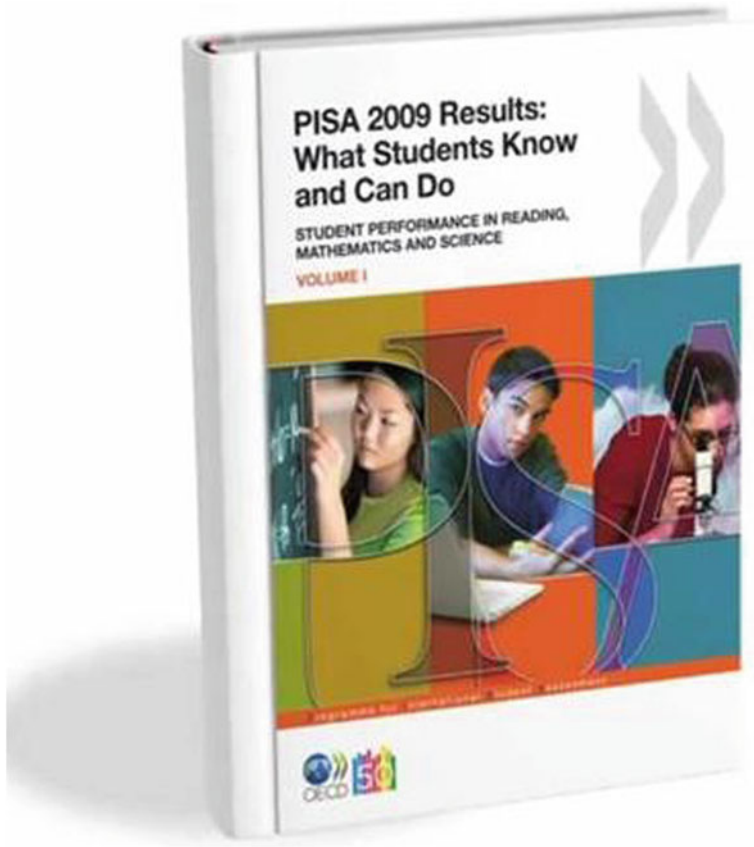
An *omnibus survey* has no specific topic at all: data on a wide variety of subjects is collected during the same interview, usually paid for by

multiple clients. Nowadays, omnibus surveys are increasingly being replaced by online access panels where clients pay for a particular survey while sharing background characteristics.

Often a distinction is made between objective questions and subjective questions. Objective questions are the home turf of official statistics and cover issues like labour situation, education, living conditions, health, etc. Subjective questions collect information on values, attitudes, and the like. In practise, this distinction cannot be sustained. Assessments of health and job preferences have a clear subjective aspect, for example. In addition, official statistics focus increasingly on issues such as satisfaction and even happiness. The UK Office for National Statistics (ONS), for instance, regularly collects data and publishes statistics on 'Measuring Subjective Wellbeing in the UK'. Finally, even objective, hard statistics have a subjective component (e.g. how many rooms are in your house, how much time do you spend on gardening?).

Many different types of organisations collect data on attitudes, values, preferences and opinions, but from a different perspective. For example, there is a big difference between opinion polls and surveys of attitudes and values (and opinions). Although *opinion polls* could be conducted according to the same quality criteria as academic surveys of values and attitudes, in practise they are often commercial, non-probability surveys focusing on one or a few questions, providing results in just a day or so, whereas academic surveys can take a year from data collection to first availability of results.

Appendix 1a presents an overview of comparative attitude surveys organised by different types of sponsors. Other well-known survey topics are behavioural patterns, lifestyles, well-being and social belonging and affiliation (see Appendix 1b). Also common are surveys on literacy and skills (Appendix 1c) and on voting behaviour (1d).



Market researchers study brand and media tracking, consumer satisfaction and advertisement effect. As mentioned above, governments too are interested in consumer satisfaction and use surveys to assess the need for public services. Both—as academics—are interested in factors that determine decision-making.

Some surveys require respondents to keep a diary, for instance time use surveys, travel surveys or expenditure surveys. Other surveys are increasingly supplemented (or even partly replaced) by data from other sources, such as GIS data or data from public registers and administrative records. As part of some surveys, data on bio-markers are collected, such as grip strength, body-mass index and peak flow in SHARE (see Appendix 1) or blood cholesterol and saliva cortisol in the LISS panel (Avendabo et al. 2010). Election polls predict the outcome of elections, as do exit polls, where voters are asked questions about their voting.

From this overview it will be clear that almost any topic can be part of a survey, but also that there is a relationship between the target population and the topic, and the survey agency and sponsor and the topic.

2.2.3 By Whom: Survey Agency and Sponsor

Surveys are commissioned by a wide range of organisations: governments, the media, local communities, labour unions, universities, institutions, NGOs and many other diverse organizations. Survey agencies can be roughly subdivided in four groups: national statistical institutes, universities, market research agencies and not-for-profit organisations. As with the topic, there is ample evidence that the type of sponsor has an impact on the response rate (see Chap. 9 by Stoop). Most studies in this area suggest that people are more likely to participate in an academic or government survey than in a commercial survey. In addition, the topic of a survey is clearly related to the type of sponsor: national statistical institutes do not run exit polls, and market research organisations conduct a lot of consumer research.

In practise, all kinds of combinations of sponsors and data collectors can occur. For instance, television networks can start their own online panels, and market research agencies collect data for national statistical institutes or universities. In the European Social Survey (ESS), an academic cross-national survey (see Chap. 15 on Repeated Cross-Sectional Surveys by Stoop and Harrison), each country selects a survey agency to collect data in that county. ESS data are therefore collected by each of the four types of survey agencies mentioned above (see www.europeansocialsurvey.org: 'Project information'—participating countries). It could however be argued that in the world of surveys, statistics, academia and market research are three different continents (and not-for-profit organisations a kind of island in between). In the world of (official) statistics, sampling is the key element of surveys (see for instance the history of the International Association of Survey Statisticians (http://isi.cbs.nl/iass/allUK.htm). Surveys run by national statistical institutes are almost always based on probability samples, whereas market research organisations increasingly use non-probability samples from online panels (see e.g. Yeager et al. 2011). An instructive overview of the differences between academia and survey research agencies is given by Smith (2009, 2011), summarised in Box 2.4. In the Netherlands and Flanders, a recent initiative is trying to bring together the different approaches to survey research in the Dutch Language Platform for Survey Research (www.npso.net).

Box 2.4: Survey research, academia and research agencies (based on Smith 2009, 2011)

Smith (2009) sees a major divide in the UK between two kinds of knowledge held by survey experts in research agencies and in academia, and feels that this is to the detriment of survey research. He contests that agency practitioners are strong on *knowing how* while academics are strong on *knowing that*. Market researchers have practical skills, but lack theoretical knowledge whereas academics know the theory

but lack practical skills and may therefore have unrealistic expectations about the sorts of data a survey can reasonably be expected to collect. Smith (2009, p. 720) points out three significant problems:

1. Practitioners make needless mistakes because they lack depth in their understanding of how survey errors work.
2. The bulk of surveys in the UK (those not using random probability samples for a start) receive almost no serious academic attention, and suffer as a result.
3. Academic commentary and expectations can be very unrealistic.

He also comes up with a number of possible solutions, although he is rather pessimistic about whether they will be picked up:

- Having academics take secondments in agencies and agency staff take academic secondments.
- Establishing formal links between agencies and academic departments with resource sharing.
- Encouraging academics and agency practitioners to coauthor papers.
- Improving the quality of formal survey training for both academics and practitioners.

In a subsequent paper, Smith (2011) discusses how academics' knowledge might be transferred more effectively, and how it might translate into better survey practise in research agencies. One conclusion he draws from attending an academic seminar on survey non-response and attrition is that he had to try to translate research findings into possible practical recommendations himself, and is not sure whether he drew the right conclusions. The second example he gives is a questionnaire training course taught by Jon Krosnick. This course presented the relevant evidence, but also highlighted some practical implications. Smith (2011) sadly realises that despite the vast question design literature, survey practitioners still write questions in the way they were taught long ago, resulting in questions that are simply bad. So, to improve survey quality, effective ways have to be found to translate academic knowledge into survey questions. Academics should focus on spelling out the practical implications of their findings, and survey agencies should change their practise in line with the results of the academic research.

2.2.4 How: Survey Mode

The best-known distinction between different types of surveys is the survey mode. Section 15.1.3 in Chap. 15 on Repeated Cross-Sectional Surveys describes the main types based on the distinction between interview surveys (face-to-face and telephone) and self-completion surveys (mail and online). Face-to-face surveys are usually rather expensive and thus most often used by academics and statisticians. Interviewers are especially helpful when the survey is long, more than one person in the household has to be interviewed or when additional information has to be collected. Recently, however, interesting experiments have been run in web surveys where respondents themselves collected blood and saliva samples and used online weighting scales (Avendabo et al. 2010).

In many surveys today, multiple modes are used. This might involve a drop-off self-completion questionnaire following a face-to-face survey, or a mixed-mode approach where web, telephone and face-to-face are deployed sequentially to maximise coverage and minimise costs. De Leeuw (2005) gives a useful overview of different modalities of mixing modes.

Commercial organisations make increasing use of online access panels. We use the term 'panel' here not to mean a single sample of people who are monitored over time—as in a longitudinal survey—but in the sense of being a permanent pool of respondents from whom repeated representative (quota) samples can be drawn. The UK organisation YouGov was a pioneer in this field (see Box 2.5).

Box 2.5: Example of an online access panel: YouGov (based on information from http://www.yougov.co.uk/about/about-methodology.asp, accessed on 24 January 2012)

Registration

In order to register with YouGov, each panel member completes a detailed profiling questionnaire and sets up an account name and password. This questionnaire enables YouGov to select a sample that reflects the population. For example, the pool divides into 56% men, 44% women; but in a sample for national political surveys, 52% women and 48% men are selected.

Recruitment and Incentives

The pool is recruited from a wide variety of sources: through targeted campaigns via non-political websites, by specialist recruitment agencies, and people can also join the panel through the open site, although the self-recruited sample is identified as such and is not generally used for published political polls.

Respondents receive a small incentive for completing YouGov surveys, to ensure that responses are not tilted towards those passionately interested in the subject of the particular survey. Incentives typically range from 50p to £1 per survey, through cash payments to an online account which pays out when the panel member reaches £50, as well as occasional cash prizes.

Conducting Surveys

When YouGov conducts a survey, selected panel members are invited by email to complete the survey by clicking on an Internet link. In order to complete the survey they must log in and provide their password. This ensures that the right people complete the survey, and enables their answers to be matched to the demographics they provided when registering with YouGov.

Response rates of at least 40% are normally achieved within 24 h and 60% within 72 h. Little difference has been detected between early and later responses, once the data have been weighted to demographic and attitudinal variables, including past votes and newspaper readership.

Although online access panels are rather popular among commercial agencies (and are inexpensive compared to surveys based on probability sampling), concerns about the non-probability sampling procedures are growing (see Sect. 2.2; Yeager et al. 2011; The American Association for Public Opinion Research 2011). As long as there is no evidence that '… the factors that determine a population member's presence or absence in the sample are all uncorrelated with the variables of interest in a study, or if they can be fully accounted for by making adjustments before or after data collection…' (Yeager et al. 2011, p. 711), the assumption that a sample from an online panel represents the target population cannot be sustained.

Probability-based online samples, on the other hand, such as the Dutch LISS panel (www.lissdata.nl), are a useful addition to scientific research. The LISS panel consists of 5,000 households, comprising 8,000 individuals. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Households that would not otherwise be able to participate are provided with a computer and Internet connection. A special immigrant panel is available in addition to the LISS panel. This immigrant panel comprises around 1,600 households (2,400 individuals), of which 1,100 households (1,700 individuals) are of non-Dutch origin.

2.2.5 When: *Cross-Sections* and Panels

For some surveys, data are collected only once. These are usually called *cross-sections*. In many cases, however, changes over time are an

important part of the research question. In these cases the survey is repeated at regular intervals; this may be every month, every year or every few years. This is usually called a *repeated cross-section*, meaning that a different sample is approached each time. Sometimes this is also called a longitudinal survey, highlighting the fact that the focus is on longitudinal comparison. Technically, however, the term 'longitudinal' is best reserved for a *longitudinal panel*. Here, the same group of respondents is approached at regular time intervals. This makes it possible to measure change at an individual level. Since panel members tend to drop out, and because a panel no longer represents the target population after a few years, a *rotating panel* design can be used (see Box 2.6). New panel members participate in a fixed number of waves. A group of new panel members is recruited for each wave, making it possible to draw conclusions on individual changes and on the population at a given moment.

Box 2.6: Rotating Panel Design, Labour Force Survey (Eurostat 2011, p. 7)

All the participating countries except Belgium and Luxembourg use a rotating panel design for the samples. The number of panels (waves) ranges from two to eight. All panel designs provid for an overlap between successive quarters, except for Germany and Switzerland, which only have a year-to-year overlap. The most common panel design with a quarterly overlap in 2009, adopted by 12 participating countries, is 2-(2)-2, where sampled units are interviewed for two consecutive quarters, are then left out of the sample for the next two waves and are included again two more times. Other common rotation patterns, each used by six countries, are 'in for 5' and 'in for 6' waves, where each panel is interviewed consecutively for five or six quarters before permanently leaving the sample. Three other rotation schemes are used by one or at most two countries.

2.2.6 Where: Regional, National, Cross-National and International Surveys

A survey among the inhabitants of a specific community can be called a regional survey. When all inhabitants of a country belong to the target population, this can be called a national survey. By stratifying regions, it is possible to make sure that study outcomes for both regions and the entire nation can be published.

In theory, international surveys sample from multiple countries and the target population is the combined population of the countries under study. In practise, however, international surveys are rare, because sampling frames seldom cover more than one country and because countries are obvious units of comparison. Consequently, most international surveys are really cross-national surveys: an independent sample is drawn in each participating country, and the results of the national data files are combined afterwards into a harmonised cross-national data file.

Two strategies for harmonisation are used in cross-national studies, namely *input harmonisation* and *output harmonisation* (Körner and Meyer 2005). Input harmonisation means that the instrument is as identical as possible in each participating country: the same fieldwork approach, the same survey mode, the same questions (but translated), and so forth. Output harmonisation allows countries to use their preferred survey mode. Ex-post output harmonisation means that different questions can be used, or that some countries can derive variables from questionnaires and others from registers, as long as the same definitions are used. Ex-ante output harmonisation means that the questionnaire has to be identical in each country, but that the data collection method may differ.

Appendix 1 gives an overview of cross-national or comparative surveys. One of the advantages of these surveys is that in most cases the data are available for external parties, or at least for academic use.

A special example of a multi-national (and also multi-actor) survey is a survey in which migrant families are interviewed in both the sending and receiving countries (see Box 2.7).

Box 2.7: 500 Families: Migration histories of turks in Europe (taken from http://www.norface.org/migration11.html, 18 January 2012)

Social, economic and cultural integration of first generation immigrants and their children has been the focus of extensive research in Europe and elsewhere. However, much remains unknown about the multi-generational transmission of social, cultural, religious and economic resources, and behaviours. Furthermore, while transnational studies on intra- and international migration processes are well established in the US, they are scarce in Europe. Finally, immigrants are mainly compared to other immigrants and/or natives, whereas studies comparing immigrants to those left behind or those who have re-emigrated to the origin country are an exception. This project will treat these research lacunae and will extend existing research on international migration processes and intergenerational mobility by implementing a unique research design based on 500 Turkish families, their immigrant descendants in Europe and those who remained in Turkey. It reconstructs basic migration, family and socio-economic histories of complete lineages generated by a sample of 500 ancestors, born in Turkey between 1930 and 1940 in selected high migration sending areas; and to personally interviewing approximately 6,000 family members over up to four generations in Turkey and Europe, investigating patterns of intergenerational transmission of resources and values and their intersection with family migration trajectories.

2.2.7 Why: Fit for Purpose

Cross-classifying the different classification criteria of surveys do not make any sense because the different classifications are related, as has been mentioned before. In addition, in many cases a trade-off has to be made between accuracy, speed and costs. Exit polls require speed, and cannot be mail surveys; statisticians need to produce exact figures and cover the general population, so non-probability-based online surveys do not suffice.

Even so, no two surveys are ever exactly the same (even when they are intended to be). Like any other product, the form of a survey will be affected by any number of design considerations. Like many products, it will be created with a clear purpose in mind. To this extent, to borrow Le Corbusier's famous dictum, 'form [at least partially] follows function'. This is related to one of the most important survey quality criteria: fit for purpose. A survey is good when it serves the purpose for which it has been designed. Sometimes speed is predominant, sometimes precision of the outcomes and sometimes the comparability of countries. Each purpose has an impact on the final form of the survey.

However, there are other factors which also influence the shape and form of surveys. The aspirations and whims of the survey architects or designers will be visible in its appearance. The survey may seek to resolve well-known problems or weaknesses by using different materials or a new production process. It might build on the success of similar products, or it might be experimental in some way, testing the feasibility of a new process, product or price point. And, in common with most production process involving human beings, things go wrong with surveys.

Like any product or object, then, survey design is a compromise between the sponsor or client and the designers. Along the way, its form may be affected by production problems and cost constraints. Early feedback from market testing may reveal that there are aspects of the product that users find difficult or off-putting, so it has to be re-engineered to make it more acceptable or to improve its quality. Many of these design aspects will be covered in the other chapters of this book.

2.3 Appendix 1: Comparative Surveys

(a) Comparative surveys on attitudes, values, beliefs and opinions

European Social Survey ESS www.europeansocialsurvey.org	Academically driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. Biennial, first Round in 2002, covers more than 30 European countries.
International Social Survey Programme ISSP www.gesis.org/en/issp/overview/	Continuing annual programme of cross-national collaboration on surveys covering topics important for social science research. Brings together pre-existing social science projects and coordinates research goals, thereby adding a cross-national, cross-cultural perspective to the individual national studies.
European Values Study EVS www.europeanvaluesstudy.eu	Large-scale, cross-national, and longitudinal survey research programme focusing on basic human values. Provides insights into the ideas, beliefs, preferences, attitudes, values and opinions of citizens all over Europe. Data are collected every ten years on how Europeans think about life, family, work, religion, politics and society.
World Values Survey WVS www.worldvaluessurvey.org	Explores people's values and beliefs, how they change over time and what social and political impact they have; carried out in almost 100 countries. Data are collected every five years on support for democracy, tolerance of foreigners and ethnic minorities, support for gender equality, the role of religion and changing levels of religiosity, the impact of globalsation, attitudes towards the environment, work, family, politics, national identity, culture, diversity, insecurity and subjective wellbeing.
Eurobarometer http://ec.europa.eu/public_opinion/index_en.htm www.gesis.org/en/eurobarometer-data-service	The Eurobarometer programme monitors public opinion in the European Union. It consists of four survey instruments/series: the Standard Eurobarometer, the Special Eurobarometer, the Flash Eurobarometer, and the Central and Eastern and Candidate Countries Eurobarometer.
Afrobarometer www.afrobarometer.org	Research project that measures the social, political and economic atmosphere in a dozen African countries. Trends in public attitudes are tracked over time. Results are shared with decision makers, policy advocates, civic educators, journalists, researchers, donors and investors as well as average Africans who wish to become more informed and active citizens.
Latinobarómetro www.latinobarometro.org	Annual public opinion survey that involves some 19,000 interviews in 18 Latin American countries, representing more than 400 million inhabitants. Latinobarómetro Corporation researches the development of democracy and economies as well as societies, using indicators of opinion, attitudes, behaviour and values. Its results are used by social and political actors, international organizations, governments and the media.

(continued)

(continued)

AsiaBarometer www.asiabarometer.org	Comparative survey in Asia, covering East, Southeast, South and Central Asia. It focuses on the daily lives of ordinary people (bumi putra) and their relationships to family, neighbourhood, workplace, social and political institutions and the marketplace.
Survey of Health, Ageing and Retirement in Europe SHARE www.share-project.org	Multidisciplinary and cross-national panel database of microdata on health, socioeconomic status and social and family networks of more than 45,000 individuals aged 50 years or over. Started in 2004, now covering 13 countries.

(b) Comparative surveys on living conditions

European Foundation for the Improvement of Living and Working Conditions EUROFOUND www.eurofound.europa.eu/surveys	Eurofound has developed three regularly repeated surveys to contribute to the planning and establishment of better living and working conditions. The surveys offer a unique source of comparative information on the quality of living and working conditions across the EU. • European Working Conditions Survey (EWCS) • European Quality of Life Survey (EQLS) • European Company Survey (ECS)
Household Finance and Consumption Survey HFCS www.ecb.int/home/html/researcher_hfcn.en.html	The HFCS collects household-level data on household finances and consumption. It covers the following household characteristics at micro-level: demographics, real and financial assets, liabilities, consumption and saving, income and employment, future pension entitlements, intergenerational transfers and gifts, and risk attitudes. Data available in 2013.
Eurostat microdata http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction	Access to anonymised microdata available at Eurostat only for scientific purposes. The following microdata are disseminated free of charge: • European Community Household Panel (ECHP) • Labour Force Survey (LFS) • Community Innovation Survey (CIS) • Adult Education Survey(AES) • European Union Survey on Income and Living Conditions (EU-SILC) • Structure of Earnings Survey (SES) • Farm Structure Survey (FSS)
European Community Household Panel ECHP http://epunet.essex.ac.uk/echp.php http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/echp	Harmonised cross-national longitudinal survey focusing on household income and living conditions. It also includes items on health, education, housing, migration, demographics and employment characteristics. ECHP is now finished.
EU Labour Force Survey EU LFS http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/lfs	Conducted in the 27 Member States of the European Union, three candidate countries and three countries of the European Free Trade Association (EFTA). Large household sample survey providing quarterly results on the labour participation of people aged 15 years and over, as well as on persons outside the labour force.
EU Statistics on Income and Living Conditions EU-SILC http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc	Instrument aimed at collecting timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. Provides comparable, cross-sectional and longitudinal multi-dimensional data on income, poverty, social exclusion and living conditions in the European Union. Cross-sectional data and longitudinal data.

(c) Surveys on literacy and skills

Adult Literacy and Lifeskills Survey ALL http://nces.ed.gov/surveys/all/	International comparative study designed to provide participating countries, including the United States, with information about the skills of their adult populations. ALL measures the literacy and numeracy skills of a nationally representative sample from each participating country.
Programme for the International Assessment of Adult Competencies PIAAC www.oecd.org/piaac	International survey of adult skills, collaboration between governments, an international consortium of organisations and the OECD; results to be published in 2013. Measures skills and competencies needed for individuals (15–65 years) to participate in society and for economies to prosper.
Trends in International Mathematics and Science Study TIMSS Progress in International Reading Literacy PIRLS http://timss.bc.edu/	The TIMSS and PIRLS International Study Centre is dedicated to conducting comparative studies in educational achievement. It serves as the international hub for the IEA's mathematics, science and reading assessments. • TIMSS, every four years, 4th and 8th grade • PIRLS, every five years, 4th grade

(d) Information on elections

Comparative Study of Electoral Systems CSES www.cses.org	Collaborative programme of research among election study teams from around the world. Participating countries include a common module of survey questions in their post-election studies. The resulting data are deposited along with voting, demographic, district and macro variables.
Infrastructure for Research on Electoral Democracy PIREDEU www.gesis.org/en/institute/competence-centers/rdc-international-survey-programmes/piredeu	Collaborative project on "Providing an Infrastructure for Research on Electoral Democracy in the European Union", coordinated by the European University Institute (EUI) and its Robert Schuman Centre for Advanced Studies (RSCAS).

References

Avendabo, M., Mackenbach, J. P., & Scherpenzeel, A. (2010). Can biomarkers be collected in an internet survey? A pilot study in the LISS panel. In M. Das, P. Ester & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: Advances in applied methods and research strategies*. Routledge: Taylor and Francis Group.

De Leeuw Edith, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233–255.

Eurostat. (2009). *Harmonised European time use surveys. 2008 Guidelines*. Luxembourg: Publications Office of the European Union.

Eurostat. (2011). *Quality report of the European Union Labour Force Survey 2009*. Luxembourg: Publications Office of the European Union.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Häder, S., & Lynn, P. (2007). How representative can a multi-nation survey be? In J. Roger, R. Caroline, F. Rory & E. Gillian (Eds.) *Measuring attitudes cross-nationally. Lessons from the European Social Survey* (pp. 33–52). London: Sage.

Kish, L. (1997). 'Foreword'. In T. Lê & V. Verma, *DHS Analytical Reports No. 3: An Analysis of Sample Designs and Sampling Errors of the Demographic and Health Surveys*, Calverton, Maryland USA: Macro International Inc.

Körner, T., & Meyer, I. (2005). Harmonising socio-demographic information in household surveys of official statistics: Experiences from the Federal Statistical Office Germany. In J. H. P. Hoffmeyer-Zlotnik & J. A. Harkness (Eds.) *Methodological aspects in cross-national research. ZUMA-Nachrichten spezial band 11*. pp. 149–162. Mannheim

Passenger Focus. (2010). *National passenger survey: Detailed technical survey overview*. Autumn 2010 (Wave 23), London.

Smith, P. (2009). Survey research: Two types of knowledge. Viewpoint. *International Journal of Market Research, 51*(6), 719–721.

Smith, P. (2011). Lessons from Academia. Viewpoint. *International Journal of Market Research, 53*(4), 451–453.

Stoop, I., Spit, J., de Klerk, M., & Woittiez, I. (2002). In pursuit of an elusive population. A survey among persons with a mental handicap. Paper presented at the International Conference on Improving Surveys, Copenhagen, August 2002.

AAPOR. The American Association for Public Opinion Research. (2011). *AAPOR Report on Online Panels.* Prepared for the AAPOR Executive Council by a Task Force operating under the auspices of the AAPOR Standards Committee. AAPOR. Retrieved from www.aapor.org/Content/aapor/Resources/ForMedia/PressReleases/AAPORReleasesReportonOnlineSurveyPanels/default.htm.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., et al. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75*(4), 709–747.

Survey Research Ethics

3

Robert W. Oldendick

The chapters in this book describe many of the important considerations in conducting survey research, from sampling and questionnaire design to determining the most effective mode of data collection and analysis. But underlying all these considerations is the question of ethics in survey research. The most elegant data collection design or most sophisticated statistical analysis of survey data must take into account a variety of ethical concerns, including protecting human subjects from all forms of abuse, guarding the privacy of information, and presenting results that accurately reflect the information provided by respondents.

3.1 Ethical Considerations I: Historical Foundations

The concern over ethical considerations in survey research can be traced to reported abuses of human subjects in medical experimentation that took place during the Second World War. The Nuremberg War Crime Trials and the resultant Nuremberg Code established many of the principles that form the basis for various codes of research, including the voluntary consent of human subjects, the requirement to avoid unnecessary physical and mental suffering, and the ability of subjects to end their participation in a study.

One study that is often cited as an example of the need for standards in the conduct of research is the Tuskegee Study of Untreated Syphilis in the Negro Male, often referred to as the Tuskegee Syphilis Study. In 1932, the Public Health Service, working with the Tuskegee Institute, began a study to record the natural history of syphilis in hopes of justifying treatment programs for blacks (http://www.cdc.gov/tuskegee/index.html). The study initially involved 600 black men and was conducted without any informed consent on the part of participants. This research was originally designed to last for 6 months, but actually went on for 40 years. Although the men agreed to be examined and treated, they were not given all the facts needed to provide informed consent. More importantly, the participants who had syphilis were never given sufficient treatment for their disease. In 1947, when penicillin became the widely accepted treatment for syphilis, the option of such treatment was not offered to these subjects.

Another set of studies that raised questions about the ethical aspects of social science research were those conducted by Milgram (1974). Milgram's interest, sparked by the actions of Nazi war criminals whose defense was that they "were just following orders," was in why so many people seemed to obey

R. W. Oldendick (✉)
Institute for Public Service and Policy Research, University of South Carolina, 1600 Hampton St., Room 411, Columbia, SC 29208, USA
e-mail: roldendi@mailbox.sc.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_3, © Springer Science+Business Media New York 2012

directions from an authority figure. Although Milgram's experiments involved a number of design variations, the central element of these studies was a "shock generator" with voltage switches ranging in 15 V increments from 15 to 450 V. These switches included labels such as "Moderate" (75–120 V) and "Strong" (135–180 V). The switches from 375 to 420 V were labeled "Danger: Severe Shock," while the 435 and 450 V switches were labeled "XXX" (Milgram 1974, p. 28).

Participants in these experiments were told that the study was about "memory and learning." The procedures involved having the subject "teach" a "learner"/"student" (who was described as another subject, but who in reality was working with the experimenter) word pairs on which the learner would then be tested. When the learner made a mistake, the subject would administer a shock and with each mistake would increase the shock level by 15 V. Although the results of these experiments were both controversial and somewhat surprising in terms of the percentage of subjects who administered the maximum voltage, the critical consideration from an ethical perspective involves the deception of subjects.

Subjects in these experiments were deceived in that they were led to believe that the "learner" was a subject like themselves and that it was only by chance that they were the one who was the teacher rather than the learner. They were also deceived into believing that they were administering a shock, when no such shock was actually delivered. In some cases, the lack of a response by the learner after the administration of a "shock" may have led the subject to believe that they had seriously injured or even killed the learner (Milgram 1974, p. 88).

As a response, the Department of Health, Education, and Welfare issued comprehensive regulations for the protection of human subjects—45 Code of Federal Regulations 46, subpart A. This set of regulations, commonly referred to as 45 CFR 46, forms the basis for protection of human participants in research, and its role in safeguarding the rights of human subjects will be noted throughout this chapter.

3.1.1 The Belmont Report

Another important event in reaction to these perceived abuses of human subjects was the 1974 formation of a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Commission's 1979 *Belmont Report* established the guidelines to be followed in conducting research with human subjects.

As noted in the *Belmont Report*, basic ethical principles refer to those general judgments that serve as a basic justification for the many particular ethical prescriptions and evaluations of human actions. Three basic cultural ethical principles particularly relevant to research involving human subjects are *respect for persons*, *beneficence*, and *justice*.

The *Belmont Report* explains that respect for persons means that individuals should be treated as autonomous agents who are capable of deliberation about personal goals and acting under the direction of such deliberation. Moreover, since not every individual—for example, children or the mentally incompetent—is capable of self-determination, the principle of respect for persons dictates that they be afforded special protection from harm when research involving such groups is considered.

The application of the principle of respect for persons leads to the requirement that researchers provide potential subjects with adequate information about what is involved in participation, including potential risks and benefits and the requirement that they obtain voluntary informed consent from participants. Information provided to participants should include the research procedures, their purposes, risks and anticipated benefits, alternative procedures, and a statement offering the subject the opportunity to ask questions and to withdraw at any time from the research. Agreement to participate should be given voluntarily, and such consent obtained without coercion, such as from a person in a position of authority urging cooperation or offering an excessive incentive to participate.

In sum, the subjects should understand clearly the range of risk and the voluntary nature of participation.

The *Belmont Report* describes beneficence as not only respecting people's decisions and protecting them from harm, but also making efforts to secure their well-being. In conducting research, investigators should not only do no harm but also maximize possible benefits while minimizing possible harms. In applying this principle, researchers should consider other ways of obtaining the benefits sought in the research. In addition, beneficence requires that the benefits of the proposed research outweigh the risks to subjects. Participation may involve risks of psychological harm, physical harm, legal harm, social harm, and economic harm, while benefits may accrue to individual subjects, their families, special population subgroups, or society at large. Although risk to subjects cannot be entirely eliminated, it should be minimized, and those designing and reviewing the conduct of research must carefully consider the balance of risks to benefits in studies involving human subjects.

The principle of justice requires researchers to show fairness in selecting participants. In designing a study, a researcher must consider what groups will receive the benefits of the research and which will bear its burdens. The *Belmont Report* identifies two levels of justice, the social and the individual. As noted in this report, individual justice requires that researchers "should not offer potentially beneficial research to some patients who are in their favor or select only 'undesirable' persons for risky research." Social justice requires that researchers not draw a distinction "between classes of subjects that ought, and ought not, participate in any kind of research, based on the ability of members of that class to bear burdens and on the appropriateness of placing further burdens on already burdened persons." In the context of research, justice does not imply that all groups are considered equally for participation in research. Classes of subjects, such as adults, may be preferred as research subjects over classes that are already burdened, such as children or institutionalized populations, and researchers should avoid employing subjects from potentially vulnerable populations simply because they may be more readily available in settings where research is conducted.

As noted previously, the *Belmont Report* led to the development of the federal policy for the protection of human subjects, 45 CFR 46. Fifteen federal departments and agencies have adopted the same language—known as the Common Rule—for addressing this issue (http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html). These regulations consist of four subparts: subpart A, which establishes the basic rules for conducting research with human subjects; subpart B, which provides additional protections for pregnant woman, fetuses, and neonates; subpart C, which provides additional protections for prisoners; and subpart D, which provides additional protections for children.

The Common Rule also requires that research conducted at any institution that receives federal support must be reviewed and approved by an Institutional Review Board (IRB), which is a committee responsible for reviewing the ethical aspects of proposed research. Although there are variations in how each institution's IRB operates, there are three general types of review conducted by these committees. (1) Full board review, which applies to research presenting possible risk to subjects, for example drug or medical device trials or surveys involving children or impaired populations. (2) Expedited review, which applies to studies that present minimal risk to subjects, defined as "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those encountered in daily life or in the performance of routine physical or psychological examinations or tests" (http://ohsr.od.nih.gov/guidelines/45cfr46.html). For example, expedited review would apply to studies that involved the collection of blood samples or moderate exercise by healthy volunteers. (3) Exempt review, in which the involvement of human subjects does not involve children or "at risk" populations and in which participation involves, for example, the use of survey

procedures in which data are recorded in a manner in which subjects cannot be identified and for which any disclosure of an individual's responses would not place them at risk of criminal or civil liability or be damaging to their financial standing, employability or reputation. Many of the studies using survey research that are currently conducted in the United States meet the criteria for exempt review.

3.2 Ethical Considerations II: Data Collection and Human Subjects

45 CFR 46 establishes the formal rules under which investigators should conduct research involving human subjects. A number of practical implications for the ethical conduct of survey research follow from these general principles, including:

1. *Respondents Should Provide Willing and Informed Consent*
   - Survey participants should understand that their participation is voluntary and that they are free not to answer any question they do not want to.
   - Participants should also be aware that they can withdraw from the study at any time.
   - Participants should be informed that declining to participate or withdrawing from the study will not have any negative consequences. This consideration is particularly important in situations where the respondent may have some type of relationship with the survey sponsor, such as in an evaluation of a service delivery program, in which a potential respondent may be concerned that failure to cooperate might adversely affect the services they receive.
   - The request to participate should include a realistic description of potential benefits. Many survey projects provide benefits for society at large and provide minimal, if any, direct benefits to participants; potential respondents should be informed of this.
   - Any costs to participants should be minimized. The survey should be designed so that participation does not result in any additional expenses for subjects. One cost involved in any survey is the time required to complete the interview, and ethical design considerations dictate that data collection procedures strive to minimize the time required of subjects. Any potential costs to participants, including the time required to complete the interview, should be made clear when making the request for participation.
   - In conducting research with populations in which comprehension may be limited, such as children or persons with a mental disability, potential participants should be given an opportunity to choose to participate, and assent obtained. Permission to participate should also be obtained from a third party who understands the potential subject's situation and is able to act in their best interest.
2. *Do No Harm to Participants*
   - Survey participants should understand that they are free not to answer any question they do not want to. All participation is voluntary, and if respondents choose not to answer a question—for any reason—it is their right as research participants.
   - Survey participation should involve minimum disruption to respondents. Interviews should be conducted at a time—and for face-to-face interviews, in a place—that is convenient for respondents. Participation also should not require respondents to retrieve extensive records or background materials. Survey designers should also be aware of the time that subjects are giving in providing information, and not make the questionnaire longer than required to test their hypotheses and address the purposes of the research.
   - Any potential risk to participants should be minimized. Physical harm is rarely an issue in survey participation, although in some health surveys blood samples may be taken that could result in physical discomfort. Emotional harm may be caused by asking sensitive questions or about situations, such as physical abuse, that cause respondents to become

uncomfortable. The researcher's obligation is to make respondents not feel embarrassed, ridiculed, belittled, or generally subjected to mental distress (Sudman 1998).

- In many ways, the principle of "doing no harm to participants," may be summarized as minimizing overall respondent burden, that is the length of the instrument and the difficulty of the task (Singer and Presser 2008). Researchers should strive to limit what they request of participants, both in terms of the amount of information requested and the difficulty involved in accessing and providing such information; they should conduct data collection at a time and place that is most convenient for respondents; and should avoid situations that may pose any physical harm or mental distress to participants.

3. *Minimize Deception*
   - Individuals who are asked to participate in a survey have the right to know about the study and not be deceived. Participants should be provided with (a) information on the purpose of the research; (b) the sponsor of the research; and (c) the approximate length of the interview.
   - While the integrity of the research sometimes requires deception of the respondent, such deception should be kept to a minimum. If it is absolutely necessary to use deception to test the hypothesis under consideration, a rationale for such deception must be included in the materials reviewed by the IRB, which will carefully scrutinize any such requests.
   - In those instances in which survey respondents are deceived as part of the research, they must be thoroughly debriefed at the end of data collection, with an explanation provided as to how they were deceived and why this was necessary for the purposes of the research.
4. *Protect Respondent Confidentiality*

   Individuals who participate in a survey have the expectation that they are doing so in confidence and that the information in the reports and data files from the survey will be produced in a way that will not allow any individual's responses to be identified. There is a distinction between anonymity and confidentiality. Anonymous responses are those obtained from survey participants when there is no way to link answers to the individual respondent. In many survey situations, the investigator has data that could potentially identify the respondent such as a telephone number and information on background characteristics such as age, race, and sex, or even the person's name in studies based on list samples. Ethical survey practice requires that such information be treated in a way that assures that individuals cannot be linked to their responses. To protect confidentiality, survey researchers should:
   - Require that all personnel working on a project, from the principal investigator through all interviewers, coders, and data managers, sign a pledge of confidentiality that explains the purpose and importance of keeping all information confidential, and makes clear that any breach of confidentiality is grounds for termination of employment.
   - In studies in which electronic files that are used for sampling or contacting respondents contain identifying information, they should be password protected. If such files are stored on a network, the location on which they are stored should also be password protected so that only authorized personnel can access them. Any electronic sharing of such files should also be protected to prevent any unintentional disclosure of information related to study participation.
   - Any information, such as name, address, or telephone number that could potentially link survey responses to an individual should be removed from the data files used in analysis. In some studies it is necessary to link the survey results to other information. For example, in a program evaluation in which survey information on program satisfaction is collected from respondents it may be necessary to link these data with administrative records, such as length of time in the

program, types of services received and how they were delivered, and the reason why such services are needed. In such cases, a unique identifier should be created for each individual that allows the data from these multiple sources to be linked. Information on how data from multiple sources are linked should be limited to a few people (for example, the principal investigator and the data processing manager) in order to minimize the potential for violating confidentiality.

For some studies in which researchers consider that violation of confidentiality may produce a high risk for participants, the investigator may seek a *Certificate of Confidentiality* from the Department of Health and Human Services or the Department of Justice. Such a certificate helps researchers guard the privacy of respondents by protecting against compulsory legal demands, such as court orders and subpoenas, for identifying information or identifying characteristics of a research participant. Even with a Certificate of Confidentiality a researcher may disclose information such as child abuse, reportable communicable diseases, or possible threat to the participant or to others, provided that the potential for such disclosure is provided to respondents in the process of obtaining informed consent.

5. *The Issue of Refusal Conversions*

One problem of continuing concern to survey researchers is that of declining response rates in surveys of all types (Curtin et al. 2005; Singer 2006). The effect of non-response on the representativeness of survey results has been extensively investigated and various approaches for minimizing the potential bias have been proposed (Groves and Couper 1998; Keeter et al. 2006; Olson 2006). One technique that has been fairly widely adopted—at least among academic survey researchers—is to attempt refusal conversions; that is, to contact households that initially refuse to participate and attempt to persuade them to cooperate (Tarnai and Moore 2008, p. 381). Refusals can be classified as "hard" ["If you contact me again, I'm calling my attorney;" "Don't call this (expletive) number, you (expletive)] or "soft" ("I'm not really interested in your study;" "We don't have the time for any survey"; or a refusal by a person who is potentially not the respondent). In the case of soft refusals, it is believed that contacting the household at a different time (e.g., when they may not be quite so busy), with a different interviewer (generally a more experienced person trained in refusal conversion techniques), and with a different approach (e.g., emphasizing the importance of the study) will lead to a successfully completed interview. Such conversion attempts are successful in obtaining a completed interview between 10% and 40% of the time (Tuchfarber et al. 1979; Miller and Wedeking 2003; Lavrakas 2010).

The ethical question is one of whether such conversion attempts should be made. Most informed consent procedures include a statement to the effect of, "Your cooperation is voluntary, and you may stop at any time …," and as a result some researchers have raised ethical concerns about refusal conversion attempts. While conversion attempts with hard refusals should not be made (and, in any case, are unlikely to be productive), the situation with soft refusals is more ambiguous. In many cases, such refusals are not reflective of an unwillingness to participate in the study, but more a reaction to the conditions at the time of the contact, for example, immediately after a disagreement with another member of the household, when respondents are eating dinner or engaged in some other family activity, or when the individual is particularly busy. As noted previously, in many cases where respondent selection within households is used, the person initially contacted is not the person who would be the selected respondent. In these situations, there is general agreement that the improvement in the quality of the data resulting from refusal conversion attempts outweighs the potential harm to participants,

and refusal conversion attempts are frequently included as part of the survey protocol, particular those done by academic survey researchers.[1] For further discussion on refusal conversion see Stoop and Harrison, Chap. 9.

6. *The Use of Incentives*
   Another reaction to the issue of declining survey response rates has been the increased use of incentives to foster cooperation (Singer and Bossarte 2006). The ethical question is whether, and under what conditions, the use of monetary incentives to induce participation might be coercive. As noted by Berry et al. (2008), while incentives can be seen as a useful approach that is consistent with the goal of encouraging social cooperation, which is valuable (so perhaps payment should be unrestricted), it may also undermine the requirement that research participation be voluntary. The prospect of economic reward may lead people to take on risks that are not consistent with their true values and preferences. Furthermore, providing incentives may undermine the principle of fair treatment if it leads to disadvantaged persons participating in research at a higher rate than wealthier subjects. They also note a concern that incentive payments may have undue influence on IRBs, in that to the extent they consider the payment as an intrinsic benefit of the research, they may allow higher levels of risk than they should. While there does not appear to be a generally accepted approach to the payment of research incentives to participants, any compensation should not be of a nature that it interferes with the ability of potential subjects to give informed consent, without the possibility of coercion or undue influence, and any study that involves the payment of direct compensation to minors should be reviewed with great scrutiny. Toepoel in Chap. 13 discusses the effects of incentives in survey participation, completion, and reliability of results with further detail.

In sum, the principles of respect for persons, beneficence, and justice established in the *Belmont Report* provide the framework for survey researchers to follow in their interactions with human subjects. Adherence to these principles is essential to the ethical conduct of survey research.

3.3 Ethical Considerations III: Survey Design

To this point the discussion of ethical considerations in survey research has focused on aspects of the survey process involving interaction with human subjects. We next focus on ethical considerations in the design of research.

Many of the factors considered in the ethical design of research could be thought of as "best practices," in that the ethics of conducting a survey demand that it be well designed, and much of the discussion in this section relies on the American Association for Public Opinion Research's (AAPOR) "Best Practises for Survey and Public Opinion Research" (http://www.aapor.org/Content/aapor/Resources/ForResearchers/BestPractises/default.htm).

1. *Mode of Data Collection.* One of the initial considerations in any study using a survey is choosing the mode of data collection. The basic modes of data collection—face-to-face; telephone; mail; other self administered; and the Internet—each has advantages and disadvantages related to factors such as cost, population coverage, and amount and types of data that can be collected. Many surveys use a mixed-mode, employing different approaches in attempting to collect data from respondents (Dillman et al. 2009). Since all surveys have budgetary and other resource constraints, choosing a mode of data collection generally requires some consideration of minimizing survey errors versus minimizing survey costs (Groves 1989).
2. *Sample Selection.* Once the mode of data collection has been established, the next step

[1] For more information on how one survey organization handles refusal conversion attempts, see the University of Indiana's Survey Research Center website at http://www.indiana.edu/~csr/phone_surveys.html.

in the survey process involves sample selection (see Hibberts et al. Chap. 5). Since surveys are used to collect information from some small fraction of a population in order to make inferences about this population, a key element in sample selection is *representativeness*. Selecting a representative sample involves the use of some form of *random or probability sampling*, such as simple random sampling, stratified sampling, cluster sampling, or random-digit dialing, and the type of sample used is largely dependent on the mode of data collection, the information available on potential respondents, and the resources available for conducting the study. The results of a survey in which the sample has been selected using probability methods can be reliably projected to the larger population with a known level of precision. It is critical that the sampling frame (the list from which the sample is selected) is complete; to the extent that it is not, those individuals who are not included have no possibility of being included in the sample, and the potential for bias is increased. Any "sample" that depends on volunteers or on individuals who select themselves into the study cannot ethically be reported as representing some larger population. For further discussion on the limitations of each sampling technique see Chap. 5 by Hibberts, Johnson and Hudson.

3. *Questionnaire Design*. In designing a survey questionnaire there are a myriad of factors to be considered, including whether a question should be asked as an open- or closed- item, whether it should include a filter question or not, the type of response format, the number and order of alternatives, the tone of wording, and the context in which the question is asked (Schuman and Presser 1981). Much research has shown that even slight variations in question wording or the context in which a question is asked can affect the results of a survey (Bishop et al. 1978, 1984). Both Chap. 6, by Stalans, and Chap. 7, by Gideon provide a thorough account of the importance of introduction, framing, and questionnaire designs on response rate, response accuracy, and response quality. Survey researchers are aware of the impact that question wording can have respondents and can use this information to influence results (Bishop 2004). The ethical challenge is to design a questionnaire that is unbiased and does not present questions in a way that will provide support for the hypothesis the researcher is testing or produce results that are desired by a research client. More than 60 years ago, Payne (1951) developed a "concise" checklist of "do and don'ts" for constructing survey questions, and the theme of these suggestions is that the question should reflect the substantive issue being addressed. Techniques such as presenting both sides of issue, providing balanced alternatives, considering the context in which the question is presented, and not using response choices that may increase response set bias are some of the elements that a survey researcher should consider in the development of a questionnaire. Cognitive methods, such as "think-aloud" interviews, can be used to evaluate the questions and a thorough pretest of the questionnaire and field procedures should be done in order to identify questions that are being misunderstood or are difficult for respondents, and to detect any potentially biasing effects in the questionnaire.
4. *Data Collection*. The well-designed survey will not produce quality results if the design elements are not executed properly. For face-to-face and telephone interviews, the first step involves interviewer training. Interviewers should be instructed in how to select respondents for the study and to foster their cooperation. They should also be given an item-by-item review of the questionnaire to learn how the questions should be read, what probes, if any, should be used and what definitions or explanation of terms can be provided to respondents. This review should be followed by a mock interview, practice interviews, and a debriefing session, with the goal of minimizing interviewer-to-interviewer variation in the administration of the survey.

The declining response rate for all types of surveys was mentioned previously, and a key consideration in the data collection process is

achieving a high response rate. From an ethical perspective, if you choose a sample from a complete sampling frame, select the sample using strict probability methods, but then collect information from a relatively small percentage of the respondents selected, can you substantively (and ethically) maintain that the results are representative of the population? This concern over response rate has led survey organizations to put more emphasis in interviewer training on gaining respondent cooperation and to implement strategies, such as pre-notification letters, multiple contact attempts, multi-mode approaches, the use of incentives, and customized survey introductions that should produce a higher response rate. But such techniques must be employed in the context of the protection of human subjects, who voluntarily participate in the research and who have the right to refuse to be included in a study and to terminate their participation at any time.

An additional consideration in data collection is the possibility of faked or fraudulent interviews, in which an interviewer fabricates all or part of an interview. The rate of falsification is generally considered to be quite low and can be minimized by observational methods or monitoring of interviewers, recontacting a sample of respondents for verification and through data analytic methods (Groves et al. 2004). Even more troubling from an ethical perspective is the fabrication or manipulation of survey results, as was alleged in a case involving surveys conducted for the *Daily Kos* (http://www.pollster.com/blogs/daily_kos_research_2000_atrou.php?nr=1; http://www.huffingtonpost.com/2011/05/27/daily-kos-research-2000-lawsuit_n_867775.html). In addition to the obvious violation of ethical principles that fraud entails, the revelation of such fabricated research may raise questions among the public concerning the integrity of survey research and lead to increased public mistrust of survey results.

Over the past decade, the number of surveys conducted using the Internet has increased substantially and this mode of data collection presents several ethical challenges beyond those found with other modes (Nosek et al. 2002). Since Internet surveys typically do not involve subject interaction with a researcher, how is informed consent obtained? Should special procedures be developed or—as is the case with many mail surveys—should the information needed for informed consent be provided in the invitation to participate, with consent consisting of the individual's completion of the survey? Relatedly, how does a researcher ensure that the person selected to participate is the one who actually provides the responses? Requiring identifying information for participation involves a tradeoff with anonymity and the confidentiality of responses and does not completely eliminate the problem. More specifically, how does an Internet-based survey assure that minors are not participating, since this group is provided with additional human subject protections that cannot readily be verified on the Internet? In addition, Internet-based surveys should provide selected respondents with a username and/or password for accessing the survey, and allow the survey to be completed only once by each user, to prevent an individual from participating multiple times. In Internet surveys that are "open subscription" or which use uncontrolled instrument distribution in recruiting respondents there is a possibility that an individual or group may participate more than once, potentially biasing the results.

Since information that is provided in an Internet survey must be transmitted between the devices on which the participants are entering their responses and the server on which the information is being stored, they should be encrypted so that they would be meaningless to anyone who would intercept them during transmission. Data should be transmitted to and stored on a secure server, with access password protected and limited to a small number of research personnel.

3.4 Ethical Considerations IV: The Reporting of Results

The effort involved in collecting data using ethical principles involving human subjects and ethical design principles to assure the validity of

the data can be undermined by reporting the results in a manner that does not match these standards. In analyzing and reporting the results of surveys there are a number of ethical principles that should be followed.

An excellent summary of these principles is provided in AAPOR's "Best Practises," which states that, "…excellence in survey practice requires that survey methods be fully disclosed and reported in sufficient detail to permit replication by another researcher and that all data (subject to appropriate safeguards to maintain privacy and confidentiality) be fully documented and made available for independent examination" (http://www.aapor.org/Content/aapor/Resources/ForResearchers/BestPractices/default.htm). In effect, ethical reporting of survey results requires that information be provided that will allow other researchers to evaluate and replicate the research and enable consumers to understand how the data were collected and what conclusions the study reached.

Among the elements that should be included in survey reports are the sponsorship of the survey, the questionnaire or the full text of the questions asked, and a description of the sample design (a complete list of the recommended elements to be included in survey reports can be found at (http://www.aapor.org/Content/aapor/Resources/ForResearchers/BestPractices/default.htm).

One key component of any reporting—and one that is often missing—is the response rate. Organizations, such as AAPOR and the Council of American Survey Research Organizations, have developed standard definitions and procedures for calculating response rates. Given the concern over response rates described previously, the availability of comparable measures of response rates across surveys provides one indicator of survey quality or the extent to which researchers were able to obtain cooperation from those selected to participate. Despite the importance of this measure many surveys, particularly those that are Internet-based or telephone surveys that use the interactive voice response (IVR) method of data collection, fail to report this information.

A technique that survey researchers frequently employ to adjust for potential bias resulting from non-coverage or non-response is weighting the data, generally in an attempt to have the characteristics of survey respondents match more closely some known characteristics of the population (Battaglia et al. 2006). Any weights used in the analysis should be reported, as should information on the characteristics on which the respondents differed from the population and any analysis the researchers performed to try to account for these differences.

Ethical survey reporting also requires that the findings and conclusions are consistent with the information provided by respondents. Data should be analyzed using statistical techniques that are appropriate for the type of data collected. The number of cases on which any analysis is based should be provided, together with the sampling errors associated with the reported statistics.

In extensive surveys it may not be feasible to report completely on all aspects of the study, but the results should be presented in a way that represents a balanced picture of the findings; investigators should not report only those findings that support their hypothesis or highlight those that are most favorable to the sponsor of the survey.

Any limitations of the data, including those related to non-coverage of the population, measurement error, non-response bias, and other factors that may affect the quality of the data should be acknowledged. Any potential conflicts of interest that may influence the interpretation of the results should be reported. When study results are released, ethical survey researchers should be willing to provide detailed descriptions of the procedures used in the study and to answer questions about the study that other researchers or interested individuals may have.

3.5 The Role of Professional Organizations in Ethical Survey Research

In terms of the ethical treatment of human subjects, survey researchers are subject to the oversight of Institutional Review Boards that review research protocols and assure that the rights of survey participants are protected. With respect to

the ethical design of a survey, however, no similar review group exists. While the findings of survey researchers are certainly subject to peer review when submitted for publication in an academic journal, this review process is distinct from the type of standards that we have described as constituting ethical survey design.

To some extent, this function is performed by professional associations, such as AAPOR, CASRO, the American Sociological Association, and the Market Research Association. Professional associations such as these have developed codes of ethics to cover a variety of issues (Kennedy 2001). Although these codes vary significantly in terms of the specific issues they address, they each serve to establish guidelines for the conduct of research to which their members should adhere.

In general, these codes address the conduct of research with human subjects. A section of the AAPOR code, for example, establishes principles of professional responsibilities in our dealings with people, including doing no harm to subjects, the voluntary nature of participation, and the confidentiality of information. Beyond this, the AAPOR code addresses principles of professional practice, such as developing research designs and using analytic methods that assure the reliability of results, and establishes standards for disclosing results, including reporting the sponsor of the survey, the sample design, the questions used, and the precision of the results (http://www.aapor.org/AAPOR_Code_of_Ethics/4249.htm).

In comparison to the AAPOR code, the Code of Ethics of the American Sociological Association (ASA) is much more extensive. In addition to addressing issues related to the protection of human subjects, the conduct of research, and the disclosure of results, the ASA code addresses issues related to, among other things, nondiscrimination, non-exploitation, harassment, employment decisions, plagiarism, and authorship credit (http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf).

The CASRO Code of Standards and Ethics for Survey Research establishes four fundamental ethical principles for research: respondents should be (1) willing participants; (2) informed about the survey's intentions and how their personal information and survey responses will be used and protected; (3) sufficiently satisfied with the survey experience; and (4) willing to participate again in survey research (http://www.casro.org/pdfs/10CodeOfStandards.pdf). The CASRO code places a great deal of emphasis on privacy and the confidentiality of information, and gives particular importance to Internet research and data security, especially e-mail solicitation and the use of "active agent technology," or software or hardware devices that capture the behavioral data about subjects in background mode.

Different professional associations also vary in the extent to which they have power to enforce the provisions of their codes. Some have no enforcement procedures and have codes that serve as guidelines to the conduct of research. Others have provisions that are enforceable, but have few sanctions, while others have potential sanctions that may include private or public reprimands, denial of association privileges, or termination of membership.

Differences among the codes of these professional organizations illustrate the varying interests of their membership and the different concerns that are considered to be most important to the type of research in which members are typically engaged. Despite these differences, these codes have a common goal: to provide guidance on the conduct of research. Such codes play a valuable role not only in reinforcing the principle underlying regulations for the ethical conduct of research involving human subjects but in establishing professional standards for designing research and the reporting and dissemination of the results from these investigations.

3.6 The AAPOR Transparency Initiative

In addition to codes of ethics, one approach to fostering the ethical conduct of research is the transparency of research procedures. The use of questionable ethical procedures is less likely—or at least more likely to be detected—if

information about the conduct of a study is readily available to other researchers as well as to the public. In November 2009, AAPOR launched a transparency initiative for reporting and disclosing information for public access. Among the goals of this initiative are improving and advancing survey methods and enhancing the cooperation of the public, upon whom survey research depends. By providing transparency about the methods used in collecting survey information this initiative hopes to reduce public skepticism about how surveys are conducted and the results reported.

3.7 Summary

The various uses of polls and surveys play an important role in modern society. From extensive investigations that gather information to track the correlates of disease and those that provide surveillance of public health to polls that examine the factors underlying electoral support or those that are used in the evaluation of programs, surveys provide information that is of great interest to researchers, policy-makers, affected populations, interest groups, and the general public. As this handbook's description of the various components that comprise a well-done survey have made evident, producing quality survey information is a demanding task. To do so ethically, treating survey respondents with respect and dignity and reporting research results in a manner that most faithfully addresses the underlying substantive concerns of a study, is the obligation of all those who engage in the survey process.

References

American Association for Public Opinion Research. *Best practices for survey and public opinion research*. Available at: http://www.aapor.org/Content/aapor/Resources/ForResearchers/BestPractices/default.htm.

American Sociological Association. *Code of ethics*. Available at: http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf.

Battaglia, M. P., Frankel, M. R., & Link, M. (2006). *Weighting survey data: How to identify important poststratification variables*. Paper presented at the annual meetings of the American Association for Public Opinion Research, Montreal.

Berry, S. H., Pevar, J. S., & Zander-Cotugno, M. (2008). *The use of incentives in surveys supported by federal grants*. Santa Monica, CA: RAND Corporation.

Bishop, G. F. (2004). *The illusion of public opinion: Fact and artifact in American public opinion polls*. Latham, MD: Rowman and Littlefield.

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1978). Effects of question wording and format on political attitude consistency. *Public Opinion Quarterly, 42*, 81–92.

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1984). What must my interest in politics be if I just told you 'I don't know'?" *Public Opinion Quarterly, 48*, 510–519.

Centers for Disease Control and Prevention. *Tuskegee study of untreated syphilis in the Negro male*. Available at: http://www.cdc.gov/tuskegee/index.html.

Council of American Survey Research Organizations. *Code of standards and ethics for survey research*. Available at: http://www.casro.org/codeofstandards.cfm.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 64*, 87–98.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken: Wiley.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household surveys*. New York: Wiley.

Groves, R. M., Fowler, F. J, Jr, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. Hoboken: Wiley.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly, 70*, 759–779.

Kennedy, J. M. (2001). *Ethics codes and survey researchers*. Paper presented at the annual meetings of the American Association for Public Opinion Research, Montreal.

Lavrakas, P. J. (2010). Telephone surveys. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed.). UK: Emerald Group Publishing Limited.

Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.

Miller, J. M., & Wedeking, J. (2003). The harder we try, the worse it gets? Examining the impact of refusal conversion and high callback attempts on the quality of survey data. Paper presented at the annual meetings of the American Association for Public Opinion Research, Nashville.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and*

guidelines for the protection of human subjects of research. Washington, DC: U.S. Government Printing Office. Available at: www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the internet. *Journal of Social Issues, 58*(1), 161–176.

Payne, S. (1951). *The art of asking questions*. Princeton: Princeton University Press.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly, 70*, 737–758.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Singer, E. (2006). Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*, 637–645.

Singer, E., & Bossarte, R. M. (2006). Incentives for survey participation: When are they coercive? *American Journal of Preventive Medicine, 31*(November), 411–418.

Singer, E., & Presser, S. (2008). Privacy, confidentiality, and respondent burden as factors in telephone survey nonresponse. In J. M. Lepkowski, et al. (Eds.), *Advances in telephone survey methodology* (pp. 449–470). Hoboken: Wiley.

Sudman, S. (1998) Survey research and ethics. In J. W. Alba, & J. W. Hutchinson (Eds.), Advances in consumer research, (Vol. 25). Provo, UT: Association for Consumer Research.

Tarnai, J., & Moore, D. L. (2008). Measuring and improving telephone interviewer performance and productivity. In J. M. Lepkowski, et al. (Eds.), *Advances in telephone survey methodology* (pp. 359–384). Hoboken: Wiley.

Tuchfarber, A. J., Oldendick, R. W., & Bishop, G. F. (1979). *Understanding and dealing with the refusal problem in telephone interviewing*. Paper presented at the annual meetings of the American Association for Public Opinion Research, Buck Hill Falls: PA.

United States Government, International Military Tribunal. (1949). *Trials of war criminals before the Nuremberg military tribunals under control council law no. 10,* (Vol. 2, pp. 181–182). Washington, DC: U.S. Government Printing Office.

An Overlooked Approach in Survey Research: Total Survey Error

4

René Bautista

The author is grateful to NORC Center for Excellence in Survey Research (CESR) for the grant support provided to complete this chapter. All errors and omissions, of course, remain the sole responsibility of the author.

4.1 Social Research and Survey Methods

Social scientists are constantly challenged to understand and measure human behavior. Researchers are frequently pressed to make inferences about populations of interest based on individual level data, but usually have limited resources at their disposal. The challenge is even bigger when researchers aim to obtain accurate, reliable, cost efficient, and representative results of the population of interest. Fortunately, there are scientific tools that allow social researchers to gather valuable data. Scientific surveys are useful instruments to measure concepts and behaviors based on a sample of cases.

Recent developments in the field of survey methodology are having a great impact upon the way individual data-based research is conducted in the social sciences. Advances in survey methods are changing the way data collection strategies are conceived. Ultimately, they are changing the way we interpret survey data. There are sources of error in surveys, nonetheless, that limit our ability to make inferences about populations of interest. The degree of accuracy in a survey (that is, obtaining survey measures that reflect population parameters or "true values") depends on several sources of error.

R. Bautista (✉)
NORC at the University of Chicago,
55 East Monroe St, 20th floor, Chicago,
IL 60603, USA
e-mail: Bautista-Rene@norc.org

In this chapter, we will describe potential sources of error in a survey, and we will discuss how survey design features can help us minimize the effect of error sources on survey estimates. The information presented in this chapter is based on current theories in the field of survey methodology as well as practical experience. We will discuss a usually overlooked survey methodology paradigm, the Total Survey Error (TSE). The TSE is a useful framework to understand survey errors in a comprehensive manner. Although the TSE can be discussed in mathematical terms, we will not follow a mathematical style. Instead, we introduce the TSE as way of thinking about the tradeoffs usually involved in survey research. The aim of this chapter is to provide a description of the TSE in such a way that the reader finds it useful, especially during the early planning of a survey.

4.2 The Notion of Error in Surveys

The notion of "errors" or "limits" is unavoidable in the context of scientific surveys. Although social researchers are aware of the existence of errors in surveys, the interpretation of errors tends to be incomplete and frequently oversimplified. The notion of error usually is associated with just one of several types of errors: the "sampling error." Sampling error is perhaps the most well-known source of error in surveys, but it is not the only one—as we will see later on. The sampling error occurs due to the fact that we are studying a fraction of the population randomly

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_4, © Springer Science+Business Media New York 2012

selected in lieu of the whole population. This concept is also commonly referred to as "theoretical margin of error," or simply "margin of error." The underlying idea is that of a statistical basis to create "confidence intervals" or boundaries within which the statistic of interest (whether a proportion, a mean, regression coefficient, or other statistic) is expected to lie. Theoretically, as the sample size increases, the margin of error decreases.

The margin of error has been widely used in reporting univariate survey statistics in almost any probability survey. Starting with Neyman's (1934) seminal article, the field of statistics devoted a great deal of effort to the understanding of the sampling process. With today's access to information technologies, it is relatively easy to calculate the margin of error for proportions under the assumption of simple random sampling.

Unfortunately, the routinely reporting of the sampling error in surveys has mistakenly lent credence to the idea that surveys need to be designed based on the "margin of error" alone. Even more, sometimes scientists misguidedly assume that errors other than those related to sampling will occur at random, and will cancel each other out with no impact on the accuracy of the estimate.

The sampling error is a very limited representation of errors in survey design, leading social scientists to overlook other sources of error. Nonetheless, our understanding of errors in surveys is rapidly changing. In recent decades, starting with Groves' (1989) book *Survey Errors and Survey Costs*, methodologists have begun to devote a great deal of effort to investigate sources of "nonsampling errors." Over the years, it has become clear that even if we had resources to collect information from the whole population of interest (that is, if we conduct a census to eliminate the sampling error), a poorly design measurement instrument, inadequate response categories, a faulty opening script, the appearance of the interviewer, the lack of interviewer training, the channel of communication used to gather data, and other important survey design features, may jeopardize the quality of the data. Questionable data collection strategies may result in a significant waste of time and resources—even if the "margin of error" is presumably small.

4.3 The TSE Approach

Developments in the field of survey methodology during past decades have incorporated contributions from a variety of fields besides statistics. Disciplines such as psychology, sociology, economy, communication, linguistics, cognitive sciences, computer sciences, and others, are reshaping the notion of error in surveys. Currently, the field of survey methodology is multidisciplinary, and its scientific value is reflected in how survey errors are understood in the field (see for instance, Biemer 2010; Biemer and Lyberg 2003; Groves and Lyberg 2010; Groves 1989, 2004; Leeuw et al. 2008; Lessler and Kalsbeek 1992). In survey methodology, sources of error are studied from different perspectives because errors can be different in nature, but they all have a potential effect on survey results.

The TSE paradigm is a useful integrated theoretical framework that will help us understand survey errors. This paradigm puts emphasis on data quality, and help researchers prepare and evaluate survey designs. Social scientists can organize their survey preparation using as a theoretical reference the TSE. TSE considers important sources of error altogether. Ultimately, the TSE goal is to obtain survey statistics as accurate as possible, given available resources.

The TSE is a function of two kinds of limitations: "sampling" and "nonsampling" errors.

The literature in survey methodology typically breaks down the nonsampling errors into at least three sources of error: coverage error, nonresponse error and measurement error. Each source may contribute to the total error of the survey. Thus, we define TSE as the sum of sampling and nonsampling errors, as shown in Fig. 4.1.

We will discuss later on in this chapter each component of the TSE individually.

4.4 Two Building Blocks: Bias and Variance

Each of the error sources in the TSE potentially contribute to the *bias* and *variance* of a survey statistic. *Bias* is defined as a systematic departure of our survey estimate from the "true" population value—such departure can occur in a particular direction. *Variance* is defined as the uncertainty introduced in our survey estimate to represent the "true" population value. The terms *bias* and *variance* have been adopted into the survey methodology field from the survey statistics field; in particular, they come from a concept known as the Mean Square Error (MSE). The MSE is a function of *bias* and *variance*. In short, MSE is the expected difference between an estimated statistic and the true value due to systematic (*bias*) and random variations (*variance*) over theoretical repeated events. Formally, the MSE = bias^2 + Variance.

In statistics, the MSE has been used as composite measure that helps identify unbiased mathematical procedures to estimate population parameters; these procedures are also known as *estimators*. The resulting MSEs from different procedures, or estimators, are usually compared to each other. The best estimator is said to be "unbiased" when *bias* is non-existent. The underpinnings of the MSE are beyond the scope of this chapter; nonetheless, the idea of an unbiased estimator helps us understand the logic behind the TSE: TSE aims to achieve accurate survey statistics.

In the same MSE fashion, each component of the TSE can be theoretically decomposed into bias and variance; thus the TSE is equal to the sum of sampling bias, sampling variance, coverage bias, coverage variance, nonresponse bias, nonresponse variance, measurement error bias, and measurement error variance. In this classification of survey errors, when a systematic error occurs leading to over- or under-estimate the "true" value, we are in the presence of *bias*; when errors occur in an unrelated way introducing uncertainty about the "true" value that we want to estimate, we are in the presence of *variance*.

4.5 Specification and Processing Errors

There are two other types of errors frequently mentioned in the context of the TSE framework: the *specification error* and the *processing error*. *Specification errors* occur when there is no clear connection between theoretical concepts or constructs, and survey variables. This kind of error happens when key concepts are excluded or misspecified during early stages of the study. In other words, it is the consequence of an ill-defined research protocol or the lack of a research protocol at all.

Specification errors are more frequent than one may think. Less experienced social researchers frequently start crafting a survey questionnaire without having a clear written research protocol. Needless to say, such practice leads to weak conclusions as the resulting data will not reflect specific social constructs. Specification error can be easily minimized if researchers devote more resources (i.e., time and effort) to carefully select their theories, concepts, hypotheses, and survey metrics, before developing a questionnaire. Ideally each survey variable should be related to a concept, and each concept should be related to a hypothesis, which in turn, should be derived from theoretical knowledge.

Processing errors refer to flaws that occur once the survey data have been collected. These errors are often introduced inadvertently. They include errors due to computer programing, coding of open ended questions, data entry of close ended questions, data cleaning, imputation of missing data, weighting, and data reporting or tabulation.

Processing errors can be minimized with a careful revision of protocols that will be implemented after the data collection period. Implementation of quality control and quality assurance procedures throughout the data processing stage is usually more efficient, than a full final revision of the data, once they all have been processed.

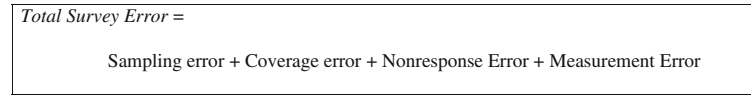


Fig. 4.1 Sources of error which can have an impact on the bias and variance of a estimate

A good practice is to document all coding rules and programing codes used during data processing. A well-documented data processing plan makes easier the identification of errors at this stage.

Specification and processing errors can drastically affect our survey estimates. In this chapter we choose not to include them as part of our theoretical equation in Fig. 4.1, because they do not have an obvious *bias* and *variance* component, as the other components have. Despite the fact that we are excluding the specification and processing errors from our TSE equation, the reader needs to be aware of their existence and to take preventive measures while planning a survey, to minimize their effects.

4.6 Sampling Error

Cost constraints commonly dictate the way empirical research is conducted in the social sciences. Since it is nearly impossible to collect data from all of the elements in the population of interest, scientific sampling has proven to be a valuable tool. Consequently, social researchers are in constant need of utilizing probability sampling to establish conclusions about a certain population.

Sampling error occurs because we do not collect information from all individuals in our population due to the fact that we analyze a sample of cases (instead of conducting a census of the whole population). Sampling error can be further divided into two components: *sampling bias* and *sampling variance*. Suppose we were to repeat several times the sampling process by drawing the same number of cases (not the same individuals) in exactly the same manner, out of the same population. Chances are that we would obtain a slightly different result every time. This is called the *sampling variance*. The basis to derive the variance is theoretical because we usually do not have resources to conduct many surveys under the same survey conditions.

A key aspect of scientific sampling is that all members in the population of interest—or target population—should have a known, non-zero, probability of being selected. When this condition is not satisfied, our sample is likely to be biased. *Sampling bias* occurs if the sampling process systematically gives no probability of selection to some individuals of the population, or gives a disproportionately small or large probability of selection, to some subjects. A way of minimizing the sampling bias is to give all subjects in the population an equal chance of selection.

If sampling bias exists due to different probabilities of selection, we could compensate for such differential probabilities of selection by means of *weights*—an adjustment that gives individual a disproportionate weight relative to their original weight. However, such adjustment may increase the uncertainty of our estimates, because subjects that were drawn with different probabilities will now have a disproportionate influence in the sampling variance.

The conceptualization and estimation of sampling variance heavily rely on theoretical statistical formulations because we do not actually draw and collect information from multiple surveys at the same time, under the same survey conditions. The computation of the sampling bias and sampling variance depends on assumptions about what would have happened if the sample were drawn numerous times—hence the name of "theoretical margin of error."

The sampling variance is a concept we use to establish the boundaries within which our estimate is expected to lie. Under this logic, the greater the number of cases in the sample, the smaller the variability. Table 4.1 illustrates the effect of the sample size on the "margin of error." These computations assume that cases were selected using simple random sampling. Also, they assume that the proportion of an estimated survey statistic is equal to 0.5,

Table 4.1 Margin of error for proportions under simple random sampling, assuming a 95 % confidence level

Number of cases	Margin of error (%)
100	± 9.8
500	± 4.4
1,000	± 3.1
5,000	± 1.4
10,000	± 1.0

and assume that the true value lies within the proposed range, 95 % of the time.

If we select 100 cases, we would have an estimate with a boundary of ±9.8 % points; but, if we set our sample size to 500 cases, we would reduce our uncertainty by almost 5 and a half percentage points (= 9.8–4.4 %). Increasing sample cases does not decrease the margin of error in a linear fashion, nonetheless.

If we compare a survey design having 1,000 cases with a survey design having 10,000 cases, we would see a reduction in the margin of error of nearly 2 % points (3.1–1.0 %). Increasing our sample size from 1,000 to 10,000 would improve our precision, but also would increase the costs of data collection efforts.

Survey designs with large number of cases represent also a great challenge in terms of survey quality. Having more cases mean hiring and training more survey personnel, doing more attempts to persuade people to participate in a survey, doing more supervision of data collection and data processing, more data management, and other burdensome logistic aspects. In some instances, it would be preferable to have a manageable, cost effective, closely monitored sample of cases, than a very large sample that can potentially increase other sources of error.

There are occasions of course, where a larger sample of cases may be needed. For instance, if the researcher wants to conduct analysis of subgroups in the population with relatively high statistical precision, then more sample cases are necessary. The point to emphasize, nonetheless, is that there are tradeoffs between feasible well-monitored samples and statistical power.

Besides increasing the sample size, *sampling variance* can be also minimized by using a procedure known as *stratification*. For instance, before drawing a sample we could divide our population into mutually exclusive groups (known as strata), based on available information. Once people are assigned to a particular stratum, we can proceed to select cases randomly from all of the strata. A population can be previously stratified into age groups, occupation, gender, literacy level, or other meaningful criteria. Stratification usually represents a type of assurance that we are randomly drawing cases from all possible groups in the population, therefore, decreasing the uncertainty of survey results.

Sometimes grouping occurs naturally in the population of interest. People may be grouped into clusters defined as classrooms, neighborhoods, offices, election precincts, and so forth. Clusters represent a way of reducing costs because there are marginal costs associated with collecting more information from the same cluster. The researcher may decide to take a sample of clusters in order to collect information and calculate inferential statistics; for example, but the researcher needs also to be aware of the fact that cluster sampling may increase the variance associated with survey estimates.

Cluster cases are usually related to one another. People in the same clusters are exposed to the same social environment, information, social influences, context, and so on; consequently, they tend to be very similar. Thus, even if data were collected from all of the individuals in the selected clusters—creating the illusion of an increased sample size—the research will not gain more knowledge from such a sample because those cases are correlated.

In practice, cost-effective scientific sampling strategies combine stratification and clustering. They may even use different selection probabilities for people from some groups. When the

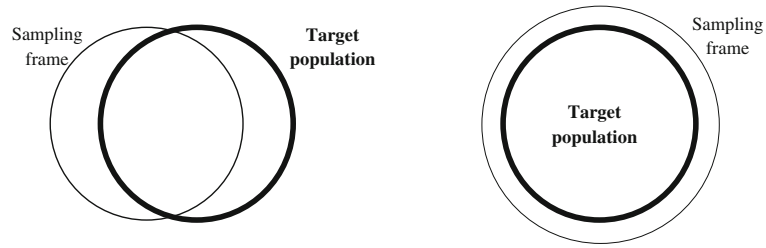


Fig. 4.2 Coverage of population by sampling frame

(a). Undercoverage **(b).** Overcoverage

selection of sample cases is made using such strategies, the sampling approach is called *complex random sampling*. When a survey is designed using a complex sampling approach, the margin of error (i.e., the variance) is not calculated in the same way as in simple random sampling. Clusters, strata, and differential probabilities of selection need to be considered in the computation of the variance.

There are tradeoffs in the selection of strategy for sampling that may have an impact on the variance of our survey estimates. While some elements may minimize the sampling error, others can potentially increase it. The researcher needs to carefully consider such tradeoffs to have a cost efficient and statistical efficient sampling design.

4.7 Coverage Error

In any scientific survey, we need a listing of elements in the population to be able to draw a sample. This list of elements in the population is known as *sampling frame*. A sampling frame is usually comprised of households, individuals, telephone records, mailing addresses, administrative records, institutions, and others. Existing lists usually serve as sampling frameworks, and represent great advantages to survey projects because information is readily available. In cases where existing sampling frames do not cover the population of interest, researchers may need to create their own sampling frame in order to cover the entire population under study.

Coverage error occurs when individuals in the population of interest are missing from the sampling frame use to draw a representative sample. Coverage error can be determined even before drawing the sample. This is possible because we can anticipate that some elements of the population are not part of the sampling frame.

Coverage error can take the form of undercoverage or overcoverage (Fig. 4.2). When we systematically exclude population elements from our sampling frame, we are dealing with undercoverage. When elements other than those described as our target population are part of the sampling frame, we are in the presence of overcoverage.

Coverage error can be broken down into *bias* and *variance*. *Coverage bias* occurs when elements in the population are systematically excluded from the sampling frame. For example, if we conduct a survey of students in a particular college, we may ask the administrative office for a list of the students' email addresses, with the purpose of using it as our sampling frame. Students who do not have an email address registered in the administrative office, will not be part of the sampling frame. Accordingly, the bias of survey statistics obtained from our study is likely to increase due to coverage error.

An easy solution to coverage bias is to recognize the limitation of the sampling frame and redefine the population of interest in light of the characteristics of the sampling frame. The researcher can describe upfront the elements in the population from which we are able to make valid statistical inferences. In our example, the target population would be college students with a valid email address registered in the administrative office.

An alternative solution for minimizing bias due to coverage error is to create a sampling frame in such a way that it includes all individuals

in the population. Clearly, this option can be very costly because it requires an enumeration of all members in the target population. An additional solution to minimize bias due to coverage is to use supplemental sampling frames.

The use of two or more sampling frames requires posterior statistical adjustments to obtain unbiased results. Survey estimates from two or more sampling frames would be combined by means of weighted averages. Nevertheless, since individuals will be disproportionally contributing to this averaged number, they are likely to increase the uncertainty around our resulting survey estimates. Put differently, in theoretical replicates of the selection process using the same sampling frames, and the same survey design, we will observe a great deal of variability on our estimates, leading to an increase in *variance* due to coverage error.

A careful selection of an up-to-date sampling frame that covers as many sampling units as possible, along with a thoughtful definition of the target population may help minimize coverage errors. If necessary, a supplementary sampling frame for those undercovered individuals may help as well. For instance, if the intention of the researcher were to draw a sample of people from a minority group for which separate sampling frames exist, then the researcher may need to design a multiple sampling frame survey. Importantly, prior to drawing the sample, the researcher needs to exclude ineligible individuals from the sampling frame to avoid misleading conclusions.

4.8 Nonresponse Error

Ideally, in a study based on scientific sampling, all selected individuals should answer all of the questions included in the data collection instrument. In practice, this rarely happens.

Nonresponse error occurs when part of the information sought is not collected because sample individuals choose not to respond; as a result, survey statistics may or may not be representative of population parameters. If sample individuals choose not to answer just some of the questions, we will experience *item nonresponse*. If sample individuals refuse to answer the entire questionnaire will be in the presence of *unit nonresponse*.

$$\text{Non response bias} = (\bar{Y}_r - \bar{Y}_m)\left(M/N\right)$$

Fig. 4.3 Computation of nonresponse bias

As it was the case with sampling and coverage error described in the two previous sections, nonresponse error can also be decomposed into *nonresponse bias* and *nonresponse variance*. *Nonresponse bias* occurs when responses obtained from those who accepted to participate in the survey are systematically different from those who chose not to participate. Such bias can increase depending on the proportion of people who chose not to participate. The proportion of people who participate in the survey is typically represented as a *response rate*.

The response rate in a survey is incorrectly considered an indicator of survey bias. Response rate alone is just one part of the story in the estimation of nonresponse bias. A basic equation can help us understand the role of (non)response rates (for details on response rate calculation, see AAPOR 2009). To obtain the nonresponse bias, one could calculate the difference between the respondent and nonrespondent means multiplied by the nonresponse rate (Fig. 4.3):

In Fig. 4.3, the expression $\bar{Y}_r - \bar{Y}_m$ represents the difference in the means of respondents ($\bar{Y}_r$) and nonrespondents ($\bar{Y}_m$), and M/N represents the ratio of nonrespondents to all sample people (M = number of respondents; N = total of respondents and nonrespondents). Although the mean of nonrespondents ($\bar{Y}_m$) cannot always be determined, the equation shows the theoretical logic of the nonresponse bias.

Using response rates alone as a way of judging the quality of a survey can be misleading. A high response rate (i.e., a small nonresponse rate) can still introduce a sizable bias in an estimate if the difference between those successfully sampled and those who refuse, is considerable. Likewise, a small response rate (i.e., a large nonresponse rate) does not

necessarily mean there is bias if the difference between respondents and nonrespondents is negligible.

Statistical adjustments could be implemented to compensate for nonresponse bias. In the case of *unit nonresponse*, weighting techniques can help reduce the difference between respondents and nonrespondents once the data have been collected. In the case of *item nonresponse*, single or multiple imputation techniques (i.e., entering model-based simulated values for data which do not exist) can help represent data that otherwise would not be available for analysis. Statistical approaches that compensate nonresponse bias can introduce uncertainty to our results. That is, we modify the bias to the expense of the variance.

If we theoretically repeat exactly the same survey over and over again, using the same survey design and the same weighting techniques, the range of possible survey outcomes is likely to be wider than the one we would have had, if those values had not been missing. The fact that we would give respondents disproportionally different weights (in the case of unit nonresponse) or simulating data using imputation techniques (in the case of item nonresponse) is likely to increases the *variance*.

There are survey design features that can help us improve our cooperation rates, whether at the item or unit level. For example, if the researcher anticipates higher levels of item nonresponse because of the sensitive nature of the questions, the researcher may consider a self-administered mode of data collection (whether paper- or computer-based instruments) instead of an interviewer-administered method.

If solutions to improve response rates include the use of information technologies, the researcher needs to be aware of the tradeoffs involved. At this time in history, younger and educated people tend to be more likely to have positive attitudes toward technology than older and less educated people, making those who are open to technologies more likely to participate in the survey using information technologies.

Offering incentives, whether monetary or nonmonetary, may also help motivate people to participate in a survey. The kind of incentives offered to sample persons depends on the kind of target population. Sometimes offering a copy of the final report is a good incentive. Other times, people are not interested in the topic of the survey and they place value on time, thus monetary incentives become more relevant in the study.

Overall, researchers should strive to reduce the burden imposed on respondents. This means that the researcher can potentially use multimode data collection approaches. Some of the information can be collected using a particular mode among some subgroups, and other subgroups may answer using a different mode of data collection. For instance, older and less educated people may be measured using a mail survey, and younger and more educated people can be surveyed using web-based modes.

Regardless of the mode of data collection, the researcher needs to be aware that respondents need some kind of reward (whether a social or economic reward), and that such reward needs to be delivered. There are tradeoffs, of course, in the use of incentives. Some respondents may rush through the survey so they can obtain the promised reward while others may be offended if money is offered. A careful selection of rewards (monetary or in-kind) as well as a carefully designed data collection protocol, can improve response rates. Toepoel (Chap. 13) provides a thorough discussion on the effect of incentives in reducing response rates and decreasing nonresponse bias, and the potential consequences of such incentives to quality of responses.

The way respondents are first contacted to participate in the survey has a great impact on their decision. To the extent that respondents perceive the survey solicitation as a legitimate request, survey cooperation is likely to happen. Also, it is important that respondents perceive that his cooperation and the corresponding exchange of information is beneficial for both, the respondent and the researcher. Likewise, if the respondent perceives that the sponsor of the survey is a prestigious institution or an authoritative figure, survey cooperation is likely to improve.

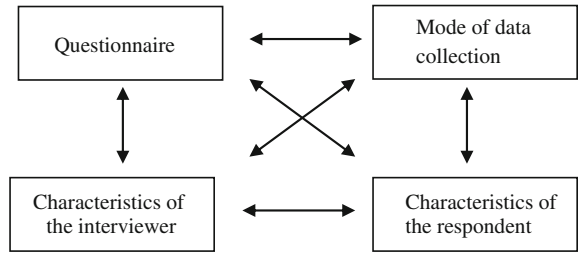


Fig. 4.4 Interacting sources of measurement error

Further, the researcher needs to carefully design a first-contact protocol as well as a follow-up protocol for refusals. These two protocols usually involve preparation of letters, friendly reminders, preparation of standardized scripts, and intense interviewer training on persuasion of sample individuals.

A follow-up protocol for nonrespondents is just as important as the first-contact protocol. The researcher needs to have these protocols prepared before starting the data collection process. Respondents may reply they are busy at the moment of contact, that they do not feel completely comfortable answering questions, that they have answered many surveys already, among many other answers. Depending on the type of refusals (whether "soft" or "core"), more experienced and better trained interviewers should try to persuade them again. Since refusal conversion is usually costly and time consuming, the researcher may consider collecting data from a subsample of nonrespondents to have at least an approximation of the bias due to nonresponse. For further discussion on why people agree or refuse to participate in surveys, see Albaum and Smith (Chap. 11), and Glaser (Chap. 12).

4.9 Measurement Error

In survey research we aim to represent social constructs with survey measures as accurate as possible. Unfortunately, several aspects related to the survey design impose restrictions to gauge the "true" value of a social concept. The "true" value can be an observable variable—which can be directly measured, such as weight, size, age, income, and others—and can be an unobservable or latent measure—and cannot be directly measured such as intelligence, happiness, quality of life, and satisfaction among others.

Measurement error occurs when there are differences between the estimated value and the "true" value due to survey design elements. Explicitly, measurement errors come from inaccuracies in responses due to four sources (Fig. 4.4): the questionnaire, the mode of data collection, characteristics of the interviewer, and characteristics of the respondent. All these sources interact and can be the cause of measurement error at the same time.

When measurement error occurs in a systematically fashion, the concept under study is misrepresented in a particular direction—whether positively or negatively. This is known as *measurement bias*. If the error occurs in a random fashion (not in a particular direction) over theoretical replicates of the same survey design, then we would likely say that we have an increase in *variance* due measurement error.

Examples of survey design aspects that can lead to measurement error include poor question wording, unclear question instructions, erroneous skip patterns, lengthy questions, inadequate response options, the topic of the questionnaire, timing, sponsorship, confusing visual designs, data collection methods, interviewer characteristics, faulty interviewer training, interviewer actions (whether indicated by the training or unforeseen behaviors), interviewer expectations, respondent reactions (whether to the topic or to the interviewer appearance), social pressure in the interviewer-respondent interaction, and respondents' memory erosion among many others.

Responses are provided based on the interpretation of interactive survey design elements. For instance, the measurement process in interviewer-administered modes requires conversational actions between interviewer and respondents to be able to complete the questionnaire. Thus, the researcher needs to decide if the interviewing protocol will utilize a fully scripted, standardized approach (i.e., having a standardized questionnaire), or if the interviewing protocol will take a flexible approach (which means the interviewer will have latitude to rephrase questions as needed).

One of the reasons for choosing standardized interviewing techniques is because interviewers represent a source of error. Different interviewers may interpret questions differently, deviating from the intended meaning of each question; therefore, they should not be allowed to change words or terms. If interviewers introduce unanticipated variability (whether systematic or random) the accuracy of the survey data would decrease.

If we choose a flexible interviewing approach is because we assume that respondents are a source of measurement error. Respondents may have different interpretations of the terms; therefore, the interviewer should intervene to provide clarification and assistance to gather accurate data.

In the end, these two competing approaches aim to improve accuracy of the data; they just identify the source of measurement error differently. Standardized interviewing aims to minimize measurement error due to the interviewer. Flexible interviewing aims to minimize measurement error due to the respondent. They both have advantages and disadvantages. The researcher needs to decide the most suitable approach depending on the anticipated error source in the study: interviewers or respondents, and depending on the challenges the researcher faces.

To date, the majority of surveys utilize standardized interviewing techniques. Under this approach, interviewers are trained to have a neutral position toward respondents' answers, read questions exactly as worded, follow established skip patterns, probe inadequate answers in a nondirective manner, record answers as expressed, and maintain an interpersonally, nonjudgmental relationship with the respondent.

Despite these guiding principles in standardized interviewing, the interviewer faces situations where respondents do not provide information needed due to a poorly developed question. Interviewers are naturally tempted to change or reword the question in order to get an answer—even when they are trained to leave the interpretation up to the respondent. This emphasizes the importance of resources allocated to questionnaire design and to interviewer training.

Seemingly irrelevant aspects of questionnaires may introduce measurement error to our survey. The order in which questions appear in a questionnaire may have an impact on the way questions are interpreted; namely, preceding questions affect the way in which subsequent questions are answered. Respondents usually provide answers in the context of an interview, not considering each item separately.

Another relevant feature in questionnaire design is the order of response options. When respondents do not have strong positions or definitive answers in some questions, the order of response categories may be a matter of highly importance. In this context, the mode of data collection may have an influence on the answer. For example, in auditory data collection modes (such as in a telephone surveys), the last response option that the respondent hears is more likely to be selected. In visual data collection modes (such as in a paper and pencil survey, or a web survey), the first visualized response category is more likely to be endorsed, relative to the rest of response options.

A solution to avoid measurement error due to the order of response categories is to rotate response options throughout the administration of the survey; however, that might not be feasible in all modes of data collection. In the context of mail surveys, for instance, it may require creating different versions of the same questionnaire. This can complicate the data entry process once the forms have returned from the field. Clearly, it will be easier to implement

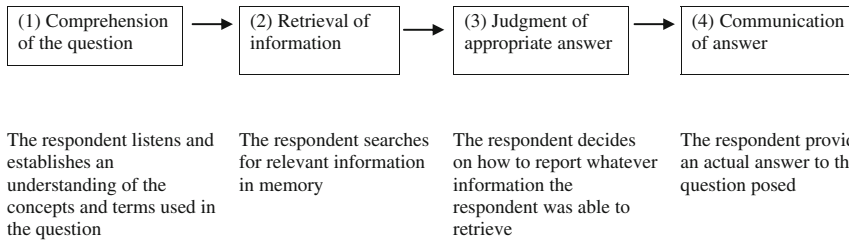


Fig. 4.5 Basic theoretical model of the survey response process

rotation of response categories in computer-assisted surveys, but that would require a significant investment in technology resources.

The selection of data collection method (telephone, Internet-based, in person, paper and pencil, computer-assisted, and others) is usually tied to other survey design decisions such as sampling frame, sampling strategy, expected response rates, and costs. An Internet-based probability survey could be feasible if we have an adequate sampling frame; but it may not be feasible due to uneven access to information technologies among members of the target population.

In general, measurement error is hard to measure because we usually do not know the true value, or because it requires advanced statistical techniques, or simply because we use qualitative methods to study and understand it. However, measurement error is one of the greatest sources of error in survey research. Among the qualitative techniques that can help us understand measurement error (in particular errors related to the respondents) are *cognitive interviews*.

Cognitive interviews are semi-structured interviews that provide insights on the cognitive process that takes place in respondents' mind while answering a survey question. This interviewing technique makes extensive use of a cognitive model which describes the steps respondents take to come up with an answer. The basic foundation of such cognitive model is presented in Fig. 4.5.

While the specifics of cognitive interviewing are beyond the scope of this chapter, the aim of a cognitive interview is to determine what the respondent experiences at each stage of the process. For example, if respondents experience difficulty in understanding some of the terms used in a particular question, or if they constantly ask for repetition of a particular question, that would suggest a problem of comprehension and the researcher would have to make improvements to the question.

If respondents are not able to remember events or specific information (e.g., visits to libraries, purchases of electronic devices, visits to the dentist, crimes, and other information of interest), for example, over a period of twelve months, but is able to remember events in the past six months, that would suggest that a problem exists at the second stage of the cognitive process. In that particular case, a possible solution is to have a shorter period of reference in the question.

If respondents are able to understand the question, retrieve relevant information, but do not feel quite sure about which response option best reflects his or her thoughts, or are confused on the meaning of each response option, it would suggest that problems are occurring at the third stage of the survey response process. Therefore, the researcher would have to modify response options or response scales to make sure he or she is gathering data from respondents accurately.

When respondents feel their answers would not conform to social norms, and they experience social pressure to communicate their answers in a different way than their true answers, it may be indicative of problems related to the fourth stage in the cognitive process model. For example, in a question about attitudes toward abortion, a respondent may not feel

comfortable communicating his or her answer to the interviewer, and may choose to modify their answer on such question.

In that case, the researcher would have to modify the question being posed and how the answers are being collected. Conceivably, self-administered methods could be more convenient for sensitive questions. Perhaps the researcher can take advantage of portable devices (i.e., mp3 players, handheld devices, laptops, and others) to administer the survey and hand the respondent a paper and pencil instrument so only the respondent knows what he or she is answering. This may help minimize the effect of measurement error due to the social pressure that exists in the interviewer-interviewee interaction. Nonetheless, incorporating technology may increase the costs of the project.

In the survey methodology field, there is a growing body of literature devoted to measurement error, and about how social researchers can potentially minimize its effects on survey estimates. Given the fact that sources of measurement error interact altogether, social scientists are encouraged to think comprehensively about these aspects when designing a survey. Solutions involving technology are rapidly changing and are becoming more affordable. When deciding on alternative methods as a way of minimizing measurement error, a revision of the latest literature on data collection methods is highly recommended.

4.10 Measuring the Total Survey Error

Social researchers constantly face a challenge on how to decide about a survey design. The first reaction, very often, is to put most of the funds available into increasing the size of the sample or to improve response rates; however, such a decision might not be necessarily the best choice all the time. There are instances where it is clearly preferable to allocate more resources on interviewer training, questionnaire design, selection of a better sampling framework, or any other survey design feature, than spending resources on getting more respondents. The TSE approach can help social scientists better inform their decisions.

A common question is how to calculate a single measure of TSE to able to compare survey designs. In that sense, the survey methodology field is still evolving. Several mathematical methods could be used to estimate the TSE for survey statistic under a particular survey design. For instance, multilevel models techniques can be used to estimate the influence of interviewer on respondents.

Latent class analysis methods are used to infer unobservable true population parameters, and to compare them to survey estimates. Structural equation models can help represent the correlation between several error sources. Also, methods dealing with inferential statistics are in continuous improvement for better variance estimation processes to account for complex sampling design features.

To date, there is no single straightforward method to compute a measure of TSE. Although the computation of an actual TSE measure may not be readily available in commercial software packages, TSE is still a very useful theoretical framework to help us understand comprehensively aspects that can jeopardize the accuracy of our results. Usually the best decisions in survey research are made when the researcher considers simultaneously the different components involved in the design.

As it should be obvious by now, there are tradeoffs between survey design elements. All decisions have an implication for accuracy, and all decisions are related to costs. There is no one-size-fits-all survey design that can be applied to all research projects. The researcher is encouraged to use the TSE paradigm to guide their discussions and decisions on survey design.

References

AAPOR. (2009). Standard definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys 2009. Available from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly, 74*(5), 817.

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. Hoboken: Wiley.

de Leeuw, E. D., Hox, J. J., Dillman, D. A., & European Association of Methodology. (2008). *International handbook of survey methodology, EAM book series*. New York: Lawrence Erlbaum Associates.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849.

Groves, R. M. (1989). *Survey errors and survey costs, Wiley series in probability and mathematical statistics. Probability and mathematical statistics*. New York: Wiley.

Groves, R. M. (2004). *Survey methodology, Wiley series in survey methodology*. Hoboken: Wiley.

Lessler, J. T., & William D. K. (1992). *Nonsampling error in surveys*, *Wiley series in probability and mathematical statistics. Probability and mathematical statistics*. New York: Wiley.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 97*(4), 558–625.

Part II

Designing The Survey

Common Survey Sampling Techniques

5

Mary Hibberts, R. Burke Johnson, and Kenneth Hudson

In this chapter we provide an introduction to probability and non-probability sampling methods commonly used in quantitative research. Although we pay special attention to sampling issues in survey research, we also briefly explain additional sampling methods used in other types of research.

5.1 Terminology Used in Sampling

Before we begin our discussion of sampling methods, it is important that you become familiar with some specialized sampling terms. First, a *sample* is a group of elements selected from a larger group in the hope that studying this smaller group (the sample) will reveal important information about the larger group (the population). The *elements* that make up samples are the basic units of the population and are most commonly "individuals." However, many different types of elements can be sampled such as "groups" (e.g., schools, grocery stores, student organizations) or "objects" (e.g., scores, library books, training videos). Populations are always larger than samples, and in most cases they are much larger than the samples drawn from them. In sampling, an uppercase N is used to denote the population size (i.e., the total number of elements in a population) and lowercase n is used to denote the sample size (i.e., the number of elements in a sample). For example, if you selected a sample of 100 individuals from a population of 1,000 then your sample size would be written as $n = 100$ and the population size would be written as $N = 1{,}000$. We will show that sampling rules tell you how to select a sample and individual sampling methods follow specific rules.

A *population* is an entire group of persons (or institutions, events, or other objects of study) that one wants to describe and understand. It is the large target group to which the researcher hopes to generalize. In order to generalize from a sample to a population, the researcher typically studies a sample that is meant to be representative of the population. In most cases, it is not practical, or even possible, to gather information from every element in a population. Therefore, researchers use samples to make inferences about the populations of interest. In other words, the population is the total group of elements that you are interested in learning more about and the sample is the group of elements you investigate directly. A few specific populations are all citizens aged 18 or older currently living in the United States, every member of a social networking website, every person at a professional baseball game, every book in a public library, and every voter registered in a particular country.

The goal of sampling is usually to select a representative sample. A *representative sample* is one that is similar to the population from which it

M. Hibberts · R. Burke Johnson (✉)
College of Education, University of South Alabama, Mobile, AL 36688-0002, USA
e-mail: BJOHNSON@usouthal.edu

K. Hudson
Sociology Department of the College of Arts and Sciences, University of South Alabama, Mobile, AL 36688-0002, USA

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_5, © Springer Science+Business Media New York 2012

is drawn in every way except that it has fewer people. Whenever one wants to generalize, it is essential that the sample is as similar to the population as possible. A *statistic* is a numerical characteristic of a sample. For example, a researcher might calculate the average salary of a sample of lawyers or the correlation between two variables (e.g., years of field experience and current salary) from sample data; these are both examples of statistics. A *parameter* is a numerical characteristic of a total population. For example, a parameter could be an average (e.g., average salary of *all* lawyers in some population), a correlation, a slope coefficient, or a percentage based on the complete set of elements rather than a sample. Sample statistics are used to estimate population parameters. Statistics calculated from samples will rarely be exactly the same as the population parameters because of random variation, but they are usually pretty close (assuming that random selection is used and one has an adequate size sample). The difference between a statistic and a parameter is called *sampling error*. It is called *sampling* error because it varies, randomly, from sample to sample when random sampling is used. It is important to note that some error is always present when using sampling methods, but the use of random sampling methods ensures that sampling error is random rather than systematic. The values of sample statistics calculated from random samples vary randomly around the true parameter. When we use an unbiased estimator (e.g., the sample mean is an unbiased estimator of the population mean) we know that random samples will vary randomly around the true population parameter and that the average of all possible sample statistics will be equal to the true population parameter. When the sampling procedure consistently produces samples that are not representative of the population from which they are drawn, the process is biased and is likely to produce a *biased sample*. With biased samples, sampling error is not random and the sample statistics are systematically different than the population parameters. Random sampling methods must be used if a researcher wants assurance that his or her sample is a non-biased or representative sample.

A *sampling distribution* shows the variability of a statistic over repeated sampling from a particular population. For example, if you repeatedly chose random samples of 500 people from a population of 3,000 people each sample would probably have a slightly different calculated statistic (e.g., average height) but most of the sample statistics would be near the population parameter and not consistently higher or lower. If you plotted all of the averages from the repeatedly drawn samples, you would have a sampling distribution. Sampling distributions are important because they show us how a sample statistic varies over repeated sampling and allows us to make probability statements. In many situations in the social sciences, the *central limit theorem* will apply. This useful theorem tells us that regardless of the shape of the population, the sampling distribution of many statistics (e.g., means, sums, and linear combinations of these such as difference between two means, difference between two proportions, regression coefficients) will be normally distributed around the true population parameter (with a known standard error) if the sample size is sufficiently large.

Sampling typically begins with a researcher locating or constructing a sampling frame. A *sampling frame* is a list of every element in the population. For example, if we are interested in drawing a sample of freshmen students at Michigan State University then our sampling frame would be a list of names of every freshman enrolled at Michigan State University. After the sample is drawn from the sampling frame the researcher contacts the potential sample members and asks them to participate.

It is common for people to refuse to participate in a research study. The proportion of people that do participate is determined by calculating the response rate. The *response rate* is the percentage of people in a sample that actually participate in the research study. The calculated response rate will rarely be 100 %. If you selected, for example, a sample of 400 individuals and only 366 participated, the response rate would be 91.5 %. The formula for the response rate is

$$\text{Response rate} = \frac{\text{Numer of people in the sample who participate in the research}}{\text{Total number of people in the sample}} \times 100$$

Logically, you want your response rate to be as high as possible in order to have a sample that is representative of the target population, and to reduce the sampling error. A 70 % or higher response rate is usually considered appropriate. However, high response rates do not necessarily guarantee that the sample will be representative of the population. That is, a sample with a high response rate might still be biased (i.e., not representative of the population) because the people who choose not to participate might be systematically different from the people who do choose to participate in the research study (see Bautista, Chap. 4 on the Total Survey Error). This is especially problematic if the non-responders are different in a way that is related to the outcomes of the study. Researchers should discuss the issues of sample selection procedures, response rates, and sample integrity when writing up reports. Generally, you should not trust research reports that exclude this information.

5.2 Probability or Random Sampling Methods

5.2.1 Simple Random Sampling

Simple random sampling is a natural starting point in a discussion of sampling because it is the simplest form of random sampling and it serves as the foundation for many other random sampling methods. Every random sampling technique uses simple random sampling at some point during the sampling process. *Simple random sampling* is a technique with the property that every element within a population has an equal chance of being included in the sample. The sample is oftentimes called a *simple random sample*. When each possible sample of a given size or every element in a population has an equal chance of being selected for the sample, the sampling method is broadly known as an *equal probability of selection method* (EPSEM)(Groves et al. 2009; Kish 1965; Peters and Eachus 1995).

An easy way to think about drawing a simple random sample is to think about the good old "hat method." The first step is to go to good hat store and buy a big, deep hat. Let us say you want to select a sample size of 100 from a population of 1,000. First, you would make one slip of paper for each individual in the population and put them into the hat. It is important that each slip of paper is identical to every other slip (e.g., size, shape, number of folds, weight). In order to make sure that each piece of paper has an equal chance of being chosen, you need to mix them all up. Reach in and choose one slip of paper (i.e., the first sample member). Then shake up the slips in the hat and draw another slip, and repeat this process until you have selected 100 individuals. This process will provide a simple random sample of 100 individuals ($n = 100$) from a population size of 1,000 ($N = 1{,}000$). In survey research, it is preferable to use sampling *without replacement* (i.e., do not put a selected slip back into the hat) because sampling without replacement is more efficient than sampling with replacement and you do not want to include the same person in your survey more than once (Groves et al. 2009). Last, identify the names of the selected sample members and contact them to participate in your study.

5.2.1.1 How to draw a simple random sample

The hat method shows how simple random sampling works; however, survey researchers do not use this method in practice. A similar approach is to use a *table of random numbers* to select a sample from a population, but this approach is rarely used. A table of random numbers is a list of numbers that fall in a random order (see Table 5.1). Because the numbers fall in a random order, a researcher can be confident that, in the long run, all digits appear with the

Table 5.1 Table of random numbers

Line/column	1	2	3	4	5	6	7	8	9	10
1	10480	15011	01536	53988	81647	02011	69179	14194	62590	36207
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12565
10	85475	36857	53342	91646	53060	59533	38867	62300	08158	17983

same frequency, and that there will be no systematic pattern of numbers in the table.

In practice, researchers often use random number generators when selecting random samples (i.e., if they do not just purchase a sample from a vendor). Here are links to three random number programs freely available on the internet:
http://www.random.org
http://www.randomizer.org
http://stattrek.com/tables/random.aspx

Random number programs randomly select a set of numbers for the desired sample size. Researchers must assign numbers to each of the elements in the sampling frame to associate each element with a number. Table 5.2 shows an example of a sampling frame where a list of people in a small population (i.e., a sampling frame) is presented with their individual identification numbers. We included age and gender data in the table so that we can calculate some simple statistics and parameters; these calculations will allow us to determine how similar our sample is to the population on these characteristics. Typically a sampling frame only contains names and identification numbers; additional information data are collected by the researcher later only for the sampled members.

Let us draw a simple random sample of ten individuals from the population of 60 shown in Table 5.2. If you wanted to use the table of random numbers then you would start anywhere in the table and move in any direction (as long as you continue in that direction), recording the numbers to use for selecting sample members. We will use a random number generator. To randomly select ten numbers from 1 to 60, we specifically went to the randomizer.org website and clicked "Use the randomizer form to instantly generate random numbers." Next, we filled in some specific information:

1. How many sets of numbers do you want to generate?
   - We entered 1 because we want one set of ten numbers.
2. How many numbers per set?
   - We entered 10 because we want ten numbers in our sample.
3. Number range?
   - We inserted 1 and 60 to indicate the range of numbers in our sampling frame.
4. Do you wish each number in a set to remain unique?
   - We clicked "yes" to indicate that we wanted sampling *without replacement* (because this is a more efficient sampling method that sampling with replacement).
5. Do you wish to sort the numbers that are generated?
   - We clicked "yes" for convenience.
6. How do you wish to view your random numbers?
   - We left the program at its default value ("place markers off") because we were not interested in knowing the order in which the numbers happened to be selected.

Table 5.2 Sampling frame with information on gender and age included

Number	Name	Age	Number	Name	Age	Number	Name	Age
01	Matthew Agee (M)	64	21* **	Nathan Hardy (M)	21	41	Allen Sanders (M)	63
02*	Fred Baker (M)	18	22	Daniel Hinton (M)	60	42	Bry Stuart (M)	33
03**	Jimmy Battista (M)	57	23	Elisabeth James (F)	43	43	Royce Simpson (M)	31
04	Maureen Blackburn (F)	30	24*	Gurupreet Khalsa (F)	18	44	Daniel Surry (M)	20
05	Kristin Brock (F)	38	25*	Adam Knight (M)	52	45**	Max Sakellaris (M)	47
06*	Walt Bryson (M)	31	26	John Knott (M)	35	46	Jessica Schober (F)	23
07	Clint Burdette (M)	52	27* **	Elizabeth Kolbusz (F)	35	47	Nora Shumway (F)	29
08*	Steve Cashore (M)	42	28	Chad McFarland (M)	20	48	Joseph Toomey (M)	42
09**	Laura Chambers (F)	21	29	Genevieve McGarty (F)	64	49	Faith Tharp (F)	23
10	Aisling Conroy (F)	21	30	Caitlyn Morroco (F)	19	50	Lisa Turner (F)	57
11	Alina Cordero (F)	55	31	Jonathan Morgan (M)	33	51**	Chuck Taylor (M)	43
12	Dean Coumanis (M)	44	32	Chelsea Mulvey (F)	33	52	Jim Van Haneghan (M)	20
13	Jane Davis (F)	50	33**	Brendan Murphy (M)	40	53*	Trent Walters (M)	26
14*	Kelsey Farkvam (F)	65	34	Ross Newell (M)	57	54*	Hunter Whidden (M)	29
15**	Joshua Foster (M)	27	35	Claire Orphan (F)	28	55	Michael Whittlef (M)	46
16	Mel Francis (M)	37	36	Susan Parise (F)	38	56	Paul Darron Wood (M)	31
17	Kathy Garrigan (F)	37	37	Elliott Quina (M)	30	57**	Patrick Waldron (M)	39
18	Ross Givens (M)	46	38	Michael Reed (M)	64	58	Dawn Yanker (F)	51
19	Michael Gordon (M)	30	39**	Mark Roberts (M)	32	59	Andrea Yohn (F)	42
20	Alexander Graf (M)	63	40	Sylvia Rogers (F)	46	60	Candice York (F)	46

Data on variables such as age and gender are usually not included in a sampling frame. Data are obtained after they are collected from the sample respondents. To do a couple of calculations, data on age and gender are provided in the columns and parentheses

* Elements selected in simple random sampling example discussed in the text

** Elements selected in systematic sampling example discussed in the text

7. Last, to obtain our set of random numbers, we clicked "Randomize Now!"

Our resulting set of random numbers was 2, 6, 8, 14, 21, 24, 25, 27, 53, and 54. As you can see in Table 5.2, the sample is composed of Fred Baker (element 2), Steve Cashore (element 8), Kelsey Farkvam, (element 14), Nathan Hardy (element 21), Gurupreet Khalsa (element 24), Adam Knight (element 25), Trent Walters (element 53), Walt Bryson (element 6), Elizabeth Kolbusz (element 27), and Hunter Whidden (element 54). This is your sample. The individuals and their numbers are marked with a single asterisk.

Simple random samples from relatively small populations can also be obtained using a spreadsheet, such as EXCEL. First, assign a consecutive case number to each of the subjects in the population, 1 through N, where N is the population size. Next, use the random number formula (in EXCEL this formula is called RAND) to assign a random number between 0 and 1 to each case. Copy and paste the values of the random numbers to fix their values. Finally, sort the cases based on the value of the random number in either ascending or descending order. Save the file and select the first n cases to obtain your sample.

After contacting participants, you would collect empirical data from the sampled individuals, and then you would conduct the statistical analysis. Remember, however, that we included data on gender and age in our example. Here is a straightforward calculation. Using age data, you can calculate the average age for the individuals in the sample. Just add the ages for the ten people and divide that number by 10 $(18 + 31 + 42 + 65 + 21 + 18 + 52 + 35 + 26 + 29)/10 = 33.7)$. The average age of the individuals in our sample is 33.7 (i.e., $\bar{X} = 33.7$), and this is our point estimate of average population age (μ).

In our particular case, the population average age is *known* to be 38.95 (i.e., $\mu = 38.95$), or about 39 years old. This parameter was calculated by adding up the ages of all the people in the population and dividing the sum by 60. Our sample statistic (33.7) was off by approximately 5 years. This might seem like a large sampling error, but remember, that our sample was quite small ($n = 10$) and larger sample sizes yield more accurate estimates of population parameters. To demonstrate this we used the random number generator and selected a larger sample of 25 individuals (elements selected: 1, 3, 7, 10, 12, 13, 14, 20, 22, 28, 31, 32, 39, 42, 43, 44, 49, 51, 52, 53, 54, 55, 56, 57, 60). The average age for the simple random sample of size 25 was 39.24, or about 39 years, which is closer to the population parameter. Sometimes a particular sample of size 10 might be closer to the population parameter than a sample of size 25 because of chance variability, but in the long run, the larger the sample the better. If you were to draw an infinite number of samples of size n from a population, determine the mean for each sample, and plot all of the empirically derived sample means, the distribution would follow a normal distribution; this distribution is called the sampling distribution of the mean. The mean of this sampling distribution would be equal to the population mean (μ), and it would be normally distributed. Because of the normal distribution, we know that the vast majority of the time the sample means will be relatively close to the population mean and more extreme values will be found less frequently.

5.2.2 Systematic Sampling

Another sampling method that includes a random process is systematic sampling. This method is used with lists (where the list is an appropriate sampling frame), the list can be either ordered or unordered. With *systematic sampling* one determines the size of the sampling interval (k), selects a random starting point between 1 and k, and then selects every kth element for inclusion in the sample. Like simple random sampling, systematic sampling is an EPSEM, but with one difference. Although each element in the population has an equal chance of inclusion in the sample, the probabilities of different combinations of sample elements being included are not equal. In order to obtain standard errors for estimators used in statistical

analysis based on systematic sampling, some assumptions must be made. If you can assume that the list is randomly ordered, standard errors assuming simple random sampling can be used. If the list is stratified (discussed below), standard errors based on stratified sampling should be used.

Here is how systematic sampling works. First, determine the sampling interval (symbolized by k) by dividing the population size by the desired sample size (i.e., N/n). Second, using a random process, select a starting point (i.e., a number between 1 and k, inclusive). Third, start with the randomly selected starting-point-element, and then select every kth element from the list. The set of selected elements constitutes the sample. This sampling method is frequently used when the sampling frame is a list (e.g., a list of names, a list of schools, a list of products, etc.).

Here is a hypothetical example for demonstration. Suppose that there are 50 employees in advertising company XYZ for which you are conducting some consulting work on group dynamics and teamwork. You determine that you want to select five employees to lead small-group projects. You have a list of the names of all 50 employees and you have assigned a number to each person on the list. Now follow the three steps. First, determine the sampling interval, k. To determine k, divide the population size ($N = 50$) by your desired sample size ($n = 5$). Dividing 50 by 5 you will see that your sampling interval is 10 ($k = 10$). Second, obtain the starting point. To do this, use one of the random number generators mentioned earlier and randomly select a number between 1 and k, that is between 1 and 10 (including 1 and 10). Perhaps you randomly selected the number 6. The first member of your sample is the person numbered 6 in your list. Third, select the rest of your systematic sample. Starting with person 6, go down the list, selecting every kth person. The second member of your sample is person 16 ($6 + 10 = 16$), the third is person 26 ($16 + 10 = 26$), the fourth is person 36 ($26 + 10 = 36$), and the fifth and final member is person 46 ($36 + 10 = 46$). Your systematic random sample is composed of persons 6, 16, 26, 36, and 46. That is five people.

Now we will select a systematic sample of ten people from the sampling frame shown in Table 5.2. Remember that earlier in this chapter we selected a simple random sample ($n = 10$) from that sampling frame, we marked the chosen individuals with single asterisks, and we found that the average age for the simple random sample was 33.7. We also determined that the average age in the population of 60 people ($N = 60$) was 38.95.

Because the sample size is 10 and the population size is 60, the sampling interval, k, is $60/10 = 6$. Using one of the random number generators, we determine that our randomly selected starting point is three. Three is our starting point and the number of the first member included in the sample. Next, starting at three, we select every 6th person until reaching the desired sample size (this occurs near the end of the list). The sample is composed of persons 3, 9, 15, 21, 27, 33, 39, 45, 51, and 57. Specifically, it is composed of Jimmy Battista, Laura Chambers, Joshua Foster, Nathan Hardy, Elizabeth Kolbusz, Brendan Murphy, Mark Roberts, Max Sakellaris, Chuck Taylor, and Patrick Waldron. We marked these 10 people in Table 5.2 with double asterisks. The average age in this sample is ($57 + 21 + 27 + 21 + 35 + 40 + 32 + 47 + 43 + 39$) divided by 10 which is equal to 36.2 years.

In this case, the average age in our systematic sample (36.2) was a better estimate of the population average (i.e., 38.95) than the average age found in the simple random sample (33.7). It might not always work out that way though; sometimes simple random sampling will yield better population estimates and sometimes systematic sampling will provide a better estimate of the population parameter. In general, if you are sampling from a sampling frame that is ordered randomly, the results from simple random sampling and systematic sampling will be very similar (Groves et al. 2009; Tryfos 1996).

The list (sampling frame) can be organized in different ways. If you order or *stratify* the list

according to levels of a categorical variable or according to the values of a quantitative variable (producing strata that are internally homogeneous and reducing the standard error) then the sampling will be more efficient than basic systematic sampling (Scheaffer et al. 1996). Groves et al. (2009) and Kalton (1983) call this process (of ordering a list and taking a systematic sample) *implicit stratification*; one might also choose to call it stratified-systematic sampling. For example, if you have a list of names organized in ascending or descending order of intelligence scores or a list organized by categories such as level in an organization or gender, then you are probably better off using systematic sampling procedures than simple random sampling. When we say "better," we mean that, in the long run, systematic sampling with this type of sampling frame will tend to result in samples that are more representative of the target population than simple random sampling methods. Systematic sampling produces representative samples in general because it is an equal probability of selection method; that is, each individual in the population has an equal chance of being included in the sample (Kalton 1983).

There is one potential problem, called periodicity, associated with systematic sampling. *Periodicity* is a cyclical pattern in a sampling frame that can cause systematic sampling to produce a nonrepresentative sample. Consider, for example, the hypothetical sampling frame presented in Table 5.3. The sampling frame consists of officers (i.e., presidents and vice presidents) from ten special interest groups (SIGs) in a national professional association. Now, let us say that we wanted to draw a systematic sample of five SIG officers. First, we would calculate our sampling interval. The population size is 20 SIG officers and we want a sample of 5. Therefore, the sampling interval k is $20/5 = 4$. We will select every fourth SIG officer to be included in the sample. Next, we would need to use a random number generator to pick the starting point. For demonstration purposes, let us suppose that the random number generator provided 2 as the starting point. We start with the second element and select every fourth element until we have the sample of 5 SIG officers. The sample is composed of elements 2, 6, 10, 14, and 18.

Table 5.3 A periodic or cyclical sampling frame

Element 1	President 1
Element 2	Vice President 1
Element 3	President 2
Element 4	Vice President 2
Element 5	President 3
Element 6	Vice President 3
Element 7	President 4
Element 8	Vice President 4
Element 9	President 5
Element 10	Vice President 5
Element 11	President 6
Element 12	Vice President 6
Element 13	President 7
Element 14	Vice President 7
Element 15	President 8
Element 16	Vice President 8
Element 17	President 9
Element 18	Vice President 9
Element 19	President 10
Element 20	Vice President 10

However, there is a problem with the sample just selected. Looking at Table 5.3, notice that all of the individuals selected for inclusion in the sample are vice presidents! The method produced a *biased* sample that does not include a single president. The sample is not representative of the target population. The organization of the sampling frame was cyclical, or "periodic," because each vice president was always directly preceded and followed by a president. The lesson here is that you must always examine your sampling frame carefully for patterns and periodicity. If you cannot remove the cyclical pattern, you should *not* use systematic sampling. One solution in this case would have been to randomize the elements in the sampling frame. A superior solution in this case and in general is to stratify the sampling frame; in this case the

problem would be solved by listing all of the presidents and then listing all of the vice presidents.

5.2.3 Stratified Random Sampling

Stratified sampling is a sampling method in which a population is divided into mutually exclusive groups (called strata), and then simple random samples or systematic samples are selected from each of the groups (each stratum). For example, you could divide a population of high school students into freshmen, sophomores, juniors, and seniors and then take random samples from each group. We call the variable that we order the population on the *stratification variable*. In the example just mentioned, the stratification variable would be high school year. Stratification variables can be categorical (e.g., ethnicity, religion, gender) or quantitative (e.g., family income, age, intelligence, time) and researchers can organize a sampling frame with one or, more frequently, multiple stratification variables.

5.2.3.1 Proportional stratified sampling

Proportional stratified sampling (also called *proportionate* stratification and stratified *sampling with proportional allocation*) is one of the most commonly applied forms of stratified sampling. In proportional stratified sampling, the researcher creates a stratified sampling frame, determines the sample sizes for the strata such that the sample sizes are proportional to the population strata sizes, and then selects a random sample of the appropriate size from each stratum. As a result, the sample strata will be proportional to their sizes in the population and random in every other respect. If gender proportions in the population are 0.45 male and 0.55 female, those proportions will be forced upon the sample. The stratification of the sample will mirror the population stratification on the stratification variables used, and it will mirror a simple random sample in every other way (if random sampling is used to obtain the elements from each stratum). Take for example our grade level strata from above. (We will use percentages instead of proportions this time.) If grade is the stratification variable then the proportions of freshmen, sophomores, juniors, and seniors in our sample will be the same as the proportions of freshmen, sophomores, juniors, and seniors in the actual high school population. Therefore, if the target high school population consists of 30 % freshmen, 28 % sophomores, 20 % juniors, and 22 % seniors, then 30 % of the students in the sample will be randomly selected from the freshman stratum, 28 % of the students in the sample will be randomly selected from the sophomore stratum, and so forth so that the chosen sample is proportional to the population. This is why we call it "proportional" stratified sampling. The proportions in the sample are made to be the same as the proportions in the total population on certain characteristics.

When using proportional stratified sampling you must know the population proportions on the stratification variables. However, a key benefit of using proportional stratified sampling is it requires fewer individuals than simple random sampling (Kalton 1983, p. 21). That is, proportional stratified sampling tends to be a little more efficient than simple random sampling because when you draw a proportional stratified sample the proportions in the sample on the stratification variable will be perfectly, or close to perfectly, representative of the proportions of the same variable in the population and otherwise the sample is a random sample. It is desirable in stratified sampling to maximize the homogeneity of elements within each stratum; this is in contrast to the next sampling method called cluster sampling where it is desirable to have heterogeneity within each cluster.

You can use many different and multiple stratification variables to stratify your sample such as gender, political affiliation, religion, income level, and education. If you choose to select your sample using stratified random sampling you can be confident that your sample will also be representative of the population on *all other variables* that are not used as stratification variables because you are still using simple random sampling within each population stratum. It is important to note that *proportional stratified sampling is an equal probability of*

selection method (EPSM), which means that every individual in the population has an equal chance of being included in the sample. As a result, proportional stratified sampling produces samples that are representative of the population from which they are drawn.

5.2.3.2 Disproportional stratified sampling

Thus far, we have focused our discussion on proportional stratified sampling. There are, however, situations when a researcher might need to select a disproportional (or disproportionate) stratified sample (also called disproportionate stratification and *stratified sampling with disproportionate allocation*). A *disproportional stratified sample* is a type of stratified sampling in which the sample proportions are made to be different from the population proportions on the stratification variable. For example, a researcher might be interested in selecting a sample of inmates from a state prison where the majority of the inmate population is male. The researcher, however, wants good sample sizes for both males and females and decides to select a sample that is half male and half female. In this case, she would need to oversample the female population and undersample the male inmate population; her final sample would be disproportionate to the original population, and her method would *not* be an EPSEM. This type of sampling is sometimes used when the interest is to compare groups rather than making statistical generalizations from the sample to the population. Generalizing to the total population is possible with disproportional stratified sampling, but you will have to use weighting procedures in the analysis (discussed below). Another situation that might require disproportionate sampling is when you have very small groups in a population. If this is the case, you might need to oversample the small groups to make sure that you have sufficient sample sizes.

An important point to understand here is that when you use disproportional stratified sampling, you cannot make generalizations directly from your sample to the population unless you use weighting procedures because the sample strata do not accurately reflect the proportions in the population. Weighting is a statistical procedure used during data analysis that gives greater weight to larger strata and less weight to smaller strata to reflect the true proportions in the population. You can, however, make comparisons between groups without worrying about weighting.

5.2.4 Cluster Sampling

A third major type of random sampling is called *cluster sampling*. This type of sampling involves the random selection of *clusters* of elements (e.g., cities, businesses, churches, schools) from a sampling frame of clusters rather than selection of individual elements (e.g., mayors, workers, churchgoers, students). A *cluster* is a collective unit that includes multiple elements. For example, a neighborhood is a cluster because it is composed of many families or, more specifically, individuals. A family could also be considered a cluster because it is composed of multiple members.

Up until this point, the sampling methods have involved sampling single units rather than collective units. As a general rule, cluster sampling always involves randomly selecting clusters at some point, rather than single-unit elements. In cluster sampling, for example, a researcher might randomly select schools from a school district or region. The simplest form of cluster sampling is exactly like simple random sampling with one major difference—clusters are randomly selected instead of individual units/elements.

Cluster sampling requires a larger sample size than simple random sampling and stratified sampling and it is less accurate for a given sample size and thus may produce a larger sampling error. This weakness can be overcome by increasing the sample size and the number of clusters. Still, there are many situations in which cluster sampling is the ideal method. For example, cluster sampling is usually preferred when the target population is geographically diverse and the researcher needs to conduct in-person interviews. In this situation, cluster

sampling can reduce travel costs, interviewer costs, and decrease the amount of time it takes to physically interview the entire sample. For example, suppose your population was geographically spread out such as all individuals in the United States. Cluster sampling would be far more efficient than physically driving to every individual's house selected from simple random sampling!

There are also situations where, as a researcher, you do not have access to a sampling frame of all the elements in a target population. Cluster sampling can be useful in this situation because you may be able to identify naturally occurring groups of sampling elements such as mental health clinics, classrooms, schools, parishes, and voting districts. Lists of these clusters might be available, more easily accessible, or easier to develop than lists of every individual in the population. Once you have selected a random sample of clusters from a population then you only need to locate or develop sampling frames for those particular clusters. There is no need to construct a sampling frame for the entire population.

Now we will show two kinds of basic cluster sampling. The simplest case of cluster sampling is one-stage cluster sampling. *One-stage cluster sampling* involves randomly selecting a set of clusters from the population and then including all of the individuals from the chosen clusters in the sample. For example, you might take a random sample of 100 hospitals from all of the hospitals in Texas. You can use one of the *random* sampling methods previously described in this chapter (simple random sampling, systematic sampling, or stratified sampling) to randomly select the sample of clusters (e.g., hospitals) from the population of all clusters. After randomly selecting the clusters from the population, all the individual elements (e.g., doctors, patients, ambulances) in those clusters are included in the sample. Therefore, in one-stage cluster sampling, *sampling* is only conducted at one stage (i.e., the cluster level).

In *two-stage cluster sampling*, sampling occurs at two stages rather than simply at the cluster level. In stage one, a set of clusters is randomly selected from all of the clusters. In stage two, a random sample of elements is drawn from each of the clusters selected in stage one. For example, 50 schools (clusters) could be randomly sampled from a list of schools. If all teachers in the 50 schools were included as in a one-stage cluster sample, the sample size might be 1,000 (50 schools $\times$ 20 teachers per school $=$ 1,000 teachers). Just as before, however, we want to select a sample size of 500. Therefore, at stage two, ten teachers could be randomly selected from each of the 50 schools using a random sampling method. The outcome would be a two-stage cluster random sample of 500 teachers. In this example we assumed that all of the schools in the population consisted of approximately 20 teachers per school. If clusters are of equal size then the simple cluster methods just described are EPSEM.

However, the size of the clusters are not always equal or approximately equal. Luckily, we have procedures to help us select clusters when cluster sizes are not equal called *probability proportional to size cluster sampling* (PPS). With PPS, each cluster's chance of being included in the sample is proportional to its size (i.e., larger clusters are given a greater chance of being selected than smaller clusters). For example, if a school district with 8 schools was comprised of four large and four small schools, where the large schools have 40 teachers and the small schools have 20 teachers, the large schools will be twice as likely to be chosen as the small schools. In practice this is accomplished by listing the large schools twice in the school sampling frame. This would mean that a large school would be twice as likely to be chosen as a small school. In the second stage, if ten teachers are chosen from each of the schools sampled in the first stage, each teacher in the district will have an equal probability of being selected into the sample.

In large national multi-stage cluster samples it is sometimes necessary to ensure that some sampling clusters will always be included in the final sample. For example, this type of cluster sampling is used in the Current Population Survey (CPS). The CPS is conducted every

month by the U.S. Census Bureau for the Department of Labor. The data collected in the CPS are used to calculate the monthly unemployment rate for the United States. Certain sample clusters, like the one that includes New York City, are always included in the CPS. Here is how cluster sampling was used in the February 2005 CPS (U.S. Census Bureau 2005):

1. The United States was divided into geographic areas or clusters called Primary Sampling Units (PSUs).
2. The PSUs in each state were grouped into strata. Some of the large PSUs were assigned their own group. All of these large PSUS were selected into the sample.
3. One PSU was then selected from the remaining strata. The probability of selection assigned to these PSUs was proportional to its size. Like the clusters in the school example above, the larger size of the PSU, the greater the probability it will be selected into the sample. A total of 741 PSUs were selected in this, the first stage of the cluster sample. In the second stage, around 70,000 housing units are selected from the sampled PSUs. About 60,000 of the housing units were occupied. About 144,000 people living in these housing units were included in the CPS sample.

In large national multi-stage clusters samples like the CPS, some groups are over represented. Consequently the sample must be weighted before it can be used to describe the characteristics of the U.S. population.

5.3 Using Sample Weights

Disproportionate stratified samples and some multi-stage cluster samples must be weighted if you want to use the samples to make an inference to the population from which the samples were obtained. Because some strata or groups are overrepresented in the sample while other strata or groups are underrepresented, it is necessary to weight the cases in the sample so that the group proportions in the sample will be the same as they are in the population. The sample cases are weighted by multiplying each case by its respective sample weight. The sample weights represent the multiplicative inverse of the probability of that case being selected into the sample. Let us illustrate with an example.

Consider a hypothetical town of 20,000 people where the population consists of five different race-ethnicity groups. The frequency and relative frequency of each group is shown in Table 5.4. In order to ensure an adequate number of people from each group in the analysis, a random sample of 200 is obtained from each one. Note that each group is now equally represented in the sample, which is not true of the population from which the sample was taken. However, weighting the cases in the sample will make the sample proportions the same as the population proportions. In this case the sample weights are easy to compute. First, calculate the probability of selection for each group. For example there are 3000 Hispanics in the town population and 200 in the sample. Thus the probability that any Hispanic in the town will be selected into the sample is $(200/3{,}000 = 0.067)$. The multiplicative inverse of this probability is $(3{,}000/200 = 15)$. Note that the product of any number multiplied by its multiplicative inverse is equal to 1. If we multiply 200 by 15 the product is 3,000, which is the original number of Hispanics in the population. Multiplying each person in the sample by their respective weight will reproduce the original population proportions.

Many national data sets that are available for public use are based on stratified random sampling where certain groups are oversampled. This is frequently done to ensure the samples include sufficient numbers of minorities that would be underrepresented in a simple or strategic random sample. These data usually come with sample weights that are calculated by the researchers collecting the data. Although the weights in these samples are based on the logic we have just described, their calculation is usually more complex and takes into account a number of factors such as race, age, sex, and the rate of nonresponse to the survey. Most of the standard statistical packages such as SPSS, SAS, and STATA have procedures for weighting

Table 5.4 How to compute sample weights for a disproportion stratified sample

	Population Frequency	Relative Frequency	Sample Proportion	Sample Frequency	Selection Probability	Sample Weight	Original Frequencies
White	12,000	0.60	0.20	200	0.017	60	12,000
Black	3,000	0.15	0.20	200	0.067	15	3,000
Hispanic	3,000	0.15	0.20	200	0.067	15	3,000
Asians	1,000	0.05	0.20	200	0.200	5	1,000
Mixed race	1,000	0.05	0.20	200	0.200	5	1,000
Population N	20,000						
Sample n	1,000						

sample data. Note that some programs, such as SPSS, weight cases to reproduce the population size, other programs such as SAS weight cases to the sample size. With both SPSS and SAS the weighted sample characteristics will reflect the characteristics of the population.

Many researchers will also combine stratified random sampling with multi-stage cluster sampling. For example, the Current Population Survey (CPS) described above is based on a national multi-stage cluster samples that also oversamples blacks. This ensures there are sufficient numbers of black respondents to produce a reliable estimate of the black unemployment rate. In this case the purpose of the oversample is to ensure there are enough Hispanic respondents in the sample to obtain reliable estimates of the health-related characteristics of this population. The weights which accompany these data also correct for the effects of cluster sampling. For example, some geographical areas may be assigned a different probability of selection in order to ensure an adequate representation of targeted minorities or the inclusion of large metropolitan areas (Singleton and Straits 2010).

5.4 Statistical Analyses with Weighted Data

The use of sample weights in computing descriptive statistics is fairly common. Whenever weights are used to calculate means, standard deviations, or relative frequencies (i.e. proportions) researchers should report the unweighted rather than the weighted sample size. This will permit others to assess the adequacy of the sample size. When computing inferential statistics with weighted data, the use of caution is in order. This is because weighting can artificially reduce the variance in certain types of variables. Remember, when you weight cases you are essentially creating multiple identical copies of the same case. This will result in an underestimate of the standard error and inflated significance tests. When this happens you increase your risk of concluding that significant differences exist where there are none (a Type I error). Therefore, when conducting statistical analysis, you should use specialized software, such as SUDDAN or STATA, that use sampling information to correct the standard errors.

Many researchers forgo the use of weights altogether when conducting statistical tests. Instead they identify each of the variables that are used to compute the sample weights. These variables are then included as covariates in multivariate models along with other variables that are theoretically important for the analysis. For example, if sample weights are computed using information regarding respondents sex, race, and age, these three variables would be included as control variables in any statistical procedures. Researchers may also use these

Table 5.5 Survey research sites providing useful sampling information and links

Address	Name
www.isr.umich.edu/src	Survey Research Center at University of Michigan's Institute for Social Research
www.norc.uchicago.edu/homepage.htm	The National Opinion Research Center at the University of Chicago
www.wws.princeton.edu/~psrc	Survey Research Center at Princeton University
http://statlab.stat.yale.edu	Social Science Statistical Lab at Yale
www.indiana.edu/~csr	Indiana University Center for Survey Research
http://www.irss.unc.edu/odum/jsp/home.jsp	The Odum Institute for Research in Social Science at the University of North Carolina
http://filebox.vt.edu/centers/survey/index.html	Virginia Tech Center for Survey Research
http://members.bellatlantic.net/~abelson	These two sites provide many useful survey research links. www.ku.edu/cwis/units/coms2/po/
www.gallup.com	The Gallup Organization
www.surveysampling.com/ssi_home.html	Survey Sampling Inc
www.aapor.org	The American Association for Public Opinion Research
www.ncpp.org	The National Council on Public Polls

variables to stratify the analysis, especially where there are global interaction effects. For example, if the effects of most model variables differ by sex, the researcher should probably estimate separate models for men and women. Before moving to other sampling methods, note that you can find many examples and useful information on the Internet by visiting survey research organizations. For your convenience, we include a table with web links to some organizations that routinely use random sampling methods (Table 5.5).

5.5 Nonrandom Sampling Methods

Although nonrandom sampling methods are rarely appropriate in survey research, we will briefly cover them here to provide a relatively comprehensive review of sampling methods used in research (Christensen et al. 2011).

5.5.1 Convenience Sampling

Convenience sampling is used when a researcher chooses his or her sample simply by including people who are available or can conveniently be recruited to participate in the study. An important note about convenience sampling is that you cannot make statistical generalizations from research that relies on convenience sampling. In fact, when convenience sampling is used it is difficult if not impossible to identify what particular population a "sample" comes from. The problem with convenience is that whatever the population might be, one can be certain that not every element has an equal chance of being included in the study. It is *not* an equal probability of selection method (EPSEM).

Convenience sampling is to be avoided *always* in survey research. If you conduct a different type of research (e.g., experimental research) and rely on convenience sampling, it is important that you provide a detailed description of the characteristics of the sample (e.g., demographic characteristics) in your research report. Some researchers attempt to describe a hypothetical population that a convenience sample might reasonably be associated with. In the end, however, it is up to the reader to carefully examine the characteristics of reported convenience sample characteristics and make a decision about whom they believe the sample might represent.

You might be surprised to learn that the majority of experimental researchers do *not* select random samples. Rather, they tend to use convenience samples. For example, published research studies are often conducted with undergraduate students enrolled in introductory psychology classes. Here is an example from published research article that relied on a convenience sample. Notice how the author includes specific details about the characteristics of his sample. The excerpt comes from a study by Foster (2008) about investment, commitment, and personality related to romantic relationships:

> One-hundred forty-four undergraduates (M_{age} = 19.64 years, SD = 2.57; 87 % female; 83 % Caucasian) participated in this study in exchange for research participation credit. All participants were involved in romantic relationships that were at least 3 months old (M_{length} = 1.81 years, SD = 1.29). The vast majority reported their relationships as dating (94 %) and exclusive (96 %; i.e., neither they nor their partners were dating anyone else) (p. 215).

5.5.2 Quota Sampling

Quota sampling is a type of nonrandom sampling that involves the identification of groups within a population, the number of people that should be included in each group to form the "ideal sample," and then the selection of that sample using convenience sampling. In other words, the researcher decides how many people to include in a particular sample, as well as subgroups of that sample, and then uses convenience sampling until he or she "meets the quota" for each group. Sometimes, a researcher might want a sample that is proportional to the subgroups of a natural population. In this case, the sampling procedure might appear similar to proportional stratified sampling; however, there is a major difference. Quota sampling relies on convenience sampling to select the elements in the groups, while proportional stratified sampling utilizes *simple random sampling* techniques to select the cases within the groups/strata. Therefore, generalizations from quota sampling are not as warranted.

5.5.3 Purposive Sampling

Another nonrandom sampling technique frequently used in experimental research is called purposive sampling. An experimental researcher uses *purposive sampling* when he or she knows the characteristics of the target population and then seeks out specific individuals who have those characteristics to include in the sample. For example, perhaps a researcher is interested in conducting research on short-term memory processing of young adults who have suffered closed head brain injuries. Individuals meeting this specific criterion are difficult to find. The researcher might visit local hospitals and rehabilitation centers to locate individuals who have the desired characteristics and ask them to participate. The researcher would continue this process until an adequate number of people have agreed to participate in the study. In short, purposive sampling involves seeking out specific individuals meeting specific criteria to participate in a research study. Here is an example of purposive sampling from a published research article:

> Six young women with learning disabilities from six different rural and nonrural school districts in a Northwestern state composed the sample for this study.…We employed purposive sampling methods to select participants….Participants were nominated by special education teachers and specifically selected to meet certain purposive sampling criteria, including that they (a) received special education services in a public high school, (b) graduated from high school with a standard high school diploma, (c) participated in a school to-work training program for at least 1 year, (d) were eligible for vocational rehabilitation services and participated in the development of an individualized plan for employment, and (e) were employed at least 30 h per week at the time of initial data collection (Lindstrom and Benz 2002, p. 69).

Purposive sampling has the same limitations as all nonrandom sampling techniques. Because purposive sampling is not an EPSEM, generalizations from one specific research study to a population would be ill-advised. An ideal purposive sampling situation would involve identifying individuals with target characteristics and then randomly selecting a sample of those

people. Unfortunately, this is not practical or possible in many situations.

5.5.4 Referral Sampling

Researchers who are conducting research with difficult to sample populations often rely on some form of referral sampling to recruit subjects for their study. Two types of referral sampling methods are network sampling and snowball sampling (Singleton and Strait 2010).

Network Sampling When using network sampling the researcher begins by obtaining a probability sample of some large population that is likely to have some connection to the target population. Members of the initial probability sample are then asked to provide contact information for members of the target population. For example, a researcher who wants to obtain a sample of student musicians might begin with simple random sample of university students. Students selected in the initial probability sample would be asked if they play a musical instrument and if they know other students who also play a musical instrument. The data in the initial screening sample can be used to estimate the proportion of students who are musicians and to obtain contact information for additional cases that will increase the sample size.

Snowball Sampling In snowball sampling, also described as chain referral sampling, each research participant who volunteers to be in the research study is asked to identify one or more additional people who meet certain characteristics and might be willing to participate in the research study. For example, Corsi, Winch, Kwiatkowski, and Booth (2007) used snowball sampling to find 339 injection drug users and 262 non-injection drug users in the Denver Colorado area to participate in structured interviews to identify childhood factors related to intravenous drug use. They used drug use indicators such as drug-related arrests and treatment admissions to identify an initial sample and then relied on participant referrals to locate other potential participants. This snowball sampling method yielded approximately 20–30 new participants per month over a 30-month period.

Sometimes you might only be able to locate a few willing, available, and appropriate participants at the beginning of a study. This is especially true when the nature of the research is sensitive, controversial, or taboo. Over time, however, each new participant might lead you to new potential participants making your total sample size larger and larger. You can picture your sample as a small snowball rolling down a hill, collecting more snow with each turn, becoming larger and larger as it rolls; hence the name snowball sampling. This sampling method can be especially useful when you need to locate members of hard-to-find populations, when no sampling frame is available, or when you are trying to identify hidden populations (Sadler et al. 2010).

5.5.5 Volunteer Sampling

Perhaps the most commonly encountered form of sampling in the media is volunteer sampling (i.e., where the people ask to be included in the sample and are included). This sampling method is widely used in political talk shows where viewers are invited to call or send a text message indicating their support or opposition to a particular position or proposal. It is also found in surveys where researchers post flyers or pamphlets inviting people to participate in a study. The data collected in such samples usually have little or no value for two reasons. First, even if we define population as all those persons who are watching or listening to the broadcast, their number is unknown. Second this type of solicitation is highly vulnerable to sample selection bias (discussed below). If people watch or listen to certain shows based on their own political orientation or preferences, it is hardly surprising that most people who call or send text messages to register their vote tend to support the position of the talk show host.

5.6 Random Selection and Random Assignment

It is very important to understand the difference between *random selection* and *random assignment*. We have focused on random selection

which is just another name for random sampling. *Random selection* involves randomly selecting participants from a population to be included in a sample. To review, random sampling is like pulling a set of names out of a hat. We discussed three major methods of random sampling: systematic sampling, stratified sampling, and cluster sampling. The key idea here is this: *the purpose of random selection is to obtain a representative sample that allows you to estimate the characteristics of the population based on the characteristics of the sample*. This type of generalization is called a *statistical generalization* (Shadish et al. 2002). Making statistical generalizations is a purpose of virtually all survey research.

Random assignment involves putting participants into experimental and control groups in such a way that each individual in each group is assigned entirely by chance. Otherwise put, each subject has an equal probability of being placed in each group. A key point here is that random *assignment* is only used in experimental research, and it enables the researchers to make strong *cause and effect* conclusions from the experiment. When using random assignment, the researcher takes a group (e.g., typically a convenience sample) and randomly divides it into two or more probabilistically equivalent groups for use in an experiment. The purpose of random assignment is to create comparison groups that are the same on "all possible factors" at the beginning of the experiment. When this is the case, a researcher can introduce a treatment and then attribute statistically significant differences between groups found at the posttest to the influence of the independent variable; that is because the independent variable is the only variable on which the groups systematically differ. For example, you might randomly assign participants to either an experimental group that receives a pill or a control group that receives a placebo. Because you used random assignment you can assume that the groups are the same except for the one variable that you manipulated (i.e., the pill) and attribute changes in behavior to that variable. Technically, we know that random assignment works across repeated assignment over the long run; it is possible, but not likely, that a large difference might occur by chance in a single instance of random assignment. Note that randomly selected samples are rarely used in experimental research; thus, statistical generalizations from single experiments such as this are ill-advised. The strategy experimental researchers typically use for generalizing is replication of experimental research findings.

Replication logic is based on repeating experimental studies in different places, with different people, and under different circumstances. When the relationship is reputedly shown in diverse situations, causal attribution of the effect of the independent variable can be generalized. Although rarely done in experimental research, the strongest design would be one in which random selection and random assignment are used. This would allow the researcher to make cause and effect claims and make statistical generalizations about the effect. When interest in causation exists in survey research, multiple nonexperimental techniques are used such as theory/model testing, matching, statistical control, and the use of longitudinal data (to enable the study of change over time).

5.7 Determining the Sample Size When Random Sampling is Used

When conducting a research study, a key decision is deciding the appropriate sample size. The simplest answer is that the bigger the sample the better, but this assumes the sampling method is appropriate and implemented correctly. In inferential statistics, bigger is better because it results in smaller standard errors, greater statistical power or fewer Type II errors in hypothesis testing, and tighter or narrower confidence intervals in estimation. A *Type II error* occurs when a researcher fails to reject a false null hypothesis. (In contrast, a *Type I error* occurs when a researcher rejects a true null hypothesis; the *null hypothesis* typically states that there is no relationship in the population).

Smaller sampling error means that a sample statistic will tend to be closer to the actual population parameter. It is important to remember that if you have access to the entire population and that population is quite small (e.g., 100 elements or less) then it would seem wise not to use sampling but, instead, to include all of the elements in the study; in this case sampling error will equal zero! However, most often researchers are working with large populations and sampling methods are needed.

In this section we are assuming that you are using simple random sampling because as discussed above, some sampling methods are a little more efficient and require smaller sample sizes than simple random sampling (e.g., proportional stratified sampling, systematic sampling with implicit stratification) and some sampling methods require slightly larger sample sizes than simple random sampling (e.g., cluster sampling). We will have more to say about sample size and adjustments to the standard errors later for these more advanced sampling methods.

Here we want to explain that determining the desired sample size (a priori sample size determination) depends on a variety of factors. We will briefly list some of the factors for consideration when thinking about sample size. First, one factor it generally does *not* depend on is the size of the population, unless one is working with very small populations of only a few thousand people (Bartlett et al. 2001; Krejecie and Morgan 1970; Nguyen 2005). A common myth is that sample size is largely dependent upon the size of the population. For researchers who work with larger populations, as is the case in most survey research, the population size is largely irrelevant.

Second, you need larger sample sizes when the population is heterogeneous. Put differently, when a population is relatively homogeneous (i.e., people in the population are very similar) a smaller sample size may be sufficient. Think about it. If every person in the population were exactly alike, you would really only need one person for your sample. Third, the better the measurement, the smaller the sample size is needed. If measurement is too poor, no sample size will be sufficient because without good measurement one is essentially examining little more than random error. Fourth, when you expect a small relationship or small difference between two means or proportions, you will need a larger sample size. Fifth, the required sample size is a function of the type of the random sampling method used. For example, stratified random samples typically require smaller samples than simple random sampling and cluster sampling tends to require larger samples than simple random sampling. Sixth, if you plan on analyzing your data by subgroups, categories, or breakdowns (as in most survey research) you need to have a larger sample size. For example, if you wanted to analyze data from males and females separately, then you would need to have sufficient sample sizes for each group. This is also an important issue when the data will be used to estimate multivariate models. If the data is used to estimate a multiple regression model, you should consider how the number of variables and variable categories will affect the power of the model to detect differences in the dependent variable (Cohen 1988).

Seventh, larger sample sizes are needed when you want to make precise estimates about a population; that is, larger sample sizes produce narrower confidence intervals. In virtually all survey research confidence intervals are determined. *Confidence intervals* are interval estimates of population parameters (e.g., a mean, proportion, regression coefficient) at a specific level of confidence (e.g., 95 % confidence, 99 % confidence). For example, you might hear a news reporter say that 65 % of the people in a city support the Mayor's decision to implement a weekend curfew for individuals under 18, ±4 %, and that the confidence level is 95 %. The 4 % is the margin of error on each side of the point estimate. We like small margins of error because that means we have more precise interval estimation. The 95 % confidence level means that 95 % of the time, over the long run, this process of estimation will capture the true population parameter. For a single confidence we often state that we are 95 % confident that the population value is somewhere within the

Table 5.6 Number of research participants needed for small, medium, and large effect sizes at recommended power of 0.80 for $\alpha = 0.01$ and 0.05

	α					
	0.01			**0.05**		
Test	Small	Medium	Large	Small	Medium	Large
t test for two means[a]	586	95	38	393	64	26
Simple correlation (*r*)[b]	1,163	125	41	783	85	28
Analysis of variance[a]						
2 groups	586	95	38	393	64	26
3 groups	464	76	30	322	52	21
4 groups	388	63	25	274	45	18
5 groups	336	55	22	240	39	16
Multiple regression[b]						
2 predictors	698	97	45	481	67	30
3 predictors	780	108	50	547	76	34
4 predictors	841	118	55	599	84	38
5 predictors	901	126	59	645	91	42

Note: Effect size is the strength of relationship. Analysis of variance is used to compare two or more means for statistical significance. Multiple regression is used to predict or explain variance in a dependent variable using two or more independent variables (labeled "predictors" in table). Information from table was extracted from constructed by R. B. Johnson

[a] The sample size number is for each group. Multiply this number by the number of groups to determine the total sample size needed

[b] The sample size reported is the total sample size needed

interval of 60 and 70 % citywide agreement, but remember that, statistically speaking, the confidence is in the process not the particular interval.

Eight, you will need a larger sample size when you want to increase the power of your statistical test (i.e., when you want to decrease the probability of making a Type II error). Ninth, the required sample size varies quite a bit depending upon the type of statistical analysis you plan to conduct. This idea is shown in Table 5.6. We have set up the table for power of 0.80 which is the recommended minimum for research. Although an alpha level of 0.05 is usually fine, we also included information for an alpha level of 0.01. To use the table, we recommend you assume a small or medium effect size and use an alpha level of 0.05. Perhaps you expect to use a multiple regression with three independent or predictor variables. In this case you will need a sample size of 576 (assuming a small effect/relationship) or 67 (assuming a medium-sized relationship). Perhaps you also are interested in some group differences and expect to conduct an analysis of variance with four groups. In this case you will need a sample size of 1,096 (assuming a small effect/relationship) or 180 (assuming a medium-sized relationship).

An excellent sample size calculator that takes into account factors such as alpha level (i.e., level of significance, expected effect size, and desired power) is GPower. It is readily available on the Internet at no cost. Here is a link to the program: htpp://www.psycho.uni-duesseldorf.de/app/projects/gpower/. (If the link does not work, just search for GPower.)

Last, because your response rate will rarely be equal to 100 %, you will need to include more people than suggested by the sample sizes shown in Table 5.5. You will typically end up with a smaller sample than you initially plan for. If you can estimate your response rate, you can use the following formula to adjust your original sample size estimate. Simply divide the number of people who you want to have in your final sample by the proportion of people who are likely to respond and actually participate. This will give you the number of people to include in your original sample. Here is the formula just described:

$$\frac{\text{Desired sample size}}{\text{Proportion likely to respond}} = \text{Number of people to include in your original sample}$$

5.8 Sampling Issues in Survey Research

Research and data collection methods are not bound to any one type of sampling methodology, though, in practice, some research approaches tend to use some types of sampling more than others (Singleton and Straits 2010). For example, experimental research subjects are frequently recruited through convenience or purposive samples. Psychologists conducting experimental research often enlist the students from their classes to participate in their subject pool. Medical researchers might purposively recruit individuals who have a health condition, such as cancer, that qualifies them for participation in the study. In nonexperimental research, data collection techniques such as focus group or participant observation are used, and these usually rely on purposive or quota samples. University administrations conducting organizational research might randomly or purposively recruit college seniors to provide feedback on obstacles faced by students in completing their academic programs. Researchers using archival data will often use systematic random sampling (Singleton and Strait 2010). Most survey researchers use some form of probability sampling. Small surveys, telephone surveys, and Internet surveys might use simple random sampling or stratified random sampling stratified by region; this includes opinion research and electoral polling at all levels.

Most large-scale national surveys that utilize face-to-face interviews use multi-stage cluster samples combined with stratified random sampling. Examples include the Current Population Survey (CPS), The Public Use Micro Sample of the U.S. Census, (PUMS), the General Social Survey (GSS), the National Health and Nutrition Examination Survey (NHANES), the Panel Study of Income Dynamics (PSID), The National Longitudinal Survey of Youth (NLSY), and the National Educational Longitudinal Study (NELS). These data and supporting materials are available online through the International Consortium for Political and Social Research (ICPSR) at the University of Michigan.

5.9 Sampling Bias in Survey Research

Four key types sampling bias can occur in survey research. *Coverage bias* occurs when some groups are systematically excluded from the sampling frame. For example, people who are homeless or who have no permanent address will be underrepresented in most household surveys. *Non-response bias* can occur when certain groups of respondents fail to respond or are less likely than others to participate in the survey or answer certain survey questions. In the best case scenario, those who are excluded from the study and those who do not respond to the survey items will be similar to those who are included in the survey and to those who respond to the survey questions. But if the missing cases or missing responses are correlated with the independent or dependent variables in the study, the sample will be biased (Little and Rubin 1987).

Both coverage error and non-response tend to be associated with particular types of survey research. Historically speaking, coverage bias was a problem for telephone surveys because many low income families did not have a

telephone. Now most Americans have telephones, but they also have cellular telephones and their numbers are not publicly listed. This problem has been partly remedied by random-digit-dialing (RDD) devices. These devices, when used in conjunction with prefix numbers provided by cellular telephone providers, can reduce the coverage error associated with cell phones (see Vehovar, Slavec and Berzelak Chap. 16). Coverage bias is also a serious problem for email and web-based surveys (see Mesch, Chap. 18, and Manzo and Burke, Chap. 19). The limitation here is obvious; people who do not have computers or access to the Internet cannot participate in an Internet survey. Consequently, low income individuals and minorities are more likely to be excluded from online surveys (Singleton and Strait 2010).

Contacting potential respondents, however, does not insure they will choose to participate in the survey or that they will answer all of the survey questions. The sample response rate depends greatly on the type of survey method used. Though response rates vary from survey to survey, there is some agreement that face-to-face surveys have the highest rate of response, followed by telephone surveys, with mail surveys or self-administered surveys having the lowest response rate (Singleton and Straits 2010; Babbie 2010).

A third type of sampling bias is *sample selection bias*. Unlike non-response bias, which occurs when individuals refuse to participate, sample selection bias results from the behavior of the researcher. This type of bias can occur when the researcher has a tendency to prefer or avoid certain subjects or communities in the sample selection process. Convenience samples and quota samples have a high risk of sample selection bias. For example surveys or surveyors that avoid very wealthy neighborhoods, because they are not perceived as accessible, are likely to produce biased results on most items related to socioeconomic status.

Finally, panel studies are particularly vulnerable to *sample attrition bias* (see Toepoel, Chap. 20 on online survey panels). Because researchers conducting panel studies interview the same individuals more than once at different points in time, there are usually respondents in the initial interview who do not participate in the subsequent interviews. Thus, panels tend to shrink with every successive "wave" or interview. This can occur for a number of reasons. Keeping track of a large number of respondents over several months or years is very difficult; respondents move, get married, get divorced, change their address, and sometimes change their names. People also die (dead people are very difficult to interview!). Some populations are also more prone to sample attrition than others. Homeless people, who are difficult to locate even for an initial interview, are extremely difficult to locate for follow up interviews. Panel studies of youthful offenders may be less likely to locate who have continued their pattern of offending because they are in jail or prison.

These limitations notwithstanding, there are a variety of options that researchers can employ to reduce sample bias. First, providing interview incentives or compensation in the form of a cash payment or gift will increase the likelihood individuals will participate in a lengthy face-to-face interview. If interview incentives, however, are used, they should only be used with probability samples; individuals should not be permitted to self-select into the sample. Researchers should also only use incentives in a manner that is not coercive or exploitive (as discussed previously by Oldendick in Chap. 3 in relation to the ethics of incentives to increase response rate, and later on in this handbook by Toepoel in Chap. 13 on the effects of incentives). Second, response rates can be improved by combining different survey methods (Singleton and Strait 2010). For example when individuals or households do not respond to a mail survey, researchers should follow up with a telephone call or face-to-face visit to the person's address. Third, whenever possible the researcher should make repeated efforts to reach the individuals or households selected in the initial sample. Yielding to the temptation to "move on" to the next person or address because the respondent is hard to reach, increases the risk of non-response bias. Finally, features of probability sampling can also be combined with features non-probability sampling to reduce coverage error and avoid sample

selection bias. Coverage bias in quota sampling might be reduced by recruiting respondents in different settings or geographic locations. Convenience samples that are collected in public places can reduce sample selection bias somewhat by randomly selecting subjects from the crowd. This can be accomplished by interviewing every third, or pulling colored marbles out of pocket. The advent of social media, such as Facebook, has made it possible to locate and maintain contact with friends and acquaintances. This might be used in the future to reduce sample attrition bias in longitudinal research, especially with younger populations.

5.10 Innovative Sampling Strategies

Response rates in surveys using traditional sampling methods are declining (Singleton and Strait 2010). Consequently new and innovative sampling methods are needed to conduct social research. This is especially true for hard-to-reach populations. For example, Heckathorn (1997) has devised a method for reaching hard to sample groups, such as homeless people and drug users. He has shown that repeated iterations of multiple snowball samples will eventually converge on the characteristics that are representative of the population under consideration. Another example of innovative sampling method is used in the National Organization Study, which links a subsample of workers in the General Social Survey to their employers. In this survey worker data are linked to organization data, allowing organizational variables to be combined with individual attributes in predicting worker outcomes (Knoke and Kalleberg 1994).

References

Babbie, E. (2010). *The practice of social research* (12th ed.). Belmont, CA: Wadsworth.

Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*, 43–50.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2011). *Research methods, design, and analysis*. Boston: Pearson.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Corsi, K. F., Winch, P. J., Kwiatkowski, C. F., & Booth, R. E. (2007). Childhood factors that precede drug injection: Is there a link? *Journal of Child and Family Studies, 16*, 808–818.

Foster, J. D. (2008). Incorporating personality into the investment model: Probing commitment processes across individual differences in narcissism. *Journal of Social and Personal Relationships, 25*, 211–223.

Groves, R. M., Fowler, F. J, Jr, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: Wiley.

Heckathorn, D. D. (1997). Respondent driven sampling: A new approach to hidden populations. *Social Problems, 44*, 174–199.

Kalton, G. (1983). *Introduction to survey sampling*. Newbury Park: Sage.

Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.

Knoke, D., & Kalleberg, A. L. (1994). Job training in U.S. organizations. *American Sociological Review, 59*, 537–546.

Krejecie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*, 607–610.

Lindstrom, L. E., & Benz, M. R. (2002). Phases of career development: Case studies of young women with learning disabilities. *Exceptional Children, 69*, 67–83.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Nguyen, P. (2005). Public opinion polls, chicken soup, and sample size. *Teaching Statistics, 27*(3), 89–92.

Peters, T. J., & Eachus, J. I. (1995). Achieving equal probability of selection under various random sampling strategies. *Paediatric and Perinatal Epidemiology, 9*, 219–224.

Sadler, G., Lee, H., Lim, R., & Fullerton, J. (2010). Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing and Health Sciences, 12*(3), 369–374. doi: 10.1111/j.1442-2018.2010.00541.x.

Scheaffer, R. L., Mendenhall, W., & Ott, R. L. (1996). *Elementary survey sampling* (5th ed.). Belmont, CA: Duxbury Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Singlton, R. A., & Straits, B. C. (2010). *Approaches to social research* (5th ed.). New York: Oxford University Press.

Tryfos, P. (1996). *Sampling methods for applied research: Text and cases*. New York, NY: Wiley.

U.S. Census Bureau (2005). *Current Population Survey*. February 2005. Contingent Work Supplemental File Technical Documentation: CPS-05, Washington, DC.

Frames, Framing Effects, and Survey Responses

6

Loretta J. Stalans

Crime and the criminal justice system are omnipresent in newspaper stories and news reports. Media reporters encourage the public's fascination with crimes and criminal offenders. What the public thinks about crime and about the criminal justice system's response to crime partly depends upon how stories are framed. Frames are "ideas that interpret, define, and give meaning to social and cultural phenomenon" (Baylor 1996, p. 1). In making decisions about how to write a story, media reporters frame a story line that highlights certain considerations, and ignores or briefly covers other considerations of an issue. For example, a recent story on CNN discussed Afghanistan police enforcing the laws against farmers growing small crops of poppy, which is the major source of heroin. The reporters, using a survival frame to create a coherent story line, sympathetically discussed the farmers' decision to break the law as necessary survival; the farmers grew poppy crops because major drug lords would kill their family if they refused to grow it and they had limited options for money to support their family. The frame of survival gave specific meaning to the farmer's actions, which overshadowed other possible interpretations such as the farmers' greed, or anger toward western civilization. As this example illustrates, a story's frame may shape viewers' opinions about Afghanistan farmers.

A frame is the essence of the story, the central guiding theme, or a salient strategy that suggests how a decision should be made. Frames can be part of the survey introduction, a phrase in a question, a salient media story, or can be habitual internal ways of approaching particular issues or questions. In writing or interpreting stories about crimes and the enforcement of criminal law, there are numerous possible frames: accepting responsibility and restoring relationships, protecting public safety, opportunity for change, loss of innocence, learning survival skills, pain and despair, relief and hope, effective compassion, or providing appropriate severe sanctions for repeat offenders. Just like a picture frame draws attention to a visual image and places it in the foreground of a wall, a decision frame makes certain information more critical and guides inferences and emotions that shape expressed opinions, attitudes, or decisions.

When do people adopt a frame and use it to form opinions and attitudes? This chapter provides a detailed examination of the "framing effect." The framing effect occurs when the framing of a message causes individuals to assign greater importance to the considerations highlighted in the message in forming their opinions about an issue (Druckman 2001). It is one example of a context effect, where the

L. J. Stalans (✉)
Department of Criminal Justice and Criminology, Loyola University Chicago, 820 N. Michigan Avenue, Chicago, IL 60611, USA
e-mail: lstalan@luc.edu

L. J. Stalans
Department of Psychology and Associate Faculty of the Women's Studies and Gender Studies Program, Loyola University Chicago, Chicago, IL, USA

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_6, © Springer Science+Business Media New York 2012

salient features of an environment or a survey affect how people interpret information and form their opinions, attitudes, or perceptions. The chapter begins with a review of the research demonstrating effects of a survey's sponsorship and introduction on respondents' responses to participating in surveys and to the interpretation of questions. Next, the different types of external frames that are used in survey questions and media stories and their effects on public views are discussed. The chapter then describes the potential decision-making processes underlying the framing effects, and draws distinctions between general context effects of questions and framing effects on information processing. It discusses the individual, contextual, and attitudinal characteristics that have been shown to modify framing effects, and their implications for increasing survey response rates and reducing methodological biases in survey responses. The chapter concludes with directions for future research and policy implications about context effects, particularly the framing effect.

Both the context of the survey, its sponsorship and introduction, and the normative, cultural, and political context of the time when the survey is administered can serve to frame related issues. An understanding of the potential framing effects can provide important information about the stability of the attitudes or beliefs across time, across different populations, and across different political, social and cultural contexts. Framing effects may limit the external validity of the study to specific times, cultures, groups, and may also suggest methodological biases that challenge whether the questions are actually measuring what they intend to measure, especially for sensitive topics. For example, questions that implicitly or explicitly address racial relations and beliefs are very sensitive and have become more sensitive across time. For example, research found that only 18% of respondents in a 1972 poll compared to 35% of respondents in a 1992 refused to give an opinion on a question related to integration of public schools (Berinsky 2002). Based on selection bias analyses, Berinsky concludes:

> In both the 1970s and today, the pool of "don't know" respondents contain a mix of, first, those individuals who are unable to form coherent and well grounded opinions on the integration question and, second, those individuals who are able to form such opinions but are hesitant to express them for fear of social sanctions. Taken together, this work and my earlier work indicate that in the 1970s, the mix of "don't knows" were largely of the first type; in the present day, many more are of the second type (p. 579).

Berinsky (2002) found similar results for questions related to affirmative action. This study indicates that the wider social context of the survey may frame responses to survey questions, and relevant cultural and social norms should be considered in designing and interpreting surveys.

6.1 A Survey's Introduction Effects on Participation

Research also has shown that how researchers and interviewers frame the survey's introduction and make the request for participation may affect individuals' willingness to participate in the survey (e.g., Woodside and Dubelaar 2003). Researchers, of course, want to minimize nonresponses to surveys, as nonresponses may create selection bias and limit generalizability. Psychological research suggests several norms and decision frames, which may increase individuals' willingness to participate in surveys (Groves et al. 1992; Kahneman and Tversky 1984). In addition, prospect theory (Kahneman and Tversky 1984) and framing research (e.g., Tourangeau and Cong 2009) suggest several potential ways to frame survey requests. Box 7.1 delineates six principles that may increase cooperation with requests to participate in surveys: civic duty, reciprocity, scarcity, legitimate authority, consistency, and sensitivity to negative consequences.

Of course, most people are more likely to participate in a survey if it personally appeals to their interests or values. Leverage-salience theory suggests that different approaches or frames will differ in their persuasive appeal across individuals, and it is important to match the

salient appeal to an individual's interests, values, or habitual ways of making decisions (i.e., internal decision frames) (Groves et al. 2000). Much survey research has highlighted the importance of tailoring the request for participation to the respondent's needs, interests, or values. Few studies have investigated different introductory frames and which attitudes, values, or personal characteristics may modify the effectiveness of the principles, listed in Box 7.1, at enhancing response rates to survey requests.

Civic duty has often been one explanation for why individuals are willing to participate in surveys. The robustness of this volunteer and helping spirit sometimes allow researchers to use simple frames to increase survey response rates. For example, Mowen and Cialdini (1980) increased participation in a survey by 19% points for the group who received the appeal, "it will really help us out" at the end of the request compared to the group who did not receive this appeal. Moreover, individuals who place a high importance on civic duty may complete a survey without being given additional incentives. Survey studies have consistently showed that a token monetary incentive increases response rates, and many marketing, university, and charity organizations use this technique to enhance response rates (see Toepoel Chap. 13). However, for some groups of individuals, a monetary incentive is not very cost-effective and may be unnecessary, whereas for other groups a small investment yields big returns. For example, a small monetary incentive of five dollars increased participation by 42% points in a survey for respondents who showed little prior commitment to civic duty volunteer activities, but only showed a modest increase of 16% points for respondents who had previously participated in civic duty volunteer activities (Groves et al. 2000). Individuals who participate in civic duty volunteer activities may habitually approach requests with the decision frame of whether it can be easily done and supports their civic duty value. For individuals who do not have a chronically accessible civic duty mindset, the monetary incentive may be a persuasive appeal for several reasons. A small monetary incentive may cue the norm of reciprocity; individuals may feel obligated to return the small favor of five dollars with a small amount of their time to complete a survey (i.e., quid pro quo). Also the monetary incentive may evoke the consistency principle if they keep the money and they do not participate it may evoke inconsistency with their self-image as an honest person. Additional research should examine the normative and self-interested principles that contribute to the effectiveness of small monetary incentives and the circumstances under which these incentives do not yield cost effective returns.

Box 6.1. Description of Norms and Decision Frames Used to Increase Participation in Surveys

Psychological research on cooperative behavior and decision making suggests the following ways to increase survey participation:

Reciprocity Principle—individuals believe they are obligated to return favors, gifts, and concessions. Thus, small monetary incentives, asking to answer just a few questions after being turned down to complete the entire survey, and providing favorable information are some ways researchers have capitalized on the reciprocity principle (Groves et al. 1992).

Sensitivity to Negative Consequences—for goal or personally relevant actions information about the negative consequences if one fails to take action are more effective than information about the positive benefits if one takes action (Kahneman and Tversky 1984).

Civic Duty Principle—Helping the larger good is part of being a good citizen. Individuals, if requested from a legitimate research authority, may believe participation in survey research is part of their civic duty.

Scarcity principle—individuals believe that limited opportunities are more valuable. Thus, statements that the survey is

sent to a selective group or will only be available for a short while may increase participation (Groves et al. 1992; Porter and Whitcomb 2003).

Consistency principle—individuals prefer to be consistent in their beliefs and behaviors. Thus, if individuals agree to a similar request or express a view that supports participation, they are more likely to participate (Groves et al. 1992).

Legitimate Authority principle—individuals judge the credibility and legitimacy of the requestor, and are more likely to comply if the source is a legitimate, credible authority. For example, research on crime and justice may have higher response rates if interviewers mention that a well known university or the National Institute of Justice is the study sponsor than if the sponsor is not mentioned or is a less legitimate source such as a newspaper.

Interestingly, the reciprocity principle that may contribute to the effectiveness of monetary incentives may also be less effective than the negative consequence approach. The reciprocity principle and the sensitivity to negative consequences principle suggest two very different and potentially conflicting approaches to creating persuasive appeals to increase survey participation. Based on the reciprocity principle, telephone survey introductions and letters to request participation in a mail or Web-based survey often emphasize the benefits of participation to society and to the respondent. Incentives, especially those that are personally relevant, are thought to be most appealing. However, prospect theory (Kahneman and Tversky 1979) as well as goal framing research indicate that individuals are more sensitive to potential losses if they fail to act than the potential personal gains if they do comply with a request. Research on information processing also generally has found that people place greater importance on negative information in decisions about norms and morality. One study has tested the relative effectiveness of a frame emphasizing the potential benefits for cooperating with a survey request or a frame emphasizing the potential losses if they do not comply with the request to participate in a survey. Researchers at the end of the first interview either framed the request for the second interview as a gain ("the information you've already provided to use will be a lot more valuable if you complete the second interview") or a loss ("Unfortunately, the information you've already provided to us will be much less valuable unless you complete the second interview"). A significant increase in participation in the second interview was evident in the loss frame (87.5%) compared to the gain frame (77.9%) in a nationwide sample (Tourangeau and Cong 2009). This framing effect was robust across demographics, except it had the greatest impact on respondents in the 45–64 age category; middle-aged respondents may have more demands and less interest in participating and framing effects should increase when people are more ambivalent. Research also suggests that an e-mail request to participate in a Web-based survey introduction is more persuasive if it uses both the principle of lost opportunity and the scarcity principle than several other types of frames. E-mails that indicated individuals were part of a selected small group (i.e., the scarcity principle) and stated a specific deadline date when the Web-based survey would be discontinued (i.e., loss of opportunity) increased participation among a sample of 12,000 high school students compared to when the e-mails did not contain scarcity and loss of opportunity information (Porter and Whitcomb 2003). Moreover, this study found that personalized e-mails and e-mails from "legitimate authorities" did not significantly increase response rates; given the abundance of spam e-mails and difficulty of actually determining whether the signatures and names on e-mails are from the actual persons, it is not surprising that these techniques do not work for e-mail solicitations. The findings from these two studies are consistent with several studies showing that people place greater importance on the negative consequences if they fail to take actions than the positive benefits if they do take action (Levin et al. 1998).

The principle to act consistent with beliefs and prior actions also modifies how sponsorship identification affects survey participation in marketing research. Woodside and Dubelaar (2003) reviewed 32 tourism advertising studies; the studies investigated whether individuals who requested brochure information about trips were more likely to participate in a survey if the brand/sponsor of the trip was identified or absence from the introductory of the survey, and whether sponsorship effects depended upon whether individuals bought a trip or did not buy a trip before the survey request was made. Across the studies, nonbuyers were less likely to participate in surveys when the brand sponsorship was identified compared to absent in the survey request. Conversely, buyers were more likely to participate when the survey request included the sponsor of a trip they had purchased than when no specific sponsor was mentioned. Similarly, nonvoters compared to voters were less likely to participate in a survey when the request stated that the survey was about politics than when the request had a neutral frame (Tourangeau et al. 2010). Both of these studies support the consistency principle, though the findings also may be due to personal relevance. Nonvoters may have seen less relevance of a survey to their lives when it was framed as being about politics; however, personal relevance cannot account completely for the different actions of nonbuyers in the sponsorship present and sponsorship absence conditions.

6.2 Introduction and Sponsorship Effects on Responses to Questions

How a survey's introduction is framed and the stated sponsorship of a research study may also serve as an interpretative frame for respondents. Galesic and Tourangeau (2007) suggest that sponsorship and introductions can affect information processing in three ways. First, noting the sponsorship may result in social desirability or experimental demand biases; respondents may select from their array of possible responses the responses that they think the researcher wants to hear. Second, sponsorship may serve to stimulate recall of particular information, especially if sponsorship is related to groups with well-known positions. Finally, sponsorship may be used to interpret ambiguous and unclear questions. Only a few studies, to date, have investigated sponsorship effects (e.g., Galesic and Tourangeau 2007; Norenzayan and Swartz 1999; Plesser et al. 1992; Smith et al. 2006). Researchers should be cautious in revealing sponsors that may suggest a specific way of interpreting the survey questions or have specific positions. For example, students who were asked to provide reasons why a murder occurred were more likely to provide answers that were congruent with the sponsors' focus. Thus, students who received a questionnaire sponsored by the "Institute of Personality Research" were more likely to provide personality trait explanations and fewer societal or family explanations. Conversely, students who received a questionnaire sponsored by the "Institute of Social Research" provided more societal level explanations and fewer personality trait explanations. Similarly, a large sample of adult respondents in the workforce reported more potential sexual harassment acts and interpreted more ambiguous acts as sexual harassment when the survey sponsor was a feminist group fighting against sexual harassment than when the survey was a university investigating the atmosphere of the work environment (Galesic and Tourangeau 2007).

Research also has found that the sponsorship of the survey may affect the relationship between questions and how questions are interpreted. For example, after individuals with Parkinson Disease had agreed to participate in a 10 min survey, researchers randomly assigned them to a neutral sponsorship and frame (researchers at the University of Pennsylvania who are trying to understand the quality of life of people who live in the eastern United States) or a health-related sponsorship and frame (researchers at the University of Pennsylvania Movement Disorders who are trying to understand the quality of life of people with

Parkinson's Disease). Researchers found a significantly stronger relationship between satisfaction with health and an overall quality of life question in the health-related frame than in the neutral frame.

One important issue for society is the prevalence of intimate partner abuse and whether men and women are equally likely to intentionally physically harm their intimate partner. How surveys introduce the topic of intimate partner violence has been shown to affect the reported rates of intimate partner violence (Regan 2010). Generally, three different types of surveys have measured the prevalence of intimate partner violence: (a) Conflict Tactics Scale (see Kimmel 2002); (b) National Violence Against Women Survey; and (c) National Crime Victimization Survey. Regan (2010) examined the effects of the overall framing of the survey (conflict tactics, personal safety, or crimes) and how the question was introduced on reported levels of partner violence. She found that the personal safety frame, which is typically used in the Violence Against Women survey, yielded higher rates of intimate partner violence than introductions using a context of how conflicts are resolved or victimization from crimes. Moreover, questions that asked respondents about attacks from intimate partners produced higher reported rates of intimate partner violence, whereas questions that asked respondents about how often their partner used different tactics to resolve conflicts produced the lowest rates.

What is identified as sexual harassment in the work place also depends upon whether the survey is sponsored by a neutral group or a group with a specific agenda. Galesic and Tourangeau (2007) randomly assigned adults to receive either a Web-based survey that was sponsored by a neutral research center and framed as a "Work Atmosphere Survey" or a Web-based survey that was sponsored by a feminist group dedicated to fighting sexual harassment and framed as a "Sexual Harassment Survey." Consistent with either the desire to provide relevant information or please the researchers, individuals who received the neutral sponsorship were more likely to recall and report encountering nonsexual harassment situations (e.g., loud personal telephone calls at work) more frequently than more potentially sexual harassment situations (e.g., manager commenting that a worker looked nice in their clothes), whereas respondents who received the feminist frame showed the opposite reporting pattern (Galesic and Tourangeau 2007). Moreover, the study found that individuals receiving the feminist frame were more likely to interpret ambiguous items such as asking a coworker out for a drink or a manager stating that the person looked nice in their clothes as more sexually harassing than individuals who received the neutral frame. Neutral frames should be used to avoid biasing responses, recall, and interpretation of questions. Moreover, researchers also should examine how personally relevant frames may be used to interpret general questions and affect the strength of the correlations with other related questions.

6.3 Typology of Frames

There have been several different typologies of external frames. These typologies of frames are based on types of frames used in media reports and sometimes in survey sponsorship or introduction, which are subsequently tested for their effects in experimental research. Box 7.2 highlights the different forms of framing and provides examples of each form. This typology of framing provides the broad categories of different framing manipulations that have been used in prior research. Box 6.2 shows examples of five types of external frames: issue, decision, value, visual, and emotional frames. More research has been conducted on the effects of issue, decision, and value frames than visual or emotional frames.

The general category of issue frames, taking an evaluative stance or specific focus on an issue, includes a couple of specific types of frames. For example, one type of manipulation of an issue frame is whether the frame discusses what will be gained (gain frame) or what will be lost (a loss frame). For example, recent research

indicates that women offenders have an average recidivism rate of 45%. Before asking respondents about whether prisons are effective at rehabilitating offenders, a gain-loss framing manipulation could be implemented. Respondents in the gain framing condition could be told "research shows that within the first three years of release from prison 55% of women offenders comply with laws and are free of new arrests." Respondents in the loss framing condition could be told "research shows that within the first three years of release from prison 45% of women offenders violate a law and are arrested for a new crime." These gain-loss framing comparisons, also called attribute framing (Levin et al. 1998) are known as "pure" framing effects because the information is the same across the two conditions of the manipulation, but whether it is stated in a positive way (success rate) or a negative way (recidivism rate) differs.

Research on gain-loss framing has focused primarily on health care and advertising (see Levin et al. 1998); studies show that healthcare procedures receive higher approval rates when success rates are emphasized rather than mortality rates. Consistent with these studies, research has found that respondents judged justice principles of allocation such as merit, ability, need, and tenure more justly when they were framed in a positive way (to deliver benefits or to withhold negative consequences) compared to when they were framed in a negative way (to deliver negative consequences or to withhold benefits) (Gamliel and Peer 2006). Research also has found that respondents assigned significantly higher fines when they were told that an advertising firm had a 80% chance of knowing its advertisement was deceptive (negative frame) compared to when they were told that the firm had a 20% chance of not knowing that its advertisement was deceptive (positive frame) (Dunegan 1996). Consistent with the findings for deceptive advertisement, much research has shown that negative information is more diagnostic and receives more attention in evaluations of morality; however, in judgments of ability or competence positive information compared to negative information is given more consideration. Thus, if respondents are judging the competence or effectiveness of the police, court, or other criminal justice interventions or organizations, a gain or positive frame will be more persuasive than a loss or negative frame. However, if respondents are asked to sentence an offender or make ethical or moral judgments, a loss or negative frame will receive more consideration than a positive or gain information.

Box 6.2. Different Types of Frames

Shootings occurring at schools continue to shock the public and leave policy makers, media reporters, and the public searching for answers. Using this topic, different forms of external frames in surveys will be illustrated.

1. Issue Frames—take an evaluative position on an issue. "Who or what do you think is most to blame for why this tragedy occurred?" Respondents either received a "blaming gun law" frame ("Many people are suggesting that weak gun control laws should be blamed for the shootings.") or a blaming media violence frame ("Many people are suggesting that violence on television, in the movies, and in video games should be blamed.") (Haider-Markel and Joslyn 2001)
2. Decision Frames—focus on a particular way to make a decision. People have indicated several sanctions for carrying guns or other weapons to the school. Please indicate which of these you believe are appropriate (inclusion) or indicate which of these you believe are not appropriate.
3. Value Frames—focus on a value to discuss the issue such as privacy, safety, freedom, honesty, autonomy, self-reliance. Before respondents answered the question "What do you think about concealed handgun laws?" They were randomly assigned to either receive a "right to protect themselves" frame or a "threaten public safety" frame as part of

the introduction to the question (Haider–Markel and Joslyn 2001).

4. Visual Frames—use images and pictures to convey meaning, emotion, and an idea about a topic. For example, pictures of students running from gunfire at a school or children who are injured or parents' crying, or a teenager walking home who protects himself from a mountain lion with a handgun could be used to frame the handgun issue and school safety.
5. Emotional Frames—can use either statements or pictures to convey specific emotions. For example, respondents could be asked "if a fourteen year-old student brings a weapon to school, what sanction do you support. a school suspension, expulsion from school, charged and tried in juvenile court, or charged and tried as an adult for illegal possession of weapon?" Before asking this question, respondents were shown a picture to evoke a "sympathy frame" or a "anger frame."

Another example of issue framing focuses on a comparison of in-group cost information or out-group benefit information. This manipulation is not a "pure" framing manipulation as the information varies across the conditions, and it focuses on how stereotypes and prejudices about groups may influence response choices. In an investigation of people's views about ethnic targeted school policies, researchers asked Dutch respondents "do you agree or disagree that children with an ethnic minority background should receive extra attention at school?" Individuals either answered the question without a frame, with an in-group cost frame "some people argue that extra attention for ethnic minority children might come at the expense of attention for Dutch children" or with an out-group benefit frame "some people argue that extra attention for ethnic minority children increases their chance to get ahead in society." Respondents expressed more opposition to the extra attention policy when exposed to the in-group cost frame than the no-frame condition in both the 2000 and 2005 national survey, but the out-group benefit frame produced inconsistent and weak effects (Londen et al. 2010).

A second type of frame is a decision frame. Decision frames may often be evident in surveys as the way a question is asked, the order of choices, or may be implied through the order of related sets of questions in a survey. The manipulation of inclusion compared to exclusion decision strategy is one type of decision frame. For example, surveys may ask citizens which sanctions are most appropriate for youthful drug dealers and provide a list of ten sanctions (i.e., inclusion frame) or the survey may ask respondents to cross-out the sanctions that are not appropriate for youthful drug dealers (i.e., exclusion frame). If framing did not influence decisions, decision makers would select the same choices in both the inclusion and exclusion frames. Research, however, shows that individuals retain a greater number of options for political seats or politicians when asked to exclude candidates than when asked to pick candidates (Levin et al. 2000; Yaniv and Schul 2000; Hugenberg et al. 2006).

This framing effect occurs through two processing strategies. First, people are less likely to take action when faced with uncertainty; thus, the ambiguous or uncertain choices are retained in the exclusion frame, but are not selected in the inclusion frame (Yaniv and Schul 2000). Secondly, research shows that individuals reconsider each choice in the inclusion mind-set and devote more processing time to each choice than in the exclusion mind-set (Levin et al. 2000; Hugenberg et al. 2006). Inclusion and exclusion framing effects are very relevant for real life decisions. For example, probation officers must decide which sex offenders warrant greater supervision, and are at high risk of committing additional violent crimes. Probation officers, having high caseloads of sex offenders, must divide their supervision time by deciding which offenders are at a higher risk of committing additional sex or violent crimes while under probation supervision. With an inclusion frame, officers would select the high risk offenders from

the larger pool of all offenders on their caseload. Officers, using an exclusion framing strategy, would eliminate offenders on their caseload who were not high risk to identify the offenders who were high risk. As noted earlier, prior research findings suggest that probation officers will retain a greater number of offenders who are high risk if they approach the task using an exclusion frame than an inclusion frame.

Inclusion and exclusion framing also affects the extent to which stereotypes are used to make decisions or provide opinions. For example, when asked to select judges and politicians from a list (inclusion) or to cross-out the individuals who were not judges or politicians (exclusion), respondents had more difficulty making distinctions between women than between men in the exclusion mind-set than in an inclusion mind-set, which is known as sensitivity stereotyping. Moreover, respondents set lower criterion for labeling a man than a woman as a politician or judge in the exclusion mind-set than in the inclusion mind-set, which is known as criterion stereotyping (Hugenberg et al. 2006).

Another type of decision framing is goal framing (e.g., Levin et al. 1998); it has focused on manipulating whether the positive benefits of taking an action are highlighted or the negative consequences of not taking an action are highlighted. For example, a survey may ask victims of burglary about their willingness to participate in a restorative justice program where they meet the offender and come to an agreement on a sanction. Before the question victims could be told: (a) victims who participate in the restorative justice program have an increased chance of feeling emotionally better after their participation (positive frame); or (b) victims who choose not to participate in the restorative justice program have a decreased chance of feeling emotionally better (negative frame). The goal framing emphasizes one behavior and the potential to gain benefits or lose the benefits from the choice to participate in the behavior. Research generally finds that individuals are more likely to participate in the desired goal behavior when they receive a negative loss frame than when they receive either a positive gain frame or receive no frame (Petruvia 2010). This type of framing clearly has many implications on how information is provided about programs and how questions are asked about citizens' willingness to participate in community-based programs such as community-based policing, community notification of sex offender residences, community crime watch groups, and restorative justice or victim-offender mediation programs.

Prevention compared to promotion focused goal framing is another type of decision frame. Prevention frames focus on minimizing risks and preventing inaccurate decisions or failures (e.g., wrongly convicting an innocent person). Promotion frames emphasize action to promote successful goal attainment (e.g., arresting and convicting offenders) and place less emphasis on the potential risks of inaccurate decisions (O'Brien and Oyserman 2010). Because criminal justice decisions are focused on actions toward others for the benefit of society rather than actions to benefit oneself, the prevention focus is similar to the negative goal framing, and the promotion focus is similar to the positive goal framing. As applied to political models of the criminal justice setting, promotion focus frames are similar to the crime control model where the criminal justice system places greater importance in swift and severe justice of catching, convicting, and punishing as many guilty criminals as possible and believes that it is a greater error to have a judge or jury decide that a person who actually committed a crime is "not guilty" than it is to convict an innocent person. By contrast, the prevention focus frames are similar to the due process model which emphasizes careful and deliberate decision making so that the system seeks truth while still respecting the rights of defendants; the due process model argues that it is a much greater error to convict an innocent person than to have a judge or jury decide that a defendant who actually committed a crime is "not guilty." Conservatives and those who emphasize "law and order" are more likely to support crime control and may be more likely to have a promotion focus, whereas liberals and those who support due process and protecting

defendants' rights may be more likely to have a prevention focus.

Psychological studies have found that individuals can have chronically accessible prevention goal framing or promotion goal framing. Individuals who had chronically accessible promotion focus mind-set provided more severe punishment to students across a variety of drug crimes than did individuals who were low in promotion focus. Being high or low on internal frames of prevention focus, however, did not affect punishment decisions (O'Brien and Oyserman 2010). When a prevention or promotion focus frame was externally induced, respondents who received a promotion focused frame compared to those who received a prevention focused frame were more likely to recommend arresting a suspect, even though they were not more certain of his/her guilt. Individuals with a promotion focused frame provided a greater number of nonevidence related reasons for why the suspect should be arrested (e.g., he could flee; the family deserves justice) (O'Brien and Oyserman 2010). This study found that respondents were more likely to be consistent with the goal framing in their recommendations about what actions should be taken; prevention focus frames produced less action than promotion focused frames. However, this study did not contain a "no-frame" condition; thus, it is uncertain whether both frames affected decisions. Moreover, the study did not examine individuals' prior support for due process compared to crime control attitudes, which may vary systematically with chronically accessibility of prevention, and promotion goal internal decision frames. If due process beliefs are related to internal prevention focus framing, individuals with this mind-set may place more importance on rehabilitation, and restorative justice goals in sentencing offenders and may place less importance on deterrence and incapacitation.

Politicians and media reporters often frame controversial issues such as abortion, affirmative action, school voucher programs, flag burning, death penalty, and decisions about wars around values such as equality, freedom, compassion, or retribution. Research has manipulated whether frames direct respondents' attention toward a particular value (e.g., right to freedom of speech) or emphasize that a policy undermined or promoted a value. For example, respondents who were exposed to either a frame that suggested school vouchers promoted equality or a frame that suggested school vouchers undermined equality were significantly more likely to mention equality in their thoughts about school vouchers than were respondents who were not exposed to a frame or respondents who were exposed to both frames (Brewer and Gross 2005). Thus, exposure to only one side of the equality argument or to both sides of the equality appeared to have focused respondents' thoughts compared to the condition where respondents did not receive an equality argument. Supporting this finding, other research also suggests that respondents exposed to either an ethical interpretation or a pragmatic interpretation of a health issue were more likely to provide thoughts consistent with the frame they received (Shah et al. 1996).

Emotional frames may be manipulated and be evoked using either specific stories about individuals or pictures. Manipulations comparing episodic and thematic framing partly affect responses through eliciting particular emotions (Gross 2008). Episodic frames use specific individuals or events to provide a meaning to a topic. Thus, for example, if school shootings were the topic of the survey question, episodic frames would present a specific school shooting such as Columbine where the shooters killed themselves. Thematic framing would use text and more general statements including statistics on the number of school shootings. Thus, the episodic and thematic frame manipulates whether information is about the group as a whole or an exemplar from the group. For example, in one study, respondents were randomly assigned to read either a column arguing against mandatory minimum sentences using details about the sentencing guidelines, the increasing prison population, and the high cost of incarceration or a column focusing on a woman who received a harsh mandatory minimum sentence (i.e., episodic frame). The research found that overall the

thematic frame was more persuasive than the episodic frame; however, episodic framing increased emotional responses of empathy, sympathy, and compassion and decreased support for mandatory minimum sentences through these affective reactions (Gross 2008). Episodic and thematic framing are not "pure" framing effects as they present different information and manipulate whether the focus is at the group or individual level.

6.4 Framing Effects and Processing Information

While the survey and wider environmental context may suggest a frame, decision makers also have internal frames that may guide decisions. Thus, external frames in media stories or surveys are not automatically adopted and must compete for persuasive influence with individuals' internal frames (e.g., Edy and Meirick 2007). As the review of how survey introductions affect respondents' willingness to participate in a survey and their interpretation of questions suggested, knowledge of individuals' internal decision frames may be very valuable in increasing response rates, and in creating valid and reliable questions. Survey respondents may either consciously or unconsciously select a frame from their memory. People develop internal frames or ways to approach a decision-making task through childhood and professional socialization. These internal decision frames are associated with a web of beliefs, attitudes, values, and schemas (information about categories containing exemplars and representations of specific crimes, events, and groups) (Stalans 2002; Stalans and Finn 1995).

Because attitudes are embedded and connected to information, beliefs, events, and emotions, external decision frames may unconsciously and automatically without effortful thought make certain beliefs more accessible, which may then be used to interpret the question and form the attitudinal response for a survey question. Contextual cues may change the accessibility of certain beliefs, which is known as "priming" (see Tourangeau and Rasinski 1988). These priming effects occur even when survey questions are not biased or misleading. Thus, when the salient features of the context change, the accessibility of different information in memory also changes, which then may lead to a change in the attitudinal response. However, through repetition and habitual use of a decision frame, it may become chronically accessible without being cued from external frames. Internal decision frames are chronically accessible ways of approaching a particular decision and emphasize certain questions to ask, and the importance and interpretation of specific information.

For example, from a very young age most children think about dilemmas and conflict situations in terms of what is fair based on who is more blameworthy for the wrongful or injurious acts. When people guide their interpretations of situations using the filter of fairness based on relative blameworthiness, this way of approaching decisions has been labeled "a normative frame" (Stalans and Finn 1995). The normative frame is chronically accessible as a way to approach conflict situations as most people have habitually used it since early childhood. Through direct and indirect experiences as well as informal professional socialization, police officers between their first, and third year working as beat cops learn to shift from a normative frame to a pragmatic decision frame in making arrest decisions (e.g., Fielding 1988; Fielding and Fielding 1991; Van Maanen 1975). The pragmatic frame places more importance on concerns about efficiency and whether their actions will promote their own career interests. Research has found that laypersons and rookie police officers approached decisions about how to handle domestic violence situations using a normative frame. The rookie and lay respondents focused on the question: Who is responsible and more blameworthy for the intimate partner violence? Experienced police officers, however, generally utilized a pragmatic frame; they focused on the questions:

(a) If arrested, will the offender be convicted?; (b) which party is more dangerous and less credible?; (c) is the case worth my time to process an arrest? (Stalans and Finn 1995, 2006). Prior research has found empirical support for the effects of these internal normative and pragmatic decision frames on officers' and laypersons' interpretations of hypothetical domestic violence situations and their decisions about whether to arrest the batterer. In the hypothetical cases both disputants claimed that they acted in self-defense, but only the woman showed signs of moderately serious physical injuries. By varying the victim's mental state, Stalans and Finn (1995) were able to create a case where a normative frame suggested arrest was appropriate and the pragmatic frame suggested that arrest was not an efficient or practical option. In one condition, the wife displayed signs of a serious mental illness by talking to her dead mother during the police interview. From a normative perspective, a husband who injures a wife who has a serious mental illness and cannot control her aggressive or annoying actions is even more blameworthy than a husband who injures a wife who can control her actions. From a pragmatic perspective, a mentally ill victim is less credible and more dangerous than a normal victim, and thus arresting the husband for domestic battery will be less likely to result in a conviction. Rookie officers were more likely to arrest the male batterer if the victim was hearing voices during the police questioning than if the victim was normal; conversely, experienced officers were more likely to arrest the batterer when the victim was normal than when she was hearing voices. Demonstrating the use of a normative frame, the researchers found that rookie officers' and laypersons' ratings of whether the offender was capable of understanding the wrongfulness of his action predicted their decision to arrest or not, whereas these ratings were not important in experienced officers' decisions (Stalans and Finn 1995, 2006). This study provides evidence of how internal decision frames may shape inferences, determine the importance assigned to different considerations, and affect decisions.

6.5 Distinctions Between General Context Effects and Framing Effects

Social psychologists believe that individuals may hold several attitudes about a topic such as appropriate sanctions for criminal offenders; their responses are formed based on the nature of the beliefs, information, feelings, and values that are recalled. Attitudes are stored in memory in an associative network; a spider web is an appropriate analogy of how attitudes are stored in memory. The spider web shows that attitudes are embedded in other attitudes and are connected to images, beliefs, emotions, values, and experiences (see Stalans 2008).

Many researchers make a distinction between context effects and framing effects, and argue that separate decision-making processes may be operating. Framing effects are a subset of context effects and generally "pure" framing effects may have different information processing strategies than the typical context effect due to question ordering. Question order effects and question wording effects, especially those that cue categorical information, are often due to heuristic processing. For example, individuals in many polls have been asked the question, "Do you believe judges are too lenient, about right, or too severe in sentencing criminal offenders?" Research has found that respondents form their answer using exemplars and information that is most easily recalled or accessible; this process is known as an availability heuristic or accessibility bias (Stalans 1993); about 80% of respondents report that judges are "too lenient" and base this response on the easily recalled media stories of violent offenders in the media. Response order effects where the order of the response choices in a question affects the interpretation and response also are often due to an accessibility bias or a confirmatory bias in which respondents do not consider additional options once a reasonable satisfactory option is found (Holbrook et al. 2007). Across a wide array of dichotomous questions (e.g., who has responsibility for outcomes. The President or Congress?), Holbrook

and colleagues investigated question order effects. They found that the most response order effects were biased toward the response choice that respondents heard last, which is known as a "recency effect"; recency effects occurred mostly in questions that were more difficult to comprehend, encouraged respondents to wait until they heard all of the choices before making a judgment, occurred in longer surveys, and were more prevalent among respondents with less education (Holbrook et al. 2007). Another context effect, the extent to which the order of related questions in a survey affect the responses to the questions that occur later, also has been examined. "These results suggest that question order may be common with conceptually related items but that their impact is generally local, affecting answers to the items themselves but not answers to later questions". Several articles provide for a more complete description of how the survey context affects respondents' opinions (Tourangeau and Rasinkski 1988; Holbrook et al. 2007).

External framing effects occur when the survey instructions, question wording or immediate context of the question affect how individuals interpret the question, consider the possible response choices, and draw inferences and attributions from the media and contextual information. Framing effects generally are not due to selective recall or ease of recall, but to the greater importance placed on the considerations suggested in the frame or to selective attention to particular considerations suggested in the frame (Druckman 2001).

The Heuristic-Systematic Model of decision making suggests that individuals may process information on two different levels depending on the context and personal characteristics (Eagley and Chaiken 1993). When personal relevance is low, time constraints are high, the topic is unfamiliar, or the persons' attitudes are complex, individuals may use heuristic shortcuts and look for peripheral cues in the survey question. Thus under these conditions, individuals may provide an answer without a thoughtful consideration of all of the choices. Heuristic processing leads to less valid and stable answers, and may produce responses that do not accurately represent respondents' true opinions. Individuals may attend to the central features of the question and to survey introductions or media stories when the topic is personally relevant, they have a goal of accuracy, they have less time constraints, or they have more prior knowledge about the topic. Frames may be used as filters to interpret the information in questions in a way consistent with the evaluative nature of the frame or they may guide decisions to reduce the depth of information processing (Brewer and Gross 2005). Unlike, general context effects that operate through accessibility, there is empirical support that individuals make a conscious effort to judge the credibility of the source of the framed message. Druckman (2001) found that the framing effect was limited to credible sources, and that framed messages from less credible sources did not affect respondents' opinions, or decision process. Other research also has found that individuals may reject frames that are inconsistent with their prior beliefs or ways of making the decision. For example, men who self-identified as Republicans or Independents, were more resistance to the frame "of blaming lack of gun control" as the reason for the Columbine school shooting, and it is these groups who typically do not support gun control. By contrast, women and liberals were more likely to use the "blame lack of gun control" frame in their explanations for the Columbine shooting (Haider-Markel and Johnson 2001). Research also showed that the majority of the public framed the causes for the Columbine shooting differently than the frames used in media stories (Birkland and Lawrence 2009). Other research also suggests that respondents are more resistance to frames that do not fit with the prior attitudes in thinking about death penalty issue (Dardis et al. 2008) and about social protests (Arpan et al. 2006).

6.6 Limitations of Framing Effects: Counter-Frames

Often time in real life, media stories and conversations with people may present more than one way to frame an issue. Are the effects of

frames in the introduction of surveys or questions eliminated when respondents are presented with both or all sides of an issue? Several recent studies have addressed this question. Brewer and Gross (2005) found that presenting both the anti-school voucher equality frame and the proschool voucher equality frame significantly reduced the number of thoughts listed about school vouchers compared to the no-frame condition, and respondents were just as likely to mention equality in the counter-frame condition as in either of the single frame conditions. When researchers stimulate racial stereotype beliefs, presenting both the benefits to minority students and the costs to the majority ethnic students for extra attention in school did not attenuate the effect of the cost frame to the in-group; respondents were less supported of ethnic targeted school policies in both the in-group cost frame and the dual frame condition compared to no frame and the out-group benefit frame (Londen et al. 2010). This research suggests that counter-frames may be rejected if they are inconsistent with prior views or the issue addresses stereotypes such as race.

However, for issues where respondents have multifaceted and less formed opinions, framing helps respondents to focus on particular considerations and to use heuristic processing. For example, respondents at a Belgian University were more likely to interpret more positively a photo of asylum seekers if they read a story entitled "Stopping Human Traffickers Becomes Easier" than if they read a story entitled "Unmasking Fake Asylum-Seekers Becomes Easier." Moreover, when the story contained both frames the photo was interpreted as positively as the victim frame and significantly more positively than the intruder frame (Gorp et al. 2009).

6.7 Measures of Individual Differences in Attitude Strength

The extent to which studies and surveys assess fleeing, uninformed opinions, or more reliable and informed attitudes depends in part on whether individuals have well developed internal frames and whether questions contain cues that may trigger stereotypes or schemata. Individuals differ in the extent to which their attitudes are enduring or open to change. Much social psychological research on attitude formation and change has identified individual characteristics such as expertise (extent of knowledge about an issue), personal importance (how much an individual is personally concerned or interested in the topic), cognitive elaboration (how much they have thought about or discussed the information or topic), and other measurable features of attitudes that affect the extent to which individuals are susceptible to persuasive appeals or education materials as well as influences from the environmental context (for in-depth reviews see Stalans 2008). The impact of frames also may vary by the knowledge level of respondents, though the findings are mixed. Research on issue frames generally shows that low knowledge individuals are more susceptible to the framing manipulation where it does not have a significant effect on respondents with greater knowledge about the issue (e.g., Haider-Markel and Joslyn 2001). Other research has found no difference in the effectiveness of frames across level of involvement or extent of knowledge (e.g., Gorp et al. 2009) or has found that those with higher involvement or knowledge are more susceptible to frames (Brewer 2003; Rhee 1997).

6.8 Conclusions

How interviewers or introductory letters frame the appeal for survey requests has been shown to influence response rates and responses to questions. The research on the loss-gain frame for goal targeted behaviors indicates that appeals framed in terms of potential losses are more effective than appeals framed in terms of benefits for survey participation. Moreover, the reviewed research suggests that interviewers need to develop some screening questions to assess respondents' decision frames around survey requests. The typology of frames and the

principles that may be internal frames for requests to participate in surveys can be used to guide additional research on this topic. Additional research is needed on understanding the chronically accessible frames underlying responses to survey questions about crime, criminal justice practices, and justice as well as the information processing strategies underlying external framing effects. Screening questions to assess the importance of reciprocity, civic duty, consistency in belief and actions, and following requests of legitimate authorities may be helpful to create more effective survey introductions and appeals for participation. The framing effect has wide reaching implications for how surveys and measures are created. Clearly leading questions that present only one side of an issue should be avoided; however, as reviewed, even when both sides of an issue are presented in a survey question it may focus individuals' thoughts on the values or emotions evoked in the frames. Research also suggests that seemingly open-ended questions where first an open-ended question is asked before the choices are given may reduce response order effects in questions (e.g., What do you think about your local police department? Are officers lazy or hardworking?) (Holbrook et al. 2007). The framing effects reviewed in this chapter suggest that researchers must be aware of the normative and cultural frames beyond the survey context that may influence respondents' opinions. Moreover, counter-frames (or presenting all sides of an issue) may not eliminate framing effects, especially for issues where respondents have complex multi-faceted opinions, such as sentencing criminal offenders, and justice issues.

References

Arpan, L. M., Baker, K., Lee, Y., Jung, T., Lorusso, L., & Smith, J. (2006). News coverage of social protests and the effects of photographs and prior attitudes. *Mass Communication and Society, 9*(1), 1–20.

Baylor, T. (1996). Media framing of movement protest: The case of American Indian protest. *Social Science Journal, 33*, 241–256.

Berinsky, A. J. (2002). Political context and the survey response: The dynamics of racial policy opinion. *The Journal of Politics, 64*, 567–584.

Birkland, T. A., & Lawrence, R. G. (2009). Media framing and policy change after columbine. *American Behavioral Scientist, 52*, 1405–1425.

Brewer, P. R. (2003). Values, political knowledge, and public opinion about gay rights. *Public Opinion Quarterly, 67*, 173–201.

Brewer, P. R., & Gross, K. (2005). Values, framing, and citizens' thoughts about policy issues: Effects on content and quantity. *Political Psychology, 26*(6), 929–948.

Dardis, F. E., Baumgartner, F. R., Boydstun, A. E., De Boef, S., & Shen, F. (2008). Media framing of capital punishment and its impact on individuals' cognitive responses. *Mass Communication and Society, 11*, 115–140.

Druckman, J. N. (2001). On the limits of framing effects: Who can frame? *The Journal of Politics, 63*(4), 1041–1066.

Dunegan, K. J. (1996). Fines, frames, and images: Examining formulation effects on punishment decisions. *Organizational Behavior and Human Decision Processes, 68*(1), 58–67.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitude*. Orlando, FL: Harcourt Brace.

Edy, J. A., & Meirick, P. C. (2007). Wanted, dead or alive: Media frames, frame adoption, and support for the war in Afghanistan. *Journal of Communication, 57*, 119–141.

Fielding, N. G. (1988). *Joining forces: Police training, socialization, and occupational competence*. New York: Routledge.

Fielding, N. G., & Fielding, J. (1991). Police attitudes to crime and punishment: Certainties and dilemmas. *British Journal of Criminology, 31*, 39–51.

Galesic, M., & Tourangeau, R. (2007). What is sexual harassment? It depends on who asks! Framing effects on survey responses. *Applied Cognitive Psychology, 21*, 189–202.

Gamliel, E., & Peer, E. (2006). Positive versus negative framing affects justice judgments. *Social Justice Research, 19*(3), 307–322.

Gorp, B. V., Vettehen, P. H., & Beentjes, J. W. (2009). Challenging the frame in the news: The role of issue involvement, attitude, and competing frames. *Journal of Media Psychology, 21*(4), 161–170.

Gross, K. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology, 29*(2), 169–192.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*(4), 475–495.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and illustration. *Public Opinion Quarterly, 64*, 299–308.

Haider-Markel, D. P., & Joslyn, M. R. (2001). Gun policy, opinion, tragedy, and blame attribution: The

conditional influence of issue frames. *The Journal of Politics, 63*(2), 520–543.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly, 71*(3), 325–348.

Hugenberg, K., Bodenhausen, G. V., & McLain, M. (2006). Framing discrimination: Effects of inclusion versus exclusion mind-sets on stereotypic judgments. *Journal of Personality and Social Psychology, 91*(6), 1020–1031.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under uncertainty. *Econometrica, 47*, 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–350.

Kimmel, M. S. (2002). Gender symmetry in domestic violence: A substantive and methodological research review. *Violence Against Women, 8*(11), 1132–1363.

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes, 82*, 171–193.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149–188.

Londen, M., Coenders, M., & Scheepers, P. (2010). Effects of issue frames on aversion to ethnic-targeted school policies. *Methodology, 6*(3), 96–106.

Mowen, J. C., & Cialdin, R. B. (1980). On implementing the door-in-the-face compliance technique in a business context. *Journal of Marketing Research, 17*, 253–258.

Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology, 29*, 1011–1020.

O'Brien, B., & Oyserman, D. (2010). The shield of defense or the sword of prosecution? How self-regulatory focus relates to responses to crime. *Journal of Applied Social Psychology, 40*(8), 1849–1867.

Presser, S., Blair, J., & Triplett, T. (1992). Survey sponsorship, response rates, and response effects. *Social Science Quarterly, 73*, 699–702.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of contact type on web survey participation: Description and illustration. *Public Opinion Quarterly, 64*, 299–308.

Regan, K. V. (2010). Framing violence: The effects of survey context and question framing on reported rates of partner violence. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 71*(4-b), 2745.

Rhee, J. W. (1997). Strategy and issue frames in election champaign coverage: A social cognitive account of framing effects. *Journal of Communication, 47*, 26–48.

Shah, D. V., Domke, D., & Wackman, D. B. (1996). "To thine own self be true": Values, framing, and voter decision-making strategies. *Communication Research, 23*, 509–560.

Smith, D. M., Schwarz, N., Roberts, T., & Ubel, P. (2006). Why are you calling me? How study introductions change response patterns. *Quality of Life Research, 15*, 621–630.

Stalans, L. J. (1993). Citizens' crime stereotypes, biased recall, and punishment preferences in abstract cases: The educative role of interpersonal sources. *Law and Human Behavior, 17*(4), 451–470.

Stalans, L. J. (2002). Measuring attitudes to sentencing. In M. Hough & J. V. Roberts (Eds.), *Attitudes toward sentencing* (pp. 15–32). Cullompton, England: Willan Publishing.

Stalans, L. J. (2008). Measuring attitudes about sentencing and sentencing goals. In M. E. Oswald, S. Bieneck, & J. Hupfeld-Heinemann (Eds.), *Social psychology of punishment*. Boulder, Co: Westview Press.

Stalans, L. J., & Finn, M. A. (1995). How novice and experienced officers interpret wife assaults: Normative and efficiency frames. *Law & Society Review, 29*(2), 301–335.

Stalans, L. J., & Finn, M. A. (2006). Public's and police officers' interpretation and handling of domestic violence cases: Divergent realities. *Journal of Interpersonal Violence, 21*(9), 1129–1155.

Tourangeau, R., & Cong, Y. (2009). The framing of the survey request and panel attrition. *Public Opinion Quarterly, 73*(2), 338–348.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents. *Public Opinion Quarterly, 74*(3), 413–432.

Tourangeau, R., & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*, 299–314.

Van Maanen, J. (1975). Police socialization: A longitudinal examination of job attitudes in an urban police department. *Administrative Science Quarterly, 20*, 207–218.

Woodside, A. G., & Dubelaar, C. (2003). Increasing quality in measuring advertising effectiveness: A meta-analysis of question framing in conversion studies. *Journal of Advertising Research, 43*(6), 78–85.

Yavin, I., & Schul, Y. (2000). Acceptance and elimination procedures in choice: Noncomplementarity and the role of implied status quo. *Organizational Behavior and Human Decision Processes, 82*, 293–313.

The Art of Question Phrasing

7

Lior Gideon

A well-constructed survey is a powerful research tool. A badly written one, with complicated phrasing, poorly designed and thought of response alternatives, an incomplete introduction, and poorly framed items, however, can lead to both response and non-response bias, which may further translate to poor predication ability. These are important aspects of the total survey error (TSE), as previously discussed by Bautista (Chap. 4) and Stalans (Chap. 6), and will be the theoretical framework of this handbook.

As Neuman discusses in Chap. 14 on face-to-face interviews, writing good survey questions takes a lot of practice, and it is as much an art as it is a science. Posing the right questions for a given survey is also a challenge, as items with similar content may require different phrasing depending on the mode in which the survey is administered. In the previous chapter, Stalans added another important dimension to the structure of survey items by examining the frames and framing effects of introducing a survey and its items. Other chapters in this handbook also discuss the importance of correctly and clearly presenting survey items to enable stable measurement and generate results that are reliable and valid. Indeed, the art of question and questionnaire phrasing is woven through this whole handbook. How you choose to phrase a survey question can depend on how the survey is administered (as an interview, in writing, etc.), the culture or cultures of the subjects being surveyed, and what topics are being covered—and whether these topics are considered sensitive.

With all these factors at work, it is critical to understand that planning a questionnaire is not an easy task, as many may think. In fact, designing a good questionnaire calls for adhering to some very strict demands. In designing a survey questionnaire, researchers must carefully consider the aims of the study, its scope, and the various resources available. Questionnaires can be general and used as a guide to observations, or they can be highly structured, with the aim of quantifying the observations for future statistical analysis. Survey questionnaires can be designed to measure just one trait, or they can enjoy the flexibility of examining multiple issues. The aim of this chapter is to provide readers with basic guidelines for designing a survey questionnaire, with some pointers on how to avoid some common mistakes.

7.1 What is a Questionnaire?

Questionnaires should not be confused with interviews, though this is a common mistake. Questionnaires are self-report data collection

L. Gideon (✉)
Department of Law Police Science and CJA, John Jay College of Criminal Justice, 899 Tenth Ave, New York, NY 10019, USA
e-mail: lgideon@jjay.cuny.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_7, © Springer Science+Business Media New York 2012

instruments that can be administered by various modes (as discussed further in Part IV of this handbook), to various individuals in different locations as part of one's research project (Johnson and Christensen 2010). They are an effective tool for obtaining information on a variety of topics such as feelings, attitudes, behaviors, beliefs, values, potential actions, decision-making processes, needs, lifestyles, sense of social belonging, consumption and shopping patterns, or even simple demographic questions. Note that all of these topics can also be examined by observation and interviewing—but this is not the same as a well-structured questionnaire. Readers must understand that questionnaires are tools, while interviews are the means of communicating the various items in the questionnaire, and thus should be understood within the context of an interpersonal interaction (whether face-to-face or by telephone or Skype).

Questionnaires can include questions in a variety of formats (i.e., open-ended, close-ended, ranking, number ranking versus word ranking, relative ranking, etc.), and can ask about various topics of inquiry such as, *Where do you shop for groceries?* Or, *Who decided where you should apply for college?* Survey questions can also ask about the past—*Who prepared your taxes last year?*—or the present: *Do you support the Second Chance Act and prisoners' rehabilitation?* They can even enquire about a hypothetical future: *What would you do if someone you know stole from you?* Questionnaire items can ask for a simple yes-or-no answer, or they can be statements requesting that participants rank their responses. An example of the latter is taken from a social support study, which included the question *Please indicate for the following statement if you 'Strongly Agree'or 'Strongly Disagree': 'There is always someone there for me when I need help.'*

As straightforward as the above items may seem, phrasing them may be a difficult task, one that must not be taken lightly. In the following sections, principles on how to design a survey questionnaire and the items to appear in it will be discussed. The first stage is to properly introduce the survey.

7.2 Concept and Construct

Once the survey goal is established and clear research questions are in place, the researcher should carefully identify and define the variables of interest. These will normally be the main variables that were presented in the research aim, and the corresponding research questions. Each variable must then be defined to clarify the actual meaning. By defining the variables, the researcher takes the first step toward developing the survey items, which in fact are the constructs, or operational definitions, of what the researcher aims to observe in the empirical world by asking a specific question.

The variables of interest that will appear in the survey questionnaire should have two definitions: a conceptual definition and a construct definition. A *conceptual definition*, sometimes also referred to as a *nominal definition*, is a simple definition of a term so that everyone will know what the researcher means when he or she refers to the variables in the study. Conceptual or nominal definitions explain and define a given term, much as a dictionary does. But this does not mean a term used in a study has the same meaning that it does in the dictionary—a conceptual definition can be one that is true only within the context of a given study and is chosen specifically for a given topic. Take, for example, a survey that aims to examine "social support"—a very complex term that means a lot of things to a lot of people. The meaning of this term will also vary depending on the context of the research goal; some studies aim to examine the social support available to terminally ill individuals and their families, for instance, while others may aim to examine social support on a college campus as part of the freshman experience." In this case, we can use a commonly agreed-upon definition offered by Lin (1986)": social support is "the perceived or actual instrumental and/or expressive provision supplied by the community, social networks, and confiding partners" (p. 18). If you read carefully, you will note that some of the terms used to define social support are themselves somewhat vague and require additional definitions as well; however,

this basic conceptual definition does nonetheless explain what we mean by "social support."

Studies also require at least one *construct definition*, also referred to as an *operational definition*. This is the guide to what the researcher will be looking for in the empirical world and how the concept will be measured. A construct definition converts a concept into a workable measurement by focusing on the actual traits in the empirical world that will become the data of the study and the source of the analysis. It includes all the operations that must be conducted in order to measure the defined term. It is important that the researcher identifies the data of interest so that others will be able to replicate the study while observing similar traits. The outcome of a construct definition is then referred to as an empirical concept.

The best way to understand a construct definition is with an example: For a study interested in examining social support, the researcher must clearly identify how social support will be measured. That is, he or she must provide a construct definition for social support. In social research, social support is broadly defined by using Lin's previously presented conceptual definition, and it can be measured in a number of ways. A study could focus on family, friends, colleagues, intimate partners, or a number of other typical elements in a support network. Each of these data would provide a different result. For example, two researchers may aim to examine the social support available to substance abusers—but if one refers to family while the other refers to co-workers, the two researchers will likely receive completely different results. This in and of itself is not bad, but if they have not clarified the construct of their variables, they would engage unnecessarily in an argument about the accuracy of each other's data. A clear construct enables the researcher to determine without a doubt if a certain observation is relevant to the variable as defined by the concept of that given variable (we refer to this as face validity). This definition will also guide future readers of the study, who will be able to check the results.

The process of constructing (or "operationalizing," as some say) the variables is also an important step preceding meaningful analysis. The way in which we construct our variables later determines our ability to analyze the data, while also limiting or maximizing our statistical ability. So we must focus on levels of measurement and what types of statistical tests are appropriate for their analysis (see Gideon 2012). There are two types of constructs: a measured construct and an experimental construct. *Measured constructs* use theoretical concepts that were developed to measure specific traits or behaviors. Examples include the Drug Severity Index, the Level of Service Inventory Revised (LSI-R), the Redemption Scale devised by Maruna and King (2009), anxiety tests, and depression scales. These are constructs that stem from a long process of building an appropriate test to elicit responses that would be an indicator or a measure of the trait examined. Those who devise such measured constructs must also carefully monitor the way in which the test is administered, as well as the way in which it is interpreted and scored. *Experimental constructs*, on the other hand, specify the manipulation the researcher plans to apply during the experiment. Such manipulations must be carefully described to specify the nature of the manipulation as well as the time of its introduction. For example, consider a study analyzing the effect of a substance-abuse program. The program itself is the manipulation, and the study must specify the construct for the experiment: The program will be introduced to substance-abusing inmates after they have completed 6 months of their sentence and before their release. A good construct would also define an eligibility criterion for those inmates who can participate in the program: Inmates should have at least 12 months until their parole hearing, for instance, and have been using heroin for the past five years. The construct may further suggest that participation in the treatment may be counted only if the inmate has completed detoxification and an additional minimum of four months in group sessions. The scientific approach also requires that the conceptual definition and its construct be logically compatible. In order to be valid, the construct must correspond to the concept and measure it for the variable. It is also crucial for the

Table 7.1 Seven steps to design a survey questionnaire

	Description of the needed step
Step 1:	Define conceptual and construct variables according to the research objectives
Step 2:	Formulate preliminary survey items according to the above constructs
Step 3:	Examine preliminary questionnaire items for the following:
	3.1 Relevance to the main research topic
	3.2 Relevance to other questionnaire items
	3.3 Logical flow
	3.4 Avoid double-barreled questions
	3.5 Avoid double negative questions
	3.6 Avoid unnecessary repetitiveness
	3.7 Avoid "leading" or "loaded" questions
	3.8 Avoid biased questions
	3.9 Use simple and direct language
	3.10 Items should be short, clear, and precise
	3.11 Determine what type of item will best examine the concept (e.g., an open-ended, close-ended/multiple-choice, or ranking question)
	3.12 Make sure all possible responses are included
	3.13 Use simple items to measure complex concepts
	3.14 Triangulate items to introduce complex concepts and ensure reliability
Step 4:	Run an empirical examination in a small representative study (pilot study)
Step 5:	Correct and rephrase items according to findings from previous stage
Step 6:	Write an introduction and instructions
Step 7:	Make any final adjustments and modifications

construct and the concept to match the research goals; otherwise the items will suffer from low construct validity. The following section presents and discusses the principles of how to design a questionnaire.

7.3 How to Design a Questionnaire

God created the world in 7 days, so why should we think of designing a questionnaire in fewer than seven steps? Table 7.1 presents the stages required for designing a questionnaire and executing data collection tool.

Each of the steps will be discussed at length in the following sections of this chapter.

Step 1: Define Variables According to the Research Objectives

As previously discussed in this chapter, the first and most important step of any survey design should be clear identification of the research goal and the main variables that will be examined. Clear definition of the research question(s) and corresponding variables will guarantee that items enjoy face validity—that is, it will confirm that we are indeed examining the trait we aim to examine. It is strongly recommended that each and every item introduced to the questionnaire will serve as a clear measurement of the variables to be examined (that is, the constructs), and will also correspond with the nominal value (the concept) of the identified variable.

Depending on what type of research the survey is designed for—exploratory or confirmatory—researchers must make an attempt to cover the broad spectrum of potential questions

that may be relevant to the research question. This is so that participants will not feel as if important information is neglected from the study. With that in mind, exploratory studies should be designed with more flexibility for allowing information that is not part of the initial questionnaire. This is done so that later revisions of the items can include topics that may not have been included initially. On the other hand, in a confirmatory study, the researcher should make every attempt to finalize the questionnaire before fielding it. To avoid excluding relevant items, a thorough literature review should be completed before constructing the survey. In examining previous studies on similar topics, researchers can become familiar with definitions, theories, questions, and other tools and instruments used. This step helps to avoid a situation in which the researcher realizes only after the questionnaire is given out that the data coming in lacks validity and/or reliability. Failing to include important items may not only affect the results and the researcher's ability to fully understand the magnitude of the researched topic, but it may also reduce respondents' trust and appreciation of the survey and its potential outcomes.

Step 2: Formulate Preliminary Survey Items

As discussed above, researchers should always formulate survey items that are direct measures of the variables of interest. To aid this process, researchers should make an initial attempt to map the topics to be covered by the questionnaire items. Specifically, after conducting a thorough review or the literature or making field observations, researchers should create general titles under which specific items will appear. After general titles are in place, specific items can be developed to capture the topic presented under each title, and the way in which it corresponds with the general research goal. The researcher should also keep in mind who the potential participants are and what their limitations may be. If participants find the questions difficult to understand or offensive, or think they simply do not "make sense," they may not respond to the survey or specific items in it (Johnson and Christensen 2010). This will result in lower reliability and validity, while increasing response/non-response bias that will, in turn, increase the total survey error, as will be discussed later in this chapter.

Step 3: Examine Preliminary Questionnaire Items

As can be seen in Table 7.1, this step is the most demanding and requires the researcher to follow no less than 14 guidelines. Each guideline is designed to assist researchers in the daunting task of phrasing questionnaire items that will be reliable and valid. Compliance with these guidelines should help to minimize the total survey error that stems from weak survey items.

Since the first principle—relevance to the main research topic—was already discussed at length, the discussion will begin with the second guideline, with relevance to other questionnaire items.

7.3.1 Relevance to Other Questionnaire Items

In order for the questionnaire to include all potential options while also being exclusive, researchers should attempt to cover the full array of topics related to the given phenomenon to be examined by the survey. With that in mind, each item should be set against the other items to make sure it does not affect the others. Sometimes a specific item may create a "halo effect," meaning it can create a prejudice against another item. Poorly designed items may also unintentionally guide a response or provide a response to other items. This usually occurs when items are coupled together. For example, consider the two following survey items:

> Please indicate if you strongly support or strongly oppose the following statement: Having rich military experience should be an important requirement for a person to be elected president.
>
> Please state if you strongly agree or strongly disagree to the following statement: Presidential candidates may benefit from having some military experience.

Note the differences in emphasis between the two items. The first item emphasizes "an important requirement" while the second

question is more moderate and does not "stream" the respondent into a specific end of the scale.

Other bad practices may occur when the researcher agrees to join an *omnibus survey*. This is a survey that targets a specific demographic profile but asks questions on a variety of topics; usually a researcher would use this form of survey to cut their costs for sampling and interviewing, as the cost would be divided by all those who have interest in the survey. But such a decision may be detrimental to the reliability of the results, as a poorly designed omnibus survey may juxtapose items in a way that confuses the survey participant or makes the participant look at some of the items in the wrong light. Imagine, for instance, a researcher who is interested in the public perception of ethics in politics. It would be unfortunate if he were to join an *omnibus survey*, only to learn after the data were gathered that questions on soap and other detergents had preceded the questions about ethics in politics. This juxtaposition changes the entire tone of the survey, and makes the results anything but clean.

7.3.2 Logical Flow

Closely related to the previous requirement, researchers must make sure that items are presented in a logical order, so as to develop the gathering of information. Items on specific subtopics within the survey should appear under the same section. For example, in a survey about available enrichment programs for middle school students, all items on extracurricular activity should be presented in the same survey section, rather than scattered across different sections. Similarly, questions about community safety should be concentrated together to provide a logical flow of events. There is no point in asking about street lighting in one section and to ask another question about police presence, for example, in another section. Demographic items should not appear in the middle of the questionnaire. They must be reserved for the end of the survey or presented at the beginning, depending on the topic under examination, and the information required.

7.3.3 Double-Barreled Questions

Johnson and Christensen (2010) define "double-barreled" questions as those that "combine two or more issues or attitudes in a single item" (p. 175). One example comes from Converse and Presser (1986): "Do you think women and children should be given the first available flu shots?" (p. 13). Another double-barreled question might be, "Do you think young children should have access to cell phones and credit cards?" Each of the above examples asks for different things in one single question item. The way in which the item is presented creates confusion, and, in the case of the latter example, respondents may respond positively to the first part—of children having access to a cell phone—while not at all approving of the second part, access to credit cards. Unfortunately, the way in which the item is presented prevents the respondent from providing a clear answer. Moreover, the confusion extends to the survey results. As each of the issues may elicit a different attitude, combining them into one item makes it difficult to identify which attitude is being measured (Johnson and Christensen, 2010). Accordingly, as a general rule, it is highly recommended to avoid the following words while phrasing statement questions: *and, as well as,* and *in addition.* These will usually result in a double-barreled question.

7.3.4 Double Negative Questions

Why ask respondents if they are unfamiliar with the "negative question" rule, when you can make their life much simpler by asking, "Are you familiar with the 'negative question' rule?" Double negative questions are those that include two negatives. Asking a question with a double negative in it not only confuses the respondent, but it may also create some level of frustration, which can result in non-response. Therefore, double negative questions should be avoided at all costs. Converse and Presser (1986) point out that such double negative questions tend to creep unobserved into questions that ask for an agree/disagree response. They provide the following example:

> Please tell me whether you agree or disagree with the following statement about teachers in public schools: Teachers should not be required to supervise students in the halls, the lunchroom, and the school parking lot (p. 13).

While some may agree that teachers should not be required to supervise kids outside the classroom, the option of "disagree" may be confusing. Disagree to "should not" means agree, and thus response error may result from such confusion, which in turn adds to the total survey error by reducing the reliability and validity of the item.

It is also possible that some words researchers use in phrasing their questions will have a negative meaning. For example, according to Converse and Presser (1986), words like *control*, *restrict*, *forbid*, *ban*, *outlaw*, and *oppose* may do more harm than good. Consider their example: "*Do you favor or oppose a law outlawing guns in the state of Maryland?*" (p. 14).

7.3.5 Unnecessary Repetitiveness

While some repetition is at times desirable for reliability purposes, unnecessary repetitiveness is not only onerous for participants, but it is also time-consuming, as it prolongs the questionnaire. Repetitiveness may also be demoralizing. When you ask the same questions over and over, respondents are more likely to develop fatigue and frustration, and as a result will skip questions or stop taking the survey all together. As a result, non-response error will increase, and survey reliability will decrease. Excessive repetition may also result in respondents providing unintentional responses, which will work against the original aim of repetition, which is to measure inter-item reliability. So if you must repeat specific items that measure complex concepts, as will be discussed later in this chapter, it is recommended that such items be examined with extra care to make sure they are absolutely necessary, and that they are not placing an additional cognitive burden on respondents. Such items should be examined in the pilot stage, and may be removed before the survey goes out for its full intended exposure.

7.3.6 "Leading" or "Loaded" Questions

Leading survey items create an unnecessary bias that increases the response errors under the total survey error. According to Johanson and Christensen (2010), "a loaded question is one that contains emotionally charged words (words that create a positive or negative reaction)" (p. 174). Johnson and Christensen's examples of words that may have a negative effect on questioning include *politicians*, *communist*, *liberal*, *gay*, *drug abuser*, *drop out*, *welfare*. A leading question is similar—it is a question that is phrased in such a way that it directs the respondent toward a specific answer. An example:

> Do you agree that the government should do everything within its power to stop foreclosures?

- Yes, it should stop foreclosures.
- No, it should not stop foreclosures.
- Don't know/no opinion.

Opening the question with the phrase "do you agree" leads participants to the first response option, and thus may result in a response bias. A simple rephrasing can make this question more neutral:

> Do you think the government should find solutions to stop foreclosures?

- Yes, the government should find solutions to stop foreclosures.
- No, the government should not find solutions to stop foreclosures.
- Don't know/no opinion.

Another problematic example presented by Bonevac (cited in Johnson and Christensen 2010, p. 175) demonstrates a question that is both loaded and leading:

> Do you believe that you should keep more of your hard-earned money or that the government should get more of your money for increasing bureaucratic programs?

- Keep more of my hard-earned money.
- Give my money to increase bureaucratic governmental problems.
- Don't know/no opinion

Taking a wild guess, the above question probably resulted in very minimal variance,

if any. A question like the above will, without doubt, become useless for policy purposes. It is thus important to remember that when phrasing questions, the goal is to present neutral question and answers. Once the question is distorted, or biased, then the information received is of low quality and cannot be treated as reliable and valid. In the final analysis, a question that results in very little variance is a source for response error.

7.3.7 Biased Questions

Similar to the above principle of loaded and leading questions, a biased question is one that sets the response within the questions, or emphasizes one of the response options over the others, and by doing so "attracts" and "locks" the attention of the respondent. Consider the following example:

> The government spends billions of dollars in foreign aid to countries that are not always in agreement with our foreign policy. In your opinion, should the US government cease supporting such countries, while spending more in rebuilding the local economy?

- Yes, the US government should cease spending our money elsewhere.
- No, the US government should not cease from spending our money elsewhere.
- Don't know/no opinion.

Note that not only is the question biased, but the response options are also biased and direct the participants to the first option. Also, simple words like "spend" may create a bias, as the word "spend" may have a negative connotation to it. To avoid biased items, the researcher should make every effort to obtain responses that are undistorted by the use of particular wording. Questions should use neutral language, and there should be a broad array of response options. Also, the wording of the question itself should be carefully monitored to rule out non-neutral words—in this example, "spending" may have a negative connotation. Furthermore, researchers should refrain from presenting their own opinions or inserting unnecessary information that may affect the tone of the questions.

7.3.8 Simple and Direct Language (Brevity and Precision)

These two issues are imperative. You should always assume that although people want to help by answering your questions, they are also busy. For this reason, you should never burden them with long questions that require hours of investment in your survey. Short, to-the-point items are always the suggested way to go. Unless you absolutely must present a scenario, you will always be better off with short, focused questions. Such questions will also yield more accurate responses, and will be less likely to suffer from non-response.

7.3.9 Open-Ended Versus Closed-Ended Questions

What type of item is best for your questionnaire is a direct result of how you conceptualized and constructed your variables. If you aim to examine something that has been examined previously or that already has a wealth of literature available, it is probably best to use multiple-choice items; these are often referred to as closed-ended questions because the responses are limited. Using closed-ended items is also preferred when previous versions of the questionnaire and items were pretested and examined in preliminary pilot surveys. This is mainly due to the fact that such items receive the most relevant and accurate information; they are easy and quick to answer; and, on top of it all, they provide unity in response, and their variation is known ahead of time. However, presenting rigid response options may also create "mind blockage," locking respondents' attention to preconceived answers that may not always accurately reflect their original opinion or attitude. Also, the order in which the responses are set may affect the response.

On the other hand, open-ended items allow more flexibility, and enable the respondent to think through and provide a more detailed answer. As such, open-ended questions are more time-consuming in terms of coding and

analyzing, and by definition, they will have greater variation. It is also possible that open-ended questions will remain unanswered. Thus, when constructing survey items, it is wise to take into consideration the available information as well as the mode of administration. What is suitable for mail, e-mail, or Web-based questionnaires may not be suitable for phone surveys.

Open-ended items and closed-ended items also differ in their level of cognitive burden, especially depending on the mode of administration. Closed-ended surveys with multiple answer options may be difficult to digest in a phone survey, whereas in a mail, e-mail, or Web-based questionnaire, such items will be much easier to follow. On the other hand, open-ended questions in a phone survey or a face-to-face survey may be difficult to document accurately. When it comes to depth of information, there is an obvious advantage to open-ended items; however, such items may also provide irrelevant information. In Chap. 8, Gideon and Moskos discuss open-ended questions in more depth in the context of interviews and how structured the interview is.

Because the majority of survey questionnaires use closed-ended items, it is essential to focus on the complexity of constructing such items. Multiple-choice or close-ended items have two components: the question itself and the potential responses. These will be explored in the next sections.

7.4 Open or Closed: How to Choose

Open-ended questions are presented to the participant without any guidance from the interviewer. Such questions enable the participant to respond without any prejudice and as he or she may see fit. Closed-ended questions, on the other hand, offer participants an array of potential responses from which participants are meant to find the one that best fits their opinions and/or attitudes. Each question has an effect on the quality, reliability, and validity of responses, as discussed earlier in this chapter. What type of question is best should be considered in context, while taking into consideration the topic of the study and available information that can be used to measure the variable of interest. Questions such as what is one's gender, marital status, age, level of education, as well as many other simple demographic information items are assumed to be known for their categories, and can thus be closed-ended. Similarly, questions such as, "Who do you support for the presidency?" will also be suitable for closed-ended questions, assuming all candidate names are known to the public. Using such closed-ended format enables quick analysis. For the sake of testing the internal reliability of the questionnaire, two items examining the same trait can be used: one posed as an open-ended question and the other in a multiple-choice format. Often researchers will use a combination of the two forms of questions, presenting "Other" as the last choice in a multiple-choice question.

7.5 Direct, Implied, and Indirect Questions

A further differentiation in types of questions is whether they are direct or implied. A *direct question* asks respondents openly about their opinions, values, beliefs, and attitudes. But often, using a direct approach may embarrass or shame respondents, particularly if the item is sensitive in nature and requires reporting an undesirable behavior such as adultery. Consequently, asking direct questions may increase response and non-response errors. To solve these potential problems, methodologists have used psychological knowledge to develop implied questions. *Implied questions* work on the assumption that people find it easier to criticize others for their behavior, beliefs, and opinions. Specifically, implied questions rely on two basic principles: (1) it is easier to present others negatively; and (2) salient responses are our own and they can be applied easily to others. Here is an example of a direct question and an implied question, both seeking an opinion about the same behavior:

Direct: *Would you consider cheating on your spouse?*

a. Never b. Not likely c. Depends d. Yes

Implied: *Many people use their business trips as an opportunity for sexual adventures. From what you know or hear from others, would you say this statement…*:

a. Is not true b. May be true c. Is true. d. Is very much true.

Aside from the fact that the range of answers may not be balanced, the direct question may result in significantly more respondents selecting the "Never" option, as this presents them in a positive light. On the other hand, the corresponding implied questions enables the respondent to project the behavior to another while divulging his or her true beliefs without having to fear the embarrassment that may be associated with such undesirable behavior. Using implied questions may reduce response error and bias, but they can also draw criticism for low validity. Some researchers argue that people may know how others behave, and report accordingly. It is for this reason that implied questions are not recommended by default, and researchers interested in such items should assess their outcomes with a grain of salt.

To overcome the validity problems of the implied question, researchers tend to use indirect questions. *Indirect questions* are ones that ask about sensitive issues in disguise. Returning to the example of extramarital affairs discussed above, an indirect and somewhat humorous question on the topic could be:

> "When the cat's away, the mouse will play." Would you support this claim in regard to extramarital adventures?
>
> a. Not at all b. To some extent c. Sometimes d. Most of the time

The above example focuses on the respondent, and is thus different from the implied question presented earlier. Rather than asking about other's behavior, the question provides a more accurate reflection of the respondents own believes and attitude, while not explicitly asking the respondent for his or her actions.

7.6 Response Ranking

When respondents are asked to express their opinion about situations, public figures, policies, and the like, closed-ended items are typically used. One type of closed-ended item is *ranking*, a very common strategy in social surveys which indicates the importance or priority assigned by participants to an attitudinal object (Johnson and Christensen 2010). Respondents can be asked to rank options either numerically or verbally, in either an absolute ranking or a relative one. Note that the simplest version of ranking does not ask respondents to rank options themselves; it is simply a variation of a multiple-choice question that asks respondents to choose the most fitting response from a list of several ranked options.

Numerical ranking and *verbal ranking* do not differ in substance. Numerical ranking simply means that numbers are used to represent levels or rank. But when the nominal value of the response is crucial for insight, and the researcher does not want to risk losing it by reducing responses to numbers, verbal ranking can be used. In this case, numbers are added only later for the purpose of statistical analysis. Ranking (either numerical or verbal) enables researchers to examine correlations between items, though at least five response options are needed to provide stable statistical analysis with sufficient variation. Researchers also use ranking when the content of a question may affect number of possible responses or when the magnitude of the response is important.

It is important to note that when a researcher would like to elicit an unequivocal opinion, he or she should offer an even number of possibilities in the response. On the other hand, to examine apathy about certain topics, an odd number of possibilities should be provided to enable an "escape to the middle." Here is an example of a verbal ranking:

> How satisfied are you with your current place of employment?
>
> Highly satisfied Satisfied Not satisfied Highly unsatisfied

After data are gathered, each of the categories—highly satisfied to highly unsatisfied—will be assigned a number for statistical purposes. Here is another example of numerical ranking:

> The scale below represents satisfaction in your current place of employment, with 9 being the highest level of satisfaction and 0 the lowest. Please mark your level of satisfaction:
>
> 0—1—2—3—4—5—6—7—8—9

Note that both examples use ranking, and although one is verbal and the other is numerical, they both use an ordinal level of measurement, and thus it is not too difficult for researchers to convert one to the other by substituting numbers for words or vice versa. In fact, many ranking items that appear in questionnaires tend to combine both methods, thus presenting the numbers with their corresponding verbal attribute. An additional observation should be made, in the above numeric example, ten categories were used. While this is still ordinal levels of measurement, the wealth of response categories imitates an interval level of measurement, and hence enables researchers to calculate a mean response, which may make statistical analysis easier.

7.7 Definite Versus Relative Rank

Definite ranking is used when a more complex evaluation of items is needed. The key to definite ranking is that the respondent is asked to rank each item independent of other items. On the other hand, *relative ranking* is used when several items are ranked in relation to one another. Researchers use relative ranking when they want to know the most (or least) important items in a long list of characteristics.

Survey items using definite ranking most often employ a so-called Likert scale, a structure that allows researchers to measure the degree to which respondents hold a number of particular attitudes. This is an example of a Likert scale with six items on the subject of job satisfaction:

People look for different things in a job. The table below lists six factors in job satisfaction. For each one, please indicate its level of importance to you.

	Not important at all	Not very important	Important	Very important
Work that pays well	1	2	3	4
Work that gives a sense of accomplishment	1	2	3	4
Work where you can advance	1	2	3	4
Pleasant working environment	1	2	3	4
Steady place of work with minimal chances of being laid off	1	2	3	4
Work that provides good social benefits	1	2	3	4

This method of definite ranking enables researchers to collect separate data on each of the factors that contribute to job satisfaction. Note, however, that each factor is ranked independent of the others, so this method does not allow researchers to assess which factor is more or less important than the others. To achieve this goal, researchers can use relative ranking. This is how the previous example would be transformed into an item using relative ranking:

People look for different things in a job. Below is a list of six factors in job satisfaction. Please mark the items in the order of their importance to you, starting with 1 as the most important factor in job satisfaction, through 6 for the least important factor.

_____ Work that pays well

_____ Work that gives a sense of accomplishment

_____ Work where you can advance

_____ Pleasant working environment

_____ Steady place of work with minimal chances of being laid off

_____ Work that provides good social benefits

Using relative ranking has its own disadvantages. Experience proves time and again that the order in which the factors are presented in the list affects how the respondent ranks the factors. That is, items that appear at the top of the list in the survey item tend to be ranked highest. In fact, the longer the list, the higher the odds of bias stemming from the order of the list. Respondents become tired or too burdened to go over the entire list, and thus focus on the first few factors that seem to be of relevance to them, without looking further down the list. Accordingly, it is recommended not to use lists longer than three or four items. It is also best to combine relative ranking and absolute ranking into one question (this will be easier in electronic survey formats such as Web-based questionnaires that enjoy higher level of interactivity).

Whatever ranking system you chose, make sure the instructions provided at the beginning of the question are clear, short, and easy to follow. If the ranking system is too cumbersome and onerous to respondents, they may not comply with the instructions and provide low-quality responses, or they may skip the item altogether.

7.7.1 Are All Possible Responses Included?

If you recall our earlier discussion on the importance of concepts and constructs, this is where constructs are crucial. Good constructs will direct your measurements and allow you to identify the scope of responses that are relevant to your study and in particular to specific items in your survey. However, before you proceed with phrasing a question, you should do a thorough literature review and immerse yourself in all the existing information related to your course of inquiry. If no previous material is available, make sure that you have done thorough field work, so that you are able to present all possible scenarios and/or response options. Keep in mind that closed-ended questions must be both exhaustive and mutually exclusive, meaning they need to present all possible and relevant options while also excluding those that are less relevant, lest they create confusion. If in doubt, leave an option that allows the respondent to add information, typically phrased as "Other_____".

7.7.2 Simple Items, Complex Concepts

Simple and to the point should be the rule that guides you in phrasing a good questionnaire item. Often, though, things that are clear to you as a researcher and specialist in a given field may not be so clear to potential participants (Converse and Presser 1986). Thus, when phrasing questions, it is wise to avoid jargon and professional concepts that may confuse respondents or even alienate them, leading to non-response. Terms and concepts such as *recidivism*, *stratification*, *variance*, and *gentrification* are best avoided and replaced with simpler concepts that are more commonly used by people outside the academic circles. There is nothing smart, elegant, heroic, or impressive about using high-flown vocabulary words that most of your respondents are not fluent in. The following survey questions are examples of overlying complex phrasing:

> Would you say that conceptualizing the idea of economic independency in early adulthood is a convoluted one?
> Would you say that explaining the idea of economic independency (responsibility) for young adults is a complex task?

Honestly, most people will not even understand what these questions mean, nor will it be clear to them what the researcher wants to find out. Words such as "conceptualizing" and "convoluted" are enough of a challenge. Throw in the concept of interest "economic independency," and you have a participant who will do one of the following: (1) skip the question; (2) answer it incorrectly; or (3) request an explanation before answering, "Don't know/no opinion."

But it is entirely possible to ask about the same topics in somewhat less complex form:

> Would you say that explaining the idea of economic independency for young adults is a complex task?

While the above question is a bit less frightening, it may still be difficult for some to grasp, and

thus may need some additional explanation. A good strategy for dealing with complex concepts is to gradually introduce respondents to them by using preparatory questions; such questions can also gauge how well participants grasp the main concepts and ideas. For example, one item can ask:

> Are you familiar with the concept of "economic independency"?

A positive response to this question can lead to another question that will verify that the respondent actually understands the concept:

Which of the below best describes the concept of "economic independency"?

Other question formats of a similar nature may be used to achieve the same goal.

7.7.3 Using Triangulation to Measure Complex Concepts

While repetition is not desired, it is at times essential, particularly when the research topic examines some complex concepts. As a form of beneficial repetition, triangulating survey items is at times desirable, as well as important for establishing reliability and consistency of the questionnaire as a tool. *Triangulation* means that the researcher may present several items that all examine the same concept or idea, but each item is constructed differently: one may be an open-ended question, another may be multiple choice, and another might use ranking. Each method uses a different approach to data collection. For example, open-ended items rely on a qualitative approach that seeks to gain insight, while closed-ended items are better suited for quantitative analysis, where quick analysis is desired and the scope of potential responses is known. Using both formats to examine the same concept will show respondent's consistency and thus provide inter-measurement reliability. For instance, this triangulation technique can assure that there is no response bias due to the order in which the responses are offered—or, if respondents' results are inconsistent, triangulation can expose this bias. When using triangulation, it is wise to present the open-ended item well before the closed-ended item, so as to avoid "mind blockage," in which respondents are affected by their response to earlier items.

Step 4: Empirical Examination in a Pilot Study

After identifying the items to be included in the questionnaire, you may want to consult your colleagues, who will most likely be happy to criticize your work—this is academia, after all. But all joking aside, this is an important preliminary step you must take before the survey is piloted; it will give you a good critical sense of things you may have overlooked, and may also open your mind to some things you never thought to ask.

After all your questionnaire items are in place, it is essential to examine them in context and to document participants' reactions to specific items and to the entire questionnaire. This can be done by selecting a sample of participants from the general population of interest to take the survey in its intended mode (e.g., face to face, by phone, by mail). In addition to the planned questionnaire, participants in the pilot study will also receive questions about the questionnaire itself, asking whether specific items were more difficult than others or if certain items took longer to answer. If respondents answer in the affirmative, this usually indicates that a specific item is problematic and requires reexamination. Researchers can also use this opportunity to identify items for which no answer was given and items that may have been differently constructed. When conducting the pilot examination of a questionnaire, researchers are encouraged to include prompt questions such as:

> Is there anything else you feel is relevant to this study?
> Where there any specific items in the questionnaire that you found difficult to answer? If yes, please identify them and explain why.

These questions will guide the researcher toward adding, deleting, or reconstructing some of the items to cover topics more thoroughly and clearly and to make the questionnaire more reliable and valid.

When examining responses to the pilot study, researchers should also take a closer look at

those closed-ended items for which participants answered "Other." This suggests that additional information may be needed. If respondents consistently filled in the blank following "Other" with a particular response, this may be deemed worthy of including it as one of response options.

Running a pilot test of the questionnaire is also important to detect the more mundane errors such as typos, grammar mistakes, jumbled question order, and numbering, unnecessary repetitiveness. When researchers are immersed in writing the questions, they are often blinded by their own cognition and overlook certain details. It is also important to keep a careful log to document all responses and reactions to the survey, including the external social or political climate or other situations that may have affected the survey.

Finally, pilot testing should provide you with some indication of the reliability and validity of your items and the entire questionnaire as it corresponds to the declared aims of the study. After data from the pilot test is coded, researchers can analyze response patterns and examine triangulated items on the same topic to see how their answers correlate. Depending on the sample size of your pilot test, you may even be able to run a reliability analysis on some of the items you wish to construct as a scale. Once you see the data and corresponding numbers, a more educated decision can be made regarding what items should stay and what items are redundant and can be deleted. Alternatively, if repetitive and triangulated items are found to have low reliability, you must reexamine their construct to figure out how they can be revised to provide a more stable measure. It is also possible that you will discover no reliability, as a result of participants not paying attention to specific items, or due to the fact that the survey is too long and onerous.

Step 5: Correct and Rephrase Items according to Findings from Previous Stage

Once you have results and logs from the previous step, you should turn to revising problematic questionnaire items. Pay special attention to their wording, order of appearance, and relevance to the entire questionnaire and topic. It is not uncommon for some items to be deleted from the questionnaire while new ones are introduced. It is also possible that you will need to rephrase specific items or place the items in a different order. If many changes are made, the questionnaire should be tested again—that is, Step 4 should be repeated.

Testing the questionnaire before full dissemination also helps you to phrase the questionnaire introduction. After the pilot stage, the researcher can better estimate the time it will take for the questionnaire to be completed. The pilot study also provides the researcher with a much better frame and context, as it is assumed that researchers can use the results of their pilot to tap on contextual, environmental, and social impediments (e.g., such as the recent event of occupying Wall Street in the US, social unrest, intense political climate, etc.). Finally, after the pilot phase, the researcher should have a slightly better understanding of the population of interest and how to motivate it.

Step 6: Writing the Survey Introduction

Many first-time surveyors do not invest much effort in devising an introduction—or worse, they skip it entirely. The survey introduction is extremely important, as it helps generate interest in the survey topic and promotes confidence in the researcher and the research institution. The introduction also provides essential details on the study, which can help the survey comply with general ethical regulations, and specifically those set down by the Institutional Review Board (IRB).

Surveys introductions must include the global aim of the study, although note that providing too much information can create a potential bias in response as a result of social desirability. Another vital part of the introduction is information about who is behind the survey—for instance, a research institute, a university, or a marketing company. This will usually boost potential participants' trust in the study and its legitimacy and thus increase participation, particularly if a well-known research institution or university is involved. The introduction should also mention the length of time estimated for

completing the survey, as well as some simple instructions. This will help potential respondents judge whether they are able to commit to the study. Introductions should also include contact information, in case any questions arise, whether about the survey in general or in regard to specific survey items. These are usually requested by IRBs, as discussed earlier by Oldendick in Chap. 3.

Finally, it is important that the introduction be short and clearly worded, so as not to turn off potential participants by placing an undue cognitive burden on them. This is particularly true in self-administered modes such as mail, e-mail, or Web-based surveys, where a researcher is not present to encourage the participant past the introduction. In academic research that uses direct interaction, such as face-to-face or phone interviews, the first few words and sentences of the introduction should introduce the research institution and clearly state, "This is not an attempt to sell anything." A good introduction will help to recruit and maintain participants, while also promoting high-quality responses. Please refer to Chap. 6 by Stalans for a thorough discussion of the importance of introduction and framing in order to reduce total survey error.

Step 7: Final Adjustments and Modifications

Finally, the questionnaire looks like it is ready for final dissemination. The questionnaire has been tested and revised, and the introduction is in place. It seems like the tool is good to go, right? Not yet! Keep in mind that once the questionnaire is out and the survey is running, you will not be able to stop, adjust, or modify anything. Consequently, it is essential to fine-tune the questionnaire before you begin the major data collection phase.

You should look for simple things, such as typos, awkward, or incorrect grammar, and items that begin on one page and end on the next. Make sure all potential responses in a closed-ended item appear together and immediately after the question itself, all on the same page. Check your pilot responses to see if any participants made small corrections or comments on specific questions. If your mode of administration is via e-mail or the Web, make sure the font you have written the survey in can be easily transferred (see Mesch Chap. 18; Manzo and Burke, Chap. 19; and Aviram, Chap. 26).

If the questionnaire will be administered by phone or face-to-face, make sure you train the interviewers and run a final "dress rehearsal" interviewing session during which you measure the time the survey takes; make note of any mispronounced words, and assess how clearly items are conveyed during the interview. Finally, do not forget to thank your participants for their time and insights. Politeness goes a long way.

As noted before, once the questionnaire is out, that is it! You will not be able to change anything, so just sit back, relax, and wait for the data to pour in. If you followed the steps in this chapter carefully, you should be able to enjoy high-quality data that is reliable and valid. These will translate to minimal errors and bias.

7.8 Summary

Constructing a reliable and valid questionnaire is not as easy as it may at first seem. Alas, nothing is easy in research, particularly when you aim to conduct a thorough and meaningful study, one that will actually advance knowledge. However, if you follow this chapter's step-by-step recommendations, you will have a much easier and even enjoyable experience when it comes to analyzing the data and writing up the results, the fruits of your hard labor.

Table 7.2 presents a summary of the steps and questions that you should be asking at each step of questionnaire development. Note that everything discussed in this chapter is arranged in this table in five distinct steps, to make your work week a bit shorter—and who does not want that?

Before proceeding to the next step, make sure that you completed the previous step satisfactorily. If you had to leave any of the questions in the above table unanswered, this should be a good sign for you to go back and repeat the previous step, and make sure that your concepts

Table 7.2 Summary of steps for constructing questionnaires

Step 1: Review Relevant Literature and Previously Tested Tools
If a questionnaire that fits your needs already exists and is tested for reliability, then there is no point in reinventing the wheel. If you proceed with writing your own survey, ask yourself the following questions:
• Who is your target population?
• Do you fully understand the issues and concepts to be examined?
• What are the main variables of interest?
• What do you want to learn about the above variables from your target population?
• How will the questionnaire be administered?
Keep in mind that certain questions will not be appropriate for self-administered modes, while others are not appropriate for face-to-face and phone interviews.
Step 2: Write the Items to be Included in your Questionnaire
Think about the following:
• Have you clearly identified the concept and construct of the variables of interest?
• Have you examined other relevant and related questionnaires?
• Have you consulted others to make sure your items are clear?
• Are the questions simply presented?
• Do the questions cover the scope of issues intended?
• Are items too complex and convoluted and potentially confusing to potential participants?
• Do any items use double negatives?
Are any questions leading or loaded?
Step 3: Design the Layout and Overall Questionnaire
• Does the questionnaire have a title, clear introduction, and section directions?
• Do the title and introduction promote interest in the research?
• Do the title and introduction promote participation and completion?
• Are items logically arranged?
• Are directions in each section clear and easy to follow?
Step 4: Empirical Examination and Pilot Testing
Try the questionnaire on a sample from the target population. In the process, note:
• How long does it take participants to complete the questionnaire?
• What items are difficult for them to answer?
• Are there items that were not clear to the participants?
• What items are left unanswered?
• Is there any information that pilot group participants added?
• Are items reliable and valid?
• Do the reliability and validity of the data provide evidence that the questionnaire examines the trait intended by the research question and variables?
Step 5: Finalize and Administer the Questionnaire
• Take a final look at your creation to make sure it is suitable for dissemination in the population of interest.
• Make sure there are no typos, grammar, or items out of place, and that instructions are clear.

For a similar flowchart of how to construct questionnaires, refer to Johnson and Christensen (2010, p. 195)

and constructs are in order and that you have met the requirements before moving on.

References

Converse, J. M., & Presser, S. (1986). Survey questions: Handcrafting the standardized questionnaire. *Sage Series: Quantitative Application in the Social Sciences, 63*(7), 9–79.

Gideon, L. (2012). Theories of research methodology: Readings in methods (pp. 27–42). Dubuque, IA: Kendall Hunt.

Johnson, L. B., & Christensen, L. B. (2010). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Thousand Oaks, CA: Sage Publishing.

Lin, N. (1986). Conceptualizing social support. In N. Lin, A. Dean, & W. Edsel (Eds.), *Social Support, Life Events, and Depression* (pp. 17–30). Orlando, FL: Academic Press.

Maruna, S., & King, A. (2009). Once a criminal, always a criminal?: 'Redeemability' and the psychology of punitive public attitudes. *European Journal of Crime Policy Research, 15*, 7–24.

Recommended Websites:

Creative Research Systems: How to design your own survey: http://www.surveysystem.com/sdesign.htm.

SurveyBuilder: Create Unlimited Online Surveys: https://www.surveybuilder.com/?utm_source=google&utm_medium=cpc&utm_term=create%20survey&utm_campaign=Survey+Builder+Free.

How to Write a Good Survey: http://www.accesscable.net/~infopoll/tips.htm.

Research Methods Knowledge Base: Constructing the Survey: http://www.socialresearchmethods.net/kb/survwrit.php.

Elizabeth Martin: Survey Questionnaire Construction, U.S. Census Bureau, December 2006 (PDF): http://www.census.gov/srd/papers/pdf/rsm2006-13.pdf.

Survey Design Tutorial: http://www.statpac.com/surveys/.

Science Buddies: Designing a Survey: http://www.sciencebuddies.org/science-fair-projects/project_ideas/Soc_survey.shtml.

Interviewing

8

Lior Gideon and Peter Moskos

The focus of the previous chapter was to acquaint readers with the recommended steps of designing a survey questionnaire. A good questionnaire and survey items are essential for successful data collection, and, as discussed before, for valid analysis. However many times questionnaires and survey items will be a product of preliminary interviews or will be used to obtain information using direct interaction between the interviewer/researcher and the participants (e.g., face-to-face or telephone modes of survey administration). In such cases, it is important to understand the advantages and disadvantages of interviewing, its many forms, and the validity of each interviewing technique to the topic being examined. We begin our discussion by explaining what we mean by "interview" and then turn to explore the many topics that can be researched using interviewing techniques. Finally we proceed to a discussion of the types of interviews and their advantages and limitations. However before doing so, we find it necessary to place interviewing in context of the Total Survey Error (TSE).

L. Gideon (✉) · P. Moskos
Department of Law, Police Science and CJA,
John Jay College of Criminal Justice,
899 Tenth Avenue, New York,
NY 10019, USA
e-mail: lgideon@jjay.cuny.edu

P. Moskos
e-mail: pmoskos@jjay.cuny.edu

8.1 Interviews and Their Place in the TSE

Interviewing may sound like a simple straightforward task. However, one should not mistake it for casual conversation. Research interviews require preparation and cannot be conducted by everybody who wishes to do them. In fact, different research topics will pose specific requirements for who may be a suitable person to conduct the interview. For example, interviewing rape victims may not be an appropriate task for males. This example raises the important issue of interviewer bias and quality of response that may result in lower survey reliability. Going back to the discussion of the complexity of interviewing, researchers must remember that response errors, which are part of the total survey error, can be a direct result of who is asking the questions. Accordingly, it is important to train interviewers, and prepare them for specific survey items. It is also potentially crucial that interviewers will speak the language and dialect of the population to be interviewed as will be discussed further in this handbook by Stoop and Harrison (Chap. 15) and Arsovska (Chap. 23). Another important factor that needs to be taken into consideration is the personality of the interviewer—cold or warm personality—as well as their appearance. Since interviews rely on direct interaction between two people, the nature and experience of that

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_8, © Springer Science+Business Media New York 2012

interaction have, at times, enormous effect on the responses provided.

8.2 What are Interviews?

Interviews are one of the more common methods of data collections that use direct interaction between researcher and subjects. In their most basic form interviews are defined "…as a conversation with a purpose" (Berg 2007, p. 89). Interviews are usually a direct face-to-face questioning that aims to achieve as much in-depth information as possible. The benefits of interviews are twofold:

1. To gain more accurate understanding on previously observed behavior;
2. To achieve information on specific topics that are not, or cannot, be observed. This will usually refer to more secretive topics, such as backstage police activity, to which access is difficult to gain.

By questioning others, the researchers emphasize the weakness of their role in the examined society.

8.3 Interview Topics in Social Science

The information that can be gathered from interviews in the field of social science is rich and diverse. Depending on the aims of the study, researchers can interview people on their political affiliation and support, consumer behavior, attitudes, actions, past experience, interaction with law enforcement, victimization experience, fear of crime, perception of support, effectiveness of correctional institutions, community care givers—the list can go on and on. At times it seems there is no limit as to which topics can be covered by an interview. However, most social scientists engaged in social research use this method to gain insight on specific aspects of the behavior they are interested in. Drawing from the field of the authors—criminology and deviant behavior—the current chapter will make use of examples from the criminological and social deviance discipline. For example Becker (1963), explored the culture of jazz clubs and how jazz players and people who attend the clubs are socialized to smoke marijuana. Becker first observed the behavior but then went further by interviewing individuals, using nonstructured informal interviews (a concept that will be discussed below). Becker determined how they became marijuana users, how they learn their smoking techniques, how they become a part of a group, and when and why users become labeled as deviant. Other studies may be interested in learning about a crime-fighting technique and thus would be interested in interviewing police chiefs and police-training specialists.

Interview topics normally correspond with the level of topic sensitivity and how easy or difficult it may be to identify and reach out to potential interviewees. Some criminologists are interesting in interviewing those engaged in illegal behavior to gain insight into the nature of their crime and criminality (see Bennett 1981). Interviews also enable offenders to explain their motives and lifestyles from their own perspective (Cops and Hochstetler 2006). Conclusions based on interview data are important to students, professionals, and scholars interesting in understanding, conceptualizing, and developing theories and methods to prevent and deter future crimes. Relying on in-depth information that is gathered by interviews provides researcher with the necessary insights needed. In addition, interviews in conjunction with observations may also prompt new theories, referred to as *grounded theories*.

No matter what the interview topic is, researchers who use interviews as their main data collection method must engage in thorough preparation before entering the scene to conduct interviews. Such preparation is crucial in helping the researcher to identify potential individuals who will be used as *informants* and help the researcher gain preliminary ideas and important information about the topic and individuals to be interviewed. Such preparation can also assist the researcher in identifying which individuals should be interviewed and what questions should and should not be asked. Depending on the topic and aim, researchers should also

choose the style of interview they will use. As explained below, different types of interviewing methods will be appropriate to different types of topics.

8.4 Types of Interview

Interviews are classified according to their level of formality and how structured they are. The interviewer can openly approach subjects asking to interview them in a formal setting or suggest to them that the intention is to engage them in a conversation in which the interviewer can learn about the research subjects' experiences and thoughts. Many times, highly formal and structured interviews will use "interview surveys" as their guide to the interview. Nevertheless, it is important to note that not all surveys are interview and accordingly not all interviews use surveys. This is mainly due to the fact that some interviews will be less formal, as the interviewer uses regular "small talk" conversation as the platform to weave in questions of more substance that are of interest and importance to the study. Formal survey-based interviews and casual conversation (informal interviews) are two opposite ends of a continuum. Most researchers engaged in interviewing as their main method of data collection will use a combination of methods found on that continuum.

Since the formal interview is highly structured and relies on participants' willingness to go along with the interview, it can be a highly useful method to obtain wealth of information in an organized manner from many different individuals. Such interviews are often structured to be long and in-depth and rely on previous knowledge gained from observations, reading of previous research, documents that the researcher was exposed to, and so on. In-depth interviews enable researchers to test their research hypotheses. Highly structured and formal interviews are many times referred to *surveys*, in which the researcher gathers systematic information that can be easily and quantitatively analyzed. In fact, many researchers who use survey interview enter the information directly to their laptops, a procedure that not only saves time, but allows them to have a quick description of how different question distribute across their sample. As you already know, there is no honey without the sting, and highly structured interviews have their downside.

One of the main disadvantages of highly structured interviews lies within the nature of the interaction between the interviewer and the interviewee. Interviews are not egalitarian conversations and tend to be very unidirectional. The interviewer presents the questions and the interviewees answer the questions they are presented with. Some say this is a very paternalistic approach as the interviewer holds the power and direct the conversation, whereas the interviewee is powerless in terms of the control they have on what questions are asked and when to talk. However, in adherence with *Institutional Review Board* protocols of research ethics, interviewees are allowed to refuse to answer specific questions they do not feel comfortable with, or they can stop their interview at any given time. In any case, such situation of unbalanced conversation is not natural and it is highly likely that responses provided by the interviewee will be affected by the situation and status of the interviewer and even the setting of the interview location. Consequently, some of the information may be biased due to *social desirability*, and the validity of the results may become questionable.

In part due to these potential problems, researchers may decide to use a less formally structured interview, one that will resemble a regular conversation. Using less formal and less structured methods, the researcher poses less stress on the interviewee, who may feel more an equal participant in the process. During the conversation many topics can emerge, some of which have more relevancy to the study and other which may not be relevant at all. Many times interviewing researchers may be confronted with topics that had not previously occurred to them as important, but surface through informal conversations. This process can be one of the greatest advantages of the less structured interview vis-à-vis the highly formal and structured interview. For example, Gideon

(2007) interviewing recovering substance abusing inmates and had a set of structured questions he wanted to ask. After a few initial interviews, Gideon decided to neglect the structured format and conduct the interviews in a more informal manner. In the less formal and less structured setting, interviewees divulged a wealth of information about their interactions with their spouse before their recent incarceration. The main knowledge gained was the fact that returning directly home from prison without any marriage counseling may account for relapse to substances and further criminal involvement. At times it is better to keep an open mind when it comes to data collection.

Once in the field, researchers should make an attempt to gather as much information as they possibly can, and less formal interviews are one of the ways to go about this. Of course this too comes with a price tag. Many times interviewees in less formal and structured interviews tend to talk about things that are of interest only to them, and as a result they may divert the focus of the researcher from the research goal. Researchers can also be sidetracked by idle conversation that serves little purpose. Trained interviewers may avoid such scenarios by directing the conversation back to desirable venues, although they will not always be successful. Their success depends many times on the individuals they are dealing with and the context in which the interview takes place. For example, interviewing active substance abusers can be a very challenging endeavor. The same goes for interviewing the elderly who may simply desire attention and conversation with others.

The type of interview a researcher chooses depends on several factors. The first concerns the level of accuracy of responses. If the interviewer is seeking to learn about specific costs of different substances and what substances are used, a more formal interview could be in order. However, if the researcher is seeking to understand drug markets and the relationships between traffickers, dealers, and buyers, a less formal interview will have to take place. Another important factor is the potential for *social desirability* bias, which means that interviewees are likely to change their behavior in response to the interaction between them and the researcher. For example males interviewing other males about their sexual activity may find different results than when the same sample of men is interviewed by a female interviewer. To promote higher reliability, and depending on the sensitivity of the interview topic, the researcher may wish to consider using a less formal method to interact with the interviewee. The hope is that a less formal interview allows the interviewer to gain more trust, allowing those being interviewed to "open up."

Environmental factors may also help to account for variation in interviewees' responses (Childers and Skinner 1996). How private or public the interview is can affect the results, in particular when the topic is sensitive or controversial. Interviews in public, where other individuals are present, may affect interviewees' responses and the interview process. Many interviewers will try to overcome such problem by conducting follow-up informal private interviews; which at times may even be disguised as regular small-talk conversation.

When Gideon interviewed incarcerated inmates, each interview was affected by the conditions allowed by the facilities. For example, some facilities allowed the interview to take place in the social worker's office without any supervision while others where in the cells or in the presence of a guard. When the interviews were conducted in the presence of others, it was clear to Gideon that the inmates were acting out and thus not providing accurate information related to the research question. Those interviews were dropped from the analysis due to their biased nature (see Soham et al. 2006).

Berg (2007) identifies another type of interview, in between the formal and informal interviews, which he refers to as *semi-standardized interview* (equivalent to *semi-structured interview*) in which a number of predetermined questions and special topics are to be covered during the interview. This interview process leaves room for other topics to emerge as there is more flexibility for the interviewer to move back and forward across different items. In such semi-structured interviews the interviewer is not locked

to a specific question or topic order and can change the order of the questions according to the interview development and relevancy of the conversation. Semi-structured interview is appropriate in cases where the researcher wants to be cautious and receptive to information that may not be identified in the initial planning stages of structuring the interview. Other times, semi-structured interviews stem from the interview itself, where the researcher receives very dull information, or unelaborated information such as "yes/no" responses. Using a semi-structure approach the interviewer can prompt the interviewee to provide more detail asking additional clarifying questions such as, "I understand this was very difficult for you, can you please explain what you did afterwards…?" Being able to provide feedback to interviewees, and in particular during high sensitive studies such as those that interview rape victims, prostitutes, and illegal workers, it is not only beneficial to the information gathering process but provides interviewees with a feeling, hopefully correct, that the interviewer is genuinely interested in their stories and cares about their experiences. This in turn can translate to higher quality information, one that may not have been gathered using a full structured formal interview with surveys.

Depending on the type of interview and directness of the interviewer, various technological aids may be either an advantage or disadvantage. The formality of and how structured an interview is can also be a product of the technologies used. For example, a computerized survey will indicate a highly structured formal interview. A voice recorder without any other aids may be perceived as more formal than a simple pen and paper. An open conversation without any accompanied aids gives the feel of an informal and nonstructured interview and could be less threatening to the interviewee. Regardless, more and more studies that rely on interviews use voice recorders as they can provide major advantages to the documentation process and allow the interviewer to pay more attention to the conversations and responses. But (as always) there are downsides:

1. Transcription and coding is an extremely time consuming, laborious, and potentially expensive process.
2. The interview location may be noisy, making audio recording of limited use.
3. Machines break, or may not be turned on. Relying solely on an electronic or mechanical device is risky. Most researchers get just one chance to interview specific individuals. If the researcher relies solely on the voice record without taking notes, he or she may be in for the surprise of their life, as Gideon was unfortunate enough to discover after an hour and a half of intensive interview with one of the prison wardens. At the end of the interview it turned out that the tape-recorder had a malfunction and the tape got caught in the device. Luckily the warden was nice enough to go over major points of the conversation to make sure right and accurate documentation was available (Gideon et al. 2010).
4. People being interviewed may simply not want to risk having their opinions recorded, as they may regret saying some of the things they say. Sometimes interviewing subjects will wish for an item to be "off the record," that is told to the interviewer but not cited to them. The concept of "on" or "off the record" comes from journalism and doesn't really apply to social science research since all interviews are technically "off the record." Rarely if ever are real names and identifying characteristics used in the writing and publication of such interview data. Regardless, the mere presence of a recorder can give the perception that the interview is being "on the record," causing those being interviewed to censure themselves.

It should be made very clear—clearer than the researcher may think necessary—to those being interviewed the confidential nature of the work, how the researcher will safeguard such information, the subject of the research (i.e., usually the subject is a general topic and not the individual person per se), and that the interviewee has control of what is recorded. Such information should be explained clearly to the person being

interviewed as simply signed an "informed consent" form may protect the researcher in some limited legal way (which is not at all the purpose of informed consent) but does little to guarantee an understanding of the issues involved. Clarity in such matters not only is essential to researchers' ethical and professional obligation, it can gain trust and yield much better data.

One technique, used by Gideon, was to give the voice recording device to the inmates to hold. Gideon then explained how it worked and showed the person being interviewed how they could pause, turn off, and even eject the tape if they wished to redraw from the interview (Soham et al. 2006). After this step was taken, no concerns were expressed about the device and interviewees spoke freely about their experiences (which, as was noted, did not happen when the same people were interviewed in public areas).

When Moskos recorded an interview with a fellow trainee in the Baltimore Police Academy, he found the data to be of limited usefulness. The officer being interviewed used carefully phrased responses and had a generally on-guard tone that did not reveal any major insights. Moskos noted that when they know they are being recorded, police officers sound like the officers in the TV shows "Cops" who speak with a stilted manner and use (not always correctly) fancy legalistic words. Given the poor quality of the data in the recorded interviews, Moskos quickly abandoned the tape recorder and relied on pen and paper to take notes and generally used the most informal interview method, casual conversation. Covert recordings are rarely if ever a viable option as not only are there serious ethical issues, such practices are illegal in many states (including Moskos's research site of Maryland). Besides, in immersion-type participant observation, it is not realistic to have a recorder running at all times. There is simply too much data.

Moskos's (2008) time in the field, 20 months, is on the longer side of most participant-observation research. Such a long time-frame removed pressure to gather relevant data quickly. More sensitive topics, such as racial attitudes, could be broached when circumstances made bringing up such subjects more appropriate. There was never a point at which "interviews" formally began and ended. In being fully immersed in the police world, both on and off duty, Moskos, because he was a police officer while he conducted his research, never had to worry about access to police. On the flipside, ease of access may come at the expensive of objectivity. In the trade-off between immersion and objectivity, Moskos clearly valued the knowledge gained from immersion as more significant than any benefits from maintaining a more traditional objective research position.

Officers were generally very open and unguarded with their opinions, as they would be in any private conversation. Interesting, revealing, typical, and even sensational snippets from conservation would be written down after the fact. One advantage of such a method—relying solely on notes taken at the scene and filling in details as soon as possible—is that the researcher by default culls the vast majority of superfluous data right at the start. But when one takes notes only selectively, the researcher must presume (and hope) that missing information does not later turn out to be significant. Anything not written down is quickly forgotten and thus, in essence, never happened. Another risk of using selective quotes to illustrate points is that it demands a certain faith from the reader in trusting the researcher's ability to parse sociological and criminological significance from late-night or drunken conversations.

While Moskos was completely open about his status as researcher, the interviews themselves could be considered somewhat covert, in that they never had a clear beginning or end. Data was taken from general conversations. Though these conversations were not recorded, additional ethical considerations are raised because the researcher's figurative tape recorder was always running. The researcher has an obligation, both professionally and personally, to protect those being studied. Even the issues of willing participation can be somewhat clouded if the researcher is observing people at work. Moskos conducted both front- and back-stage research (i.e., on and off duty). While fellow

officers did not need to associate with Moskos socially, on duty it was not possible for officers not to respond to a call simply because Moskos was present.

To protect his fellow police officers—some of whom Moskos says became close friends—quotes are provided without detailed descriptive of the speaker. Nor does Moskos's book include "characters" in the traditional literary or ethnographic sense. The end result is that those familiar with the officers involved may be able to attribute specific quotes to specific officers, but one cannot attribute all quotes to a single officer. Further complicating matters is that his research site and even his coworkers are a matter of public record. Moskos was not concerned about academics or the public knowing (or caring) who said what, and he made no attempt to disguise his research site by giving Baltimore some bland pseudonym. But he was concerned about unforeseeable career harm coming to officers from *within* the police department. Since there was no recording of any conversation or interview, those quoted would always have plausible deniability. If confronted, they could simply deny they were the source. To further protect research subjects, Moskos showed those in his book the completed manuscript before publication and asked if there were any objections or mistakes. There were none. Moskos found that those quoted in his book had little objection to portrayals that were not always positive. Even police officers, not generally considered an at-risk group, can feel misunderstood and quite powerless in the glare of publicity. Officers simply wanted themselves and their working conditions to be presented honestly and in context. Many of the officers later thanked Moskos for, in effect, serving as their voice.

8.5 Types of Questions to Ask

The range of questions that can be asked during an interview is extremely wide. Questions can be simple as "how old are you?" to more complex and sensitive questions such as "have you ever engaged in sexual intercourse in exchange for drugs?" The questions asked can also be very complex in their structure and may range from very simple *multiple choice* questions to highly demanding questions that seeks in-depth description and information. Questions can be fully structured, semi-structured, and not structured. You probably can imagine that there is a connection between the type of interview and the type of questions asked. *Structured questions* are planned by the researcher ahead of time, clearly phrased, and even tested prior to the interview. A written list of the exact questions to be asked and in what order is called an "*interview* schedule." Such preparation can be useful in gaining access to an organization (and IRB approval). But one can always deviate from a list of questions and adopt different methods as to reflect new understandings or a change in field conditions.

Usually we distinguish between two types of structured questions: (1) *Open-ended questions*, questions to which the interviewer does not provide a set of potential responses; (2) *Multiple-choice questions*, question in which the interviewer is provided with few relevant alternatives as optional responses. As always, both methods raise certain concerns. Open-ended questions may attract irrelevant responses that may sidetrack the interviewer. Multiple-choice questions may block the respondent's memory or divert their attention from more accurate response that were not presented by the interviewer. Also significant is the order of potential responses in multiple choice questions, which may affect the answers, picked by the responded regardless of what he or she may really think or believe. But the main disadvantage of the structured-question format is found in the influence the interviewer has on the interviewee, as well as in the level of response validity. Not being able to express one's true thoughts because of the limitations of the question format can be very frustrating to those being interviewed. The use of structured questions reduces the potential of receiving new and surprising information. For the questions are limited to the researcher's previous knowledge and preparations.

In normal everyday conversations structured questions are not natural. So researchers seeking to use them in a study must prepare thoroughly

before they use such a method in a field interview. This preparation requires that the researcher already be familiarized with the group being examined, including its language and terminology as well as potential and appropriate questions and answers. Using wrong terminology or even wrong dialect can result in biased results. However once the items are carefully constructed and the interviews are completed, data can be analyzed relatively easy. This is even more so if the questions used are multiple choice questions.

When using structured open-ended questions, the interviewer must be conscious to the meaning of the words used. It is possible that simple concepts that mean one thing in the researcher's own culture and society mean different things in the interviewee's culture. For example, "do you ever *feel chilled* when you hear about domestic violence cases?" While the researcher may use the word "chilled" to describe a feeling of being horrified by a certain incidents, the term could easily be used to describe something else to those who understand "chill" as "relaxing" or "taking things easy." Other times slang can become an interfering factor in the interview process if the researcher is unaware of certain vocabulary used by inmates for example. This may place the researcher in the awkward situation of asking what a word or phrase means. At best, which is not necessarily bad, the researcher looks clueless. Worse would be that the researcher is assumed to be a fool and perhaps subject to ridicule. Still worse, the researcher may miss the actual meaning of what is said. Worst of all would be actively misinterpreting what the person means.

For example if a researcher asks inmates if they would like to have *fish* as part of their regular meals; the word *fish* means a newly arrived inmate, and thus can be understood differently from the original intention of the researchers. Another interesting example can be the use of the verb "schooled," which means in prison lingo that the inmate is highly knowledgeable in the ways of prison life, and is a source of reference to other inmates. Moskos (2008) points out that even police lingo can mean very different things in different places. Among Baltimore police, to "jack somebody up" means to frisk somebody aggressively on the street, but among New York City police, "jacking a person up" implies to an extra-legal beating. Moskos learned this only when he casually used the phrase with New York City police officers and saw the look of shock on their face. The researcher need not (indeed should not) present him or herself as being as smart or cool or knowledgeable as the person being interviewed. The very purpose of conducting an interview is to learn from the person being interviewed. No matter your education and experience, always assume those being interviewed know something you do not. That is why you are interviewing them in the first place—expect to be surprised and if you don not understand something, ask.

While asking interviewees for clarification is a natural part of the interviewing process, researcher should not go out to the field and interview without preparation and some sense of what to expect. Researchers need to be serious, credible, and respect the time and sincerity of those being interviewed. Asking questions that reflect a complete lack of understanding of those being interviewed serves no purpose. Even worse, it can put the interviewer in a position where interviewees lose respect, consider the interview a waste of time, and may even feed the researcher nonsense answers. Planning ahead of time with structured questions enables the researcher to ask appropriate questions, phrase the questions in such a manner that analysis of the responses will be relatively quick, and enable the researcher to test the study's hypotheses.

Different from the structured questions, *unstructured questions* are not planed ahead of time, and their phrasing is not as rigid. As a result, their flexibility becomes one of their most valuable assets as they can prompt new and other questions. Researchers can react to information provided from a previously presented question by asking other questions that are of interest. Take for example the question, "Do you think any of the officers in your district will be promoted in the near future?" Responses to that

question can prompt other related questions such as: "Who do you think will be promoted?" and "Why do you think this officer will be promoted and not another officer?" It can also lead to more revealing answers about an interviewee's chances of being promoted and why she may believe she is being held back.

Unstructured questions in an informal interview that include spontaneous questions require the interviewer to have a good knowledge and functional grasp of the language and culture of those being interviewed. This is needed to help divert and direct the conversation in the desired direction. One of the main advantages of unstructured questions is that researchers can present them almost at any time and place without advanced planning. They also do not require the researcher to have a previously thorough knowledge in the researched topic, and are very similar in nature to the exploratory designs. The researcher can enjoy the flexibility of the design and react to ongoing events in the field. Another advantage is that unstructured questions and interviews allow interviewees to speak more freely and to present their opinions, positions, and perceptions while at times providing the interviewing researcher with unique and valuable information, one that was not known to the researcher before. This is why many researchers who seek to understand a culture use this method, as did Einat (2005) in an attempt to explore the inmates' subculture through their argot. Einat spent weeks with incarcerated convicted offenders in their cells and yard. Encouraging the inmates to speak freely and fluently without any disturbance from him, the interviewer, Einat attempted to learn their culture and to understand the importance of their argot (language) in their culture: why and when it is used, by whom, and against whom? Using this method, Einat was able to identify six main categories of words: violence and use of force, sexual behavior, loyalty and squealing, the prison and staff, drugs, and others. Einat learned that these categories reflect the needs, interests, and stress experienced by the inmates. Language connected all incarcerated offenders in the amalgam of prison life allowing them to identify with each other and achieve social cohesiveness.

A disadvantage of unstructured interviews is that reliable quantitative analysis of results is not possible unless one asks the same question in the same way to different people. For instance, Moskos analyzed differences in white and black police officers toward policing. Since his interviews were unstructured, he was able to draw general conclusion about differences in attitudes from his interviews, but was unable, based on these interviews alone, to demonstrate statistically significant quantitative differences. To achieve this end, Moskos supplemented his interview data with questionnaire data. Not only can questionnaires provide some hard numbers to support qualitative data, they can help researchers confirm the reliability of their own qualitative methods.

8.6 Summary

Deciding what type of interview to use and what aids can be brought to the interview depends on many factors: the sensitivity of the topic, how familiar the interviewer is with the topic, the type and level of accuracy that is expected to be received from the interview, the nature of the interviewees, and the connection and interaction the researcher established with the individuals to be interviewed.

Using technological aids depends mainly on the researcher. It is up to the individual researcher to evaluate if he or she can rely on pen and paper (and memory) and/or how technical measures will affect the interviewees and the data gathered. Using technology has its advantages but may also present the researcher with some unforeseen challenges that must be taken in consideration prior to the interview.

Corresponding with the different types of interviewing techniques are questions types. Questions can be fully structured, semi-structured, or unstructured. Each has its own advantages and disadvantages that are appropriate to different types of research and research topics.

Fully structured questions provide the researcher with data that are quick accurate and easily ready for analysis. But sensitive and difficult topics, including most criminal justice and criminological research, require more flexibility from the interviewer and thus encourage the use of semi-structured and unstructured questions rather than fully structured interviews.

References

Becker, H. S. (1963). *Outsiders: Studies in the sociology of deviance*. New York, NY: Free Press.

Bennett, J. (1981). *Oral history and delinquency: The rhetoric of criminology*. Chicago, IL: Chicago University Press.

Berg, B. L. (2007). *Qualitative research methods for the social sciences*. Boston, MA: Pearson/Allyn and Bacon.

Childers, T. L., & Skinner, S. J. (1996). Toward a conceptualization of mail surveys response behavior. *Psychology and Marketing, 13*(2), 185–209.

Copes, H., & Hochstetler, A. (2006). Why i'll talk: Offenders' motives for participating in qualitative research. In P. Cromwell (Ed.), *In their own words* (pp. 19–28). Los-Angeles, CA: Roxbury Publishing Company.

Einat, T. (2005). *Confined language: The life and words among prison walls*. Tel-Aviv: Schocken Publishing House.

Gideon, L. (2007). Family role in the reintegration process of recovering drug addicts: A qualitative review of Israeli offenders. *International Journal of Offender Therapy and Comparative Criminology, 51*(2), 212–226.

Gideon, L., Shoam, E., & Weisburd, D. L. (2010). Changing prison to therapeutic milieu: Evidence from the Sharon prison. *The Prison Journal, 90*(2), 179–202.

Moskos, P. (2008). *Cop in the hood: My year policing Baltimore's eastern district*. Princeton, NJ: Princeton University Press.

Soham, E., Gideon, L., Weisburd, D., & Vilner, Y. (2006). When 'more' of a program is not necessary better: Drug prevention in the Sharon prison. *Israeli Law Review Journal, 39*(1), 1–23.

Part III

Response and Non-response Errors

Unit Non-Response Due to Refusal

9

Ineke Stoop

9.1 Non-Response and Survey Quality

High response rates are often considered to be the outstanding quality characteristic of surveys. According to Biemer and Lyberg (2003), non-response is a quality feature that many survey users and sponsors have heard of, and the response rate is often seen as indicative of the competence of a survey organisation. Survey researchers know by now that non-response rates are not linearly related to non-response bias (Groves 2006). They are also aware of the fact that devoting limited resources to increasing response rates with little or no impact on total survey error is not money well spent (Merkle and Edelman 2002). High response rates after excluding 'difficult' groups (minority language speakers, mobile only, far away regions) will not result in a good representation of the target population. High response rates on a poorly drafted questionnaire will not contribute to valid conclusions.

As non-response rates are increasing (De Leeuw and De Heer 2002; Couper and De Leeuw 2003) survey costs and fieldwork duration increase too, and surveys are increasingly hard to plan. Also, non-response is a key threat to survey quality when survey participation is related to the outcome variables of a survey. The higher the non-response and the more respondents differ from the non-respondents, the larger the non-response bias and the lower the validity of survey outcomes. This is both true for non-contact and for refusal.

This chapter focuses on refusal, both because the major part of non-response is usually due to refusal and because refusal is possibly related to the topic of the survey. If this is so, and if surveys on, say, political involvement are completed mainly by those who are interested in politics, the validity of survey results is indeed threatened. The chapter starts with a short introduction on non-response bias and on what constitutes a refusal. This is not as clear-cut as one might think. Refusal in mail, e-mail and Internet surveys can usually not be measured, and not opening the door to an interviewer may be an implicit refusal.

Section 9.3 explores the issue of why people cooperate; Sect. 9.4 looks at survey design issues that can influence cooperation and in particular interviewers; and Sect. 9.5 examines the decision to cooperate. These sections partly mirror Groves and Couper's (1998) model on survey cooperation, but also include other points of view. Section 9.6 raises the question of the extent to which refusal conversion can help to enhance response rates and minimise bias. It also shows how follow-up surveys or doorstep questionnaires can help to provide information about core questions of the survey. Section 9.7 returns to the major problem: non-response bias. Section 9.8 briefly summarises the results

I. Stoop (✉)
The Netherlands Institute for Social Research/SCP,
16164, 2500 BD, The Hague, The Netherlands
e-mail: i.stoop@scp.nl

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_9, © Springer Science+Business Media New York 2012

presented thus far and attempts to give some guidance in handling refusals.

9.2 Why is Refusal to Participate a Cause for Concern?

"The basis for unbiased inference from relatively small observed samples to largely unobserved populations is probability sampling" (Singer 2006, p. 637). The assumptions underlying probability sampling are violated when increasing numbers of sample members cannot be contacted, cannot participate or are unwilling to participate. This is why non-response is one of the major threats to survey quality.

Non-response can induce non-response bias in survey estimates, although empirical results have shown that higher non-response rates do not necessarily cause higher non-response bias (Groves 2006). Non-response bias will occur when the likelihood of responding is correlated with survey topics of interest. When people who travel a lot are not contacted in a mobility survey, people who are ill cannot participate in a health survey and people who do a lot of voluntary work are more likely to answer survey questions on social cohesion (and on any other topic), bias can be expected. As the larger share of survey non-response is usually due to refusal rather than non-contact or inability to participate, it is important to minimise non-response due to refusal whenever possible, to assess and analyse the relationship between survey outcomes and reasons for refusal and to try to adjust for non-response due to refusal. Evidence from the European Social Survey (Stoop et al. 2010, Matsuo et al. 2010a), a cross-national face-to-face survey that aims at high response rates and low noncontact rates, shows that refusal is a key cause for concern (see Fig. 9.1).

9.2.1 What is Non-Response Bias?

Bias occurs when respondents differ from non-respondents, or when survey participation is correlated with survey variables (see Groves 2006). Acknowledging that response rates vary greatly across different surveys and topics, it may be assumed that everyone has an unobservable 'propensity' (a probability or likelihood) of being a respondent or a non-respondent. This propensity is indicated by ρ_i. According to Bethlehem (2002) bias can be formulated as follows:

$$\text{Bias}(\overline{y}_r) \cong \frac{\sigma_{y\rho}}{\overline{\rho}}$$

where $\sigma_{y\rho}$ is the population covariance between the survey variable y and the response propensity ρ, and $\overline{\rho}$ is the mean propensity in the target population over sample realisations given the sample design, and over recruitment realisations given a recruitment protocol design.

Put differently, bias is to be expected when factors determining survey participation are correlated with the survey topics. This would mean that respondents differ from non-respondents. When response propensity and survey outcomes are correlated, bias will be larger when the non-response rate is larger.

9.2.2 What is a Refusal?

As refusal is usually a major cause of survey non-response, and as refusal may be related to the topic of the survey, it is important to pay special attention to non-response due to refusal. But what exactly is refusal? The American Association for Public Opinion Research (AAPOR) (2011a) states that refusal occurs when a housing unit or the respondent refuses to be interviewed, or breaks off an interview. The possibility of a refusal clearly depends on the survey mode. The AAPOR therefore distinguishes between in-person household surveys (often also called face-to-face surveys), RDD (random digit dialling) telephone surveys, mail surveys and Internet surveys (these modes are further discussed in Part 4 of this handbook). In face-to-face surveys a sample member can be a named individual who—once contacted—can be asked to participate by the interviewer. In other cases the sampling frame comprises households or addresses. In this case the interviewer has to select one of the household members, usually with the help of one household member who is willing to be an intermediary.

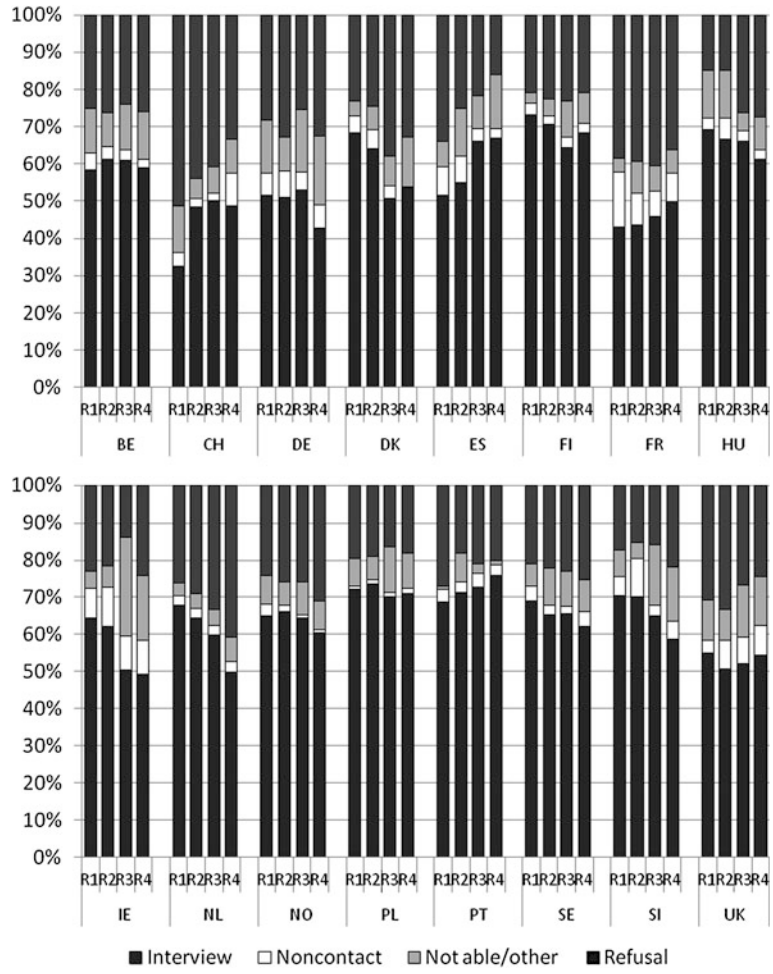


Fig. 9.1 Final disposition codes, European Social Survey Round 1 to Round 4 (only countries that participated in each round). Source: Matsuo et al. (2010a, p. 55)

When it has been assessed that the sample person belongs to the target population and is able to participate, he or she can either agree to answer the survey questions (and then complete the interview or break off) or refuse. However, what happens when a security guard or tenants' association does not grant access to the targeted housing unit? Or when a request for an interview is rejected by another household member? Or when an appointment has been made for an interview that does not result in an interview? Each of these non-interview situations is represented by a final disposition code. The AAPOR definitions also distinguish between soft, temporary and hard, permanent refusals, without clearly specifying what is meant by soft and hard.

The process in telephone surveys of named persons or mobile phone surveys is fairly similar. Someone has to pick up the 'phone and be willing to refer the interviewer to the target respondents. In RDD telephone surveys, a judgement will usually have to be made as to whether the household comprises a member of the target population and/or whether a sample member will have to be selected from eligible household members.

The AAPOR standard definitions further cover mail and Internet surveys of specifically named persons. This implies, for instance, that no response rates can be calculated for Internet surveys based on online volunteers or opt-in panels. Explicit refusal in mail and Internet

survey will rarely occur. In a mail survey it can happen that contact has been made with target respondents, they have been asked to complete a mail questionnaire and they have declined to do so. In addition, in mail surveys, entirely blank questionnaires are sometimes mailed back in the return envelope without any explanation as to why this has been done (AAPOR 2011a, p. 27). These cases will be the exception rather than the rule. Usually, it can only be assumed that not sending back a mail questionnaire is an implicit refusal. In other cases the questionnaire may not have been received, or may have been lost.

In Internet surveys, the likelihood of receiving an explicit refusal is even smaller. Explicit refusals can occur in Web surveys when the recipient replies to the e-mail invitation stating that he or she does not want to participate in the survey. Implicit refusals are those where a respondent visits the Internet survey URL and logs in with an ID and/or password, but fails to complete any of the survey items. Both types of refusals are likely to be rare (AAPOR 2011a, p. 35). In Internet surveys it is even more difficult to distinguish between different types of non-contact. Did the target respondent receive an invitational e-mail, was it classified as spam, did they read the invitation, did they simply forget to fill in the questionnaire or did they put it off until later and then forget about it?

In face-to-face and telephone surveys, the interviewer may be able to elicit some information about reasons for refusal. These are a potentially useful source of information for assessing the effect of bias (see Sect. 9.3.5). Neither in mail nor in Internet surveys is the researcher likely to be able to collect any information about the reason for refusal. Also, because these recruitment modes make it difficult to distinguish between non-contact and refusal, most research on non-response due to refusal focuses on face-to-face and telephone surveys. In practise, this chapter will focus on face-to-face surveys.

Even in these types of surveys it can be difficult to determine whether a sample unit is a non-contact, cannot cooperate or does not want to cooperate. When someone does not open the door this is a non-contact, but it could be an implicit refusal. The same holds for not picking up the phone when an unknown number is shown on the display. In addition, sample units can exaggerate language problems to avoid a downright refusal, or underreport the presence of members of the target population when the interviewer calls. These issues should be acknowledged when designing a survey. This chapter will however largely ignore these finer details.

9.3 Why Do People Cooperate?

Albaum and Smith (Chap. 11) discuss in length why some people agree to participate in surveys while others do not. Complementary to Albaum and Smith's chapter, Glaser (Chap. 12) focuses on the demographic profile of survey respondents. However, it is equally important to discuss the characteristics of those who cooperate and those who refuse, in order to gain a better understanding of non-response due to refusal.

9.3.1 Social Environment

A first factor in the analysis of cooperation is the social environment. Social environment could comprise general concepts like *survey culture*, more tangible issues like *survey scarcity* and more urbane issues such as neighbourhood characteristics. The *survey culture* of a country reflects a general attitude towards surveys and is hard to measure. One could say that the survey climate deteriorates as response rates go down. In this sense, survey culture is a tautological concept. Other reflections of the survey culture could be major events that have an impact on attitudes to surveys (such as confidentiality scandals), the economic and political climate and trust in institutions, the predominance of survey administration modes (face-to-face, telephone, mail, Internet) and the perceived legitimacy of surveys. Goyder et al. (2006) embed survey culture in general cultural factors. In their meta-analysis, cultural factors are one dimension of theories on non-response. These authors

approvingly quote Johnson et al. (2002, p. 55) "survey respondents are not merely autonomous information processors, rather, they exist within complex social matrices that influence their thoughts, feelings, and behaviours".

Survey scarcity is an issue at the individual level (see Sect. 9.3.3), but is also a characteristic of the environment. One reason that response rates were higher in the past could be that survey requests were scarce, whereas nowadays we are asked at least once a day to provide some kind of information. When surveys are a new phenomenon, high response rates are to be expected (Groves et al. 1992). Illustrative is the case of surveys among minority ethnic groups in the Netherlands, where response rates from recently immigrated groups were very high, and much higher than those of the native Dutch population (Stoop 2005). Nowadays, ethnic minority groups usually have much lower response rates, which could at least partly be due to the fact that participating in a survey is no longer perceived as an exciting new activity and a sign of being a valued member of Dutch society.

Results from the European Social Survey (Stoop et al. 2010) could be explained by the fact that response rates in new European Union (EU) member states are higher because social surveys are relatively new in these countries. Where social surveys are rare or new, people may either be pleasantly surprised that their opinions are considered important for research and policymaking, or they may be wary of unknown interviewers who seem to be prying. In former Communist countries, social surveys were scarce; in small countries like Switzerland and Luxembourg, the survey burden for individuals is likely to be much higher than in larger countries.

Earzllier studies (see Groves and Couper 1998) have identified urbanicity as a social environmental influence on survey cooperation. This could be because neighbourhood characteristics in urban areas are different, interviewing or interviewers are different or the population characteristics are different. Neighbourhood characteristics that could explain survey cooperation are population density, crime rates and social disorganisation. Other factors are multi-unit buildings in urban areas where the interviewer has to communicate via the intercom. It could also be that the fieldwork staff working in rural areas differ from those who mainly operate in large cities.

It may also be that the population composition in highly urban areas differs from that in rural areas (more single-person households, more ethnic minority groups) and that the neighbourhood effect reflects an individual effect (see e.g. Goyder et al.2002). One example of the difficulty of distinguishing between neighbourhood effects per se and composition effects is given in the study by Schräpler et al. (2010). They identified those neighbourhoods where the willingness to cooperate was highest. These were (1) high-income families in new, privately owned buildings; (2) older families in the outskirts; (3) self-employed people in new buildings; and (4) new, high quality detached houses (commuter belt). So do high-income families cooperate more, or people living in neighbourhoods where many high-income families live?

9.3.2 Who Cooperates More Often?

Analyses of survey response usually include standard background variables such as age, sex, family composition, socio-economic status, education, employment, race or ethnic group, and type of dwelling (see e.g. Stoop 2005). This is partly because these variables are often available at an aggregate level, from population statistics, or from the sampling frame. A comparison between population or sample and the realised interviews is then easily made. In practise, analyses of non-response based on these variables generally shed disappointingly little light on what kind of people refuse and why. There are several reasons for this.

First, until the beginning of this century many studies did not distinguish between contactability and cooperation (Lynn and Clarke 2001). Identifying willing survey participants generally obfuscates the fact that some may be easy to reach whereas others may be willing to

participate. Elderly people for instance may be easier to contact but less willing to cooperate than younger people. Age effects might thus cancel each other out. Second, practical reasons could cause lower survey cooperation by some groups. Lower cooperation by men in face-to-face surveys may be partly due to the fact that they are less often at home during the day. This may mean that they are less frequently approached directly by the interviewer, for whom it will be more difficult to tailor the recruitment effort. This could also result in lower survey cooperation.

Third, background characteristics are never causes of survey cooperation. At most they are correlates. When studying the relationship between age and refusal it is important to specify what 'age' stands for Stoop et al. (2010). Elderly people may have a lower educational level, be less adept at filling in complicated forms and less willing to let a stranger enter their home. Thus, they might refuse more often. They may also be less busy and have a higher sense of civic duty. This would result in higher survey cooperation. Younger people may be confronted with a wide range of stimuli that vie for their attention, hence less cooperation. On the other hand, younger people might be more curious about efforts to seek information from them and have more experience of standardised information and form-filling. Age could be a correlate of many different factors behind survey cooperation, which might explain why Groves and Couper (1998, p. 136) found few robust effects of age.

Educational level can also illustrate the complicated relationship between background characteristics and survey cooperation. Several studies have highlighted the relationship between educational level, form literacy and willingness to participate in a survey. Brehm (1993) suggests that interviews would be seen as tests and thus as more attractive by persons with a higher level of education. However, he found an overrepresentation of persons with a lower education level. Groves and Couper (1998, p. 128) hypothesised that the better educated have profited from earlier form-filling efforts and thus may be more prone to cooperate. They point to the consistent finding from the literature that the lower educated groups more often fail to participate in surveys, but find no such results in the governmental surveys they study. They also discuss the possible benefits of cooperating in a survey (p. 122), such as the enjoyment in thinking about new topics, and the possible costs, such as the cognitive burden incurred in answering and comprehending the survey questions. As the cognitive burden of cooperating in a survey might be less for the more highly educated, this might result in higher cooperation by the higher educated (see also Tourangeau and Smith1996; Holbrook et al. 2003). In the European Social Survey (Stoop et al. 2010) less well-educated groups were seriously underrepresented in eight out of 24 countries, and considerably overrepresented in four countries.

9.3.3 Reasons Behind Survey Participation

More helpful than looking at socio-demographic and socio-economic variables is trying to find out why people answer questions and participate in surveys. Basically, there are four—possibly overlapping—reasons why people participate in a survey. The first reason is that they enjoy survey participation, the second that some incentive makes participation attractive, the third that participation is important for themselves and the final one that it is important for someone else. Singer (2011) would call the first three egoistic and the fourth altruistic. She also adds another perspective: participation because of specific survey characteristics. For example: participation in a governmental survey because of a desire to help the government is altruistic, and "because it is a government survey" (ibid: 380) is related to a survey characteristic.

Survey enjoyment and survey importance can be measured directly. Several non-response researchers have measured the general attitude towards surveys, the attitude towards particular surveys and the impact of substantive survey characteristics such as topic and sponsor (Goyder 1986; Couper 1997; Campanelli et al. 1997;

Loosveldt and Storms 2008; Singer et al. 2003; Hox et al. 1995; Rogelberg et al. 2003).

So why would people like surveys? One reason is that some people like answering questions, that they enjoy giving their opinions or thinking about interesting topics (Couper and De Leeuw 2003; Loosveldt and Storms 2008), the enjoyment of thinking about interesting topics, gratification that one's opinion was sought. Political interest and knowledge may make a survey more interesting and thus increase cooperation. Lack of political interest and knowledge, on the other hand, may point to an isolated position in society in general, and thus decrease cooperation. According to Brehm (1993, p. 64) "… people who are most likely to participate in polls are also more likely to participate in society generally". Groves et al. (2004) found that political contributors showed a relatively high cooperation rate in surveys regardless of the survey topic. Rogelberg et al. (2001) found that willingness to participate in a future survey among students depended on survey enjoyment, measured by three questions 'I do not like filling out surveys', 'Surveys are fun to fill out' and 'I enjoy filling out surveys'.

If people do not like participating in surveys, external incentives can still make survey participation attractive. As handing out incentives to respondents is also a characteristic of surveys, we will return to the issue of incentives in Sect. 9.4.1. It may however be expected that monetary incentives are more effective for less well-off people. In this case the presence of an incentive is a reason for survey participation.

A third reason to participate in a survey is that people feel their opinions count. This is again the scarcity principle (Groves et al. 1992) that we discussed earlier at an aggregated level. This principle suggests that a letter emphasising the importance of individual participation, and stating that such participation cannot be substituted by another person, will be more effective at generating commitment to the survey. Many survey introductions and advance letters emphasise that the results of the survey could be used to improve health, living conditions and public services, thus highlighting the importance of participation by the respondent.

Finally, people may participate in a survey because it is important for someone else, e.g. society, science, a local community. When survey participation is induced by a sense of civic duty, the notion that social norms produce a sense of obligation to provide help or to grant the requests of others in the belief that this serves the common good. Those who score highly on civic duty will be members of community groups, volunteer for neighbourhood clean-ups, be politically active locally, assist in schools, participate in church outreach programmes, volunteer for youth athletics or social programmes and participate in surveys (see also Goyder1987).

Rogelberg et al. (2001) discuss the concept of survey value. Measured by three questions: 'a lot can be learned from information gathered from surveys', 'nothing good comes from completing a survey' and 'surveys are useful ways to gather information', Olson et al. (2011) use the social exchange perspective to predict that letters emphasising that the community will benefit from their participation will yield higher participation rates. Different types of reasons can of course be related. People might feel that something is more interesting when it is important. The satisfaction of being part of a socially useful or academically interesting enterprise indicates both survey interest and survey importance.

It also makes sense to assume that survey participation, because of its perceived importance to society, will be a factor behind the participation of socially involved and integrated people, and much less of the socially isolated. Social isolation (Groves and Couper 1998) may manifest itself in the absence of shared norms or the lack of guidance by the norms of the dominant culture. Loosveldt and Carton (2002) used direct measures of social involvement in explaining refusal in the second wave of a panel survey. Persons who scored highly on utilitarian individualism in the first wave, which means they were more interested in personal gain and less in societal well-being, were more likely to refuse participation.

Table 9.1 Reasons for survey participation, expected underrepresentation and overrepresentation of specific groups

	Over represented
Survey enjoyment	Higher cognitive skills Interest in politics Social persons (face-to-face)
Incentive	Less well-off
Importance self	Use public assistance Ventilate opinion
Importance society	Religious Socially integrated Voluntary work Interest in politics

Individual, general reasons for survey participation are summarised in Table 9.1. Whether these reasons play a role will depend on the characteristics of the survey (Sect. 9.4.1) and the interaction with an interviewer. For instance, lower participation in official surveys, or surveys seen as coming from vested interest organizations, could be expected from those who score low on civic duty, who are cynical about political institutions, and who lack trust in governmental organisations.

9.3.4 Reasons Not to Participate in a Survey

Reasons not to participate in a survey may be derived from reasons for refusal, but also from voiced doorstep concerns (Bates et al. 2008). They may to some extent mirror the reasons to participate. People may not like surveys in general, or the topic or sponsor of a particular survey. It could also be that they do not see the relevance of a survey for themselves, or lack the feeling of civic duty or social involvement that could be a reason for participation. Here, too, there will be an interaction between individual characteristics (being busy, being concerned about privacy), survey characteristics (Sect. 9.4) and interaction with the interviewer (Sect. 9.5).

Refusal may be caused by characteristics of the survey, e.g. interview length, frequency of survey requests, misuse of surveys as a sales pitch, the cognitive burden of the interview, embarrassment and inconvenience caused by survey cooperation, invasion of privacy and being treated irreverently. The first factor related to the costs of being interviewed is interview length. Bogen (1996) has studied the relationship between interview length and response rates and finds the evidence inconclusive. Dillman (2000, p. 305) suggests that a questionnaire may even be so short that it stops making sense to respondents. If interview length, or more generally the time burden of participating, were an issue, one would expect busy people to participate less than non-busy people. There is ample evidence, however, especially from the analysis of participation in time use studies, that busy people do not cooperate less often (Abraham et al. 2006; Van Ingen et al. 2009; Stoop 2007). This could be due partly to the fact that busyness is often related to having a paid job and having a family with (young) children. The social involvement of these groups could counteract time constraints.

The frequency of survey requests may result in survey fatigue, or survey saturation (Goyder 1986). The frequency of survey requests may also counteract scarcity arguments used by the interviewer in the recruitment phase (see below). Selling under the guise of a survey ('sugging', see Groves and Couper 1998) or using surveys as a sales pitch is something more and more people are becoming wary about. De Leeuw (2001) even suggested starting every telephone survey introduction with 'I am not selling anything'. In her experiments this led to a small but positive effect on response rates. Stocké and Langfeldt (2004), however, found that the misuse of survey interviews as a door-opener to sell goods and services had no effect on respondents' evaluation of surveys in general.

The cognitive burden of the survey, sensitive or difficult questions (Tourangeau and Smith 1996),

perceived invasion of privacy (Singer et al. 1993; 2003) or the feeling of being treated disrespectfully (Morton-Williams 1993, p. 105); Dillman 2000, pp. 15–16) may also be a reason for refusing cooperation. Dillman (2000, pp. 19–21) highlighted the importance of respect and trust and outlined ways to establish trust in order to overcome security concerns of potential respondents.

9.3.5 Recording Reasons for Refusal

In many interview surveys, the reasons that sample units give for unwillingness to participate are recorded. This information can help when refusal conversion is used to detect and adjust for non-response bias, for interviewer training purposes and for developing and improving recruitment strategies in other surveys. Table 9.2 gives an overview of the reasons for refusal recorded in Round 4 of the European Social Survey (Matsuo et al. 2010a).

Differences between countries in Table 9.2 will be due, at least partly, to differences in recording reasons for refusal. One example of questionable comparability is 'not interested': why would this be reported in 55% of Portuguese cases and never in Denmark? 'Not interested' is also difficult to interpret as a category: does it mean that the target person is not interested in the topic or in survey participation, or is it simply a polite way of phrasing refusal: "I am not interested".

Smith (1984, p. 486) commented on the difficulty of obtaining clear and meaningful reasons for refusal 'Review of detailed interviewer comments on final refusals indicates that reasons for refusing are not always clear and that some people seem to offer a string of reasons hoping that the interviewer will expect one and leave them alone', as did Brehm (1993, p. 51). Sceptical readers might wonder whether the refusals' reasons are just the first coherent thought that pops into their mind, and not the underlying cause of the decision. According to this thinking, the refusal might feel threatened by the interview, but tell the interviewer, 'I'm too busy.' Considering these cross-national differences, one could argue that the value of reasons for comparative research is limited. One particularly interesting outcome, incidentally, is the emphasis on privacy in Bulgaria, the Czech Republic, Russia, the Slovak Republic and Ukraine, all countries with a Communist past where academic social surveys are fairly new (and response rates fairly high).

Information about reason for refusal can be very important in the national context, for the next interviewer, analysis of bias and future survey design. Section 9.7.1 extensively discusses the possible links between types of reasons for refusal, e.g. situational or topic-related, future cooperation and non-response bias.

9.4 Survey Design and Survey Cooperation

9.4.1 General Design Issues

Several characteristics of the survey design can have an impact on the refusal rate. Relevant aspects are the sampling frame, the recruitment mode, the topic and sponsor of the survey, so-called house effects, advance letters, incentives and interviewer issues. Given the crucial importance of interviewers in face-to-face surveys they will be treated separately in the next section.

In the European Social Survey (Häder and Lynn 2007) different types of sampling frames are used: lists of individuals, households, and dwellings and random route procedures to select addresses and subsequently households and individuals. Individual sampling frames have the advantage that personalised advance letters can be sent to target respondents. The conventional wisdom (Morton-Williams 1993) is that this enhances participation, although Luiten (2011) came to opposite conclusions. An additional advantage of individual sampling frames is that the interviewer can focus directly on the target person and does not have to go through the intermediate step of respondent selection. Individual sample frames can sometimes also provide information that can be used to tailor the approach (age, sex, country of birth of target persons). Individual sampling frames become problematic if they are outdated or incorrect.

Table 9.2 Reasons for first refusal (more than one reason possible) at first refusal (%)

	Bad timing	Not interested	Subject unknown/ difficult	Waste of time	Waste of money	Privacy	Never do surveys	Cooperate too often	No trust surveys	Bad experience	Dislike subject	No approval	Not admit strangers	Other	*N*
Belgium	9.4	35.8	3.0	12.9	1.2	3.4	8.5	0.8	3.2	1.4	1.5	2.7	1.2	14.9	1,454
Bulgaria	7.8	26.1	1.4	17.2	0.4	13.1	7.1	1.1	8.8	0.7	0.2	5.1	5.8	5.3	832
Croatia	30.4	25.4	2.8	12.1	6.6	8.0	2.3	3.5	0.7	0.8	3.6	0.4	0.4	3.2	1,066
Cyprus	9.3	27.0	2.7	13.9	1.2	7.3	12.7	4.6	6.2	4.6	3.9	1.5	3.5	1.5	259
Czech Republic	5.8	40.0	1.4	11.3	1.0	12.5	9.2	0.4	6.8	1.6	1.2	2.6	3.5	2.6	914
Denmark	21.8	0.0	3.0	9.1	0.7	1.1	17.2	5.0	1.8	0.5	1.4	0.7	2.7	34.9	559
Estonia	10.0	38.6	1.6	14.5	0.7	7.2	4.9	1.3	6.4	1.0	1.8	2.3	1.7	8.1	832
Finland	17.9	35.1	3.6	10.1	0.5	2.9	5.8	1.6	1.5	0.7	2.1	1.5	0.8	15.8	999
France	3.8	52.2	1.1	16.4	0.2	2.4	7.1	0.7	2.7	0.4	0.7	2.3	3.0	7.0	1,614
Greece	4.8	39.4	6.7	11.8	4.0	10.1	8.5	3.0	2.4	1.0	2.2	1.9	3.3	0.9	672
Hungary	15.2	35.9	1.9	9.9	1.7	8.2	9.7	1.1	4.9	1.6	1.6	4.4	3.9	0.1	1,222
Ireland	13.9	37.7	2.1	4.3	3.6	1.0	8.8	1.1	1.8	0.1	1.2	2.1	1.7	20.7	1,006
Israel	13.2	35.2	4.6	7.7	1.0	11.9	7.5	0.6	4.7	1.0	2.0	1.8	1.6	7.3	506
Latvia	9.8	17.6	4.9	18.1	2.9	9.8	7.3	0.5	5.4	1.5	3.4	1.5	15.6	2.0	205
Netherlands	6.3	43.5	2.5	11.7	0.7	3.7	6.8	1.7	2.5	0.7	2.1	1.2	1.2	15.5	2,508
Norway	8.3	50.8	1.7	5.8	0.6	2.1	9.3	1.0	0.4	0.4	0.6	0.7	0.0	18.3	1,154
Poland	2.2	37.2	1.5	16.5	2.1	7.0	6.9	0.6	3.9	0.9	0.2	3.8	1.4	15.7	853
Portugal	2.5	55.2	2.4	14.5	2.3	4.7	9.3	0.0	4.5	1.1	0.9	2.1	0.6	0.0	796
Romania	10.3	36.4	10.9	8.7	6.0	2.2	8.2	1.6	1.6	3.3	0.5	2.2	4.4	3.8	184
Russia	10.1	21.8	2.2	17.5	1.1	14.4	11.2	1.0	5.6	1.0	1.3	0.0	10.8	2.1	1,101
Slovak Republic	16.8	32.7	0.9	9.9	2.0	13.0	7.0	0.0	4.9	2.0	0.9	4.6	2.4	3.0	660
Slovenia	24.8	34.0	3.2	11.7	0.2	5.8	7.7	1.1	2.1	0.6	0.6	2.1	0.6	5.6	532
Spain	10.8	45.5	2.2	14.1	1.1	3.4	7.9	0.4	3.6	0.4	2.6	1.0	0.9	6.4	1,243

(continued)

Table 9.2 (continued)

	Bad timing	Not interested	Subject unknown/ difficult	Waste of time	Waste of money	Privacy	Never do surveys	Cooperate too often	No trust surveys	Bad experience	Dislike subject	No approval	Not admit strangers	Other	*N*
Sweden	33.9	37.4	0.9	1.3	0.2	2.1	6.9	1.1	0.6	0.4	1.5	1.8	0.0	11.9	821
Switzerland	15.7	36.0	1.1	3.4	0.7	8.2	8.2	0.6	3.6	0.8	0.6	1.1	0.4	19.6	2,443
Turkey	2.3	51.4	1.2	4.9	0.0	8.5	7.4	0.5	5.2	2.1	2.0	1.6	12.0	1.0	611
Ukraine	4.4	20.1	2.8	22.6	1.2	11.9	7.2	0.3	8.1	0.6	2.5	3.9	9.0	5.4	890
United Kingdom	17.3	34.4	2.3	4.9	0.8	7.0	9.0	0.8	2.3	0.4	1.8	1.1	2.5	15.4	1,627

Taken from Matsuo et al. (2010a, p. 33). Germany and Lithuania were excluded because they used different categories for recording reasons for refusal. Multiple reasons for refusal could be recorded, but some countries recorded only one

This can result in many ineligible cases or in lots of cases that should be interviewed at their new address or who cannot be found because they left no forwarding address.

The type of sponsor and the type of survey organisation are generally assumed to have an impact on survey cooperation. Government and academic surveys achieve higher response rates, probably because they appeal more to a sense of civic duty and authority than commercial surveys (Dillman 2000, p. 20; Groves and Couper 1998, p. 139). Effects on response rates and survey outcomes stemming from the differences in procedures used by different fieldwork organisations are generally called 'house effects'. The presence of house effects is well-known in political polling; for example, see Martin et al. (2005).

In introducing surveys, it is considered good practise to emphasise that the topic of the survey is relevant to the interviewee, assuming that topic saliency is related to response behaviour (Groves and Couper 1998, pp. 286–288). A study by Groves et al. (2004) found that persons cooperated at higher rates in surveys on topics of likely interest to them. Political contributors cooperated with all the topics, however, which might point more to social involvement than to topic relevance as a determinant of survey participation.

Te Riele (2002) presented an inventory of topic-related response behaviour in surveys by Statistics Netherlands. Although she does not always distinguish between non-contact and non-cooperation, her results are very informative: people who are less mobile are less likely to participate in the mobility survey, converted refusals in the housing demand survey are less likely to want to move, persons with health problems are overrepresented in the health survey [a result which contradicts the findings by Cohen and Duffy (2002)], persons who are interested in politics are overrepresented in the national election studies, and—in general—there seems to be a relationship between social participation and survey participation.

Advance letters underscore the legitimacy of a survey, remove suspicion, communicate the value of the survey and evoke the principles of

social exchange and reciprocation, thereby positively influencing response (De Leeuw et al. 2007). Box 9.1 below contains an extract from the guidelines for response enhancement (and minimisation of refusal) in the European Social Survey (Koch et al. 2010).

Box 9.1: Guidelines for Response Enhancement

Guidelines on response enhancement in the European Social Survey (Koch et al. 2010)
Advance letters

A letter sent in advance of an interviewer call usually has a positive effect on the response rate. It can serve several purposes, addressing a variety of issues known to affect survey participation. The advance letter can be used to

- explain the purpose of the survey,
- identify the sponsor and the survey organisation,
- include or announce any gifts or incentives and provide information about them,
- alert the respondent, or household, to expect a call from an interviewer.

In most cases, interviewers value the use of an advance letter, as their first contact with the sample person or the sample household is then not totally unexpected.

If the sample frame is one of named individuals, the advance letter should be addressed personally to the selected individual. If using a sampling frame of addresses or households, the effect of an advance letter may be diluted, as the individual to be selected may not receive or read the letter.

Care should be taken to ensure that the time span between the arrival of the letter and the call of the interviewer is not too long. Sometimes the best way to do this is to instruct interviewers to send the letters in a way that matches their planned work pattern (rather than sending the letters centrally at the start of the fieldwork period).

There is a vast range of literature on survey incentives and how and why they work (e.g. Singer et al. 1999a, b; Singer 2002). A general finding (see e.g. Bethlehem et al. 2011, p. 74) is that prepaid incentives are more effective than promised incentives because obligation is a stronger motivational factor than inducement or 'bribery'. Most researchers seem to agree that a small incentive is generally seen as a nice gesture, both by interviewers and respondents, that a large incentive is not necessarily more effective than a small one, and that incentives do not have a detrimental effect on survey quality.

9.4.2 Interviewers

The recruitment mode may also determine the cooperation rate. It is generally assumed that refusing is easier by telephone than in a face-to-face survey. Holbrook et al. (2003) ascribe the higher response in face-to-face surveys partly to the reassuring effect emanating from having an interviewer on the doorstep but also observe that some respondents, especially women and the elderly, are reluctant to let a stranger into their home.

Interviewers are generally (and rightly) seen as the key agents of the researcher. They define the purpose of the survey to sampled individuals, provide them with a key rationale for participating in the survey, address their concerns and find convenient times for them to be interviewed (Durrant et al. 2010). In addition, they keep call records; collect observational data and conduct a number of other mainly administrative tasks (see e.g. Matsuo et al. 2010a). Survey interviewing is neither a well-paid nor a very high-status job. Nevertheless, the prevailing belief is that interviewer experience is a critical factor in securing cooperation (Groves and Couper 1998, pp. 200, 215).

In the past, a number of studies have tried to identify which interviewer characteristics, attitudes and behaviour are crucial in obtaining cooperation (Morton-Williams 1993; Campanelli et al. 1997; Carton 1999; De Leeuw 1999; Hox and De Leeuw 2002). These studies encountered a number of problems. First, it is hard to discover what happens on the doorstep. Second, interviewer effects are often hard to

distinguish from sample effects as fully interpenetrated designs are rare in face-to-face surveys. This means that some interviewers work in areas that usually have a lower response rate than other areas. In addition, experienced or successful interviewers may more often be assigned the difficult cases. And third, interviewer behaviour may have to be adapted to sample person characteristics (i.e. tailoring).

As the interviewer's task is comprehensive, complex and crucial for the success of a face-to-face survey, the importance of training of interviewer is generally emphasised. Box 9.2 below presents best practises with regard to interviewer training as drafted by the American Association for Public Opinion Research (2011b). A concise overview of rationales for and content of interviewer training is given by Loosveldt as part of the Round 6 Specification for Participating Countries (European Social Survey 2011). Detailed guidance on interviewer training may be found in Lesser et al. (2008), Loosveldt (2008) and Schaeffer et al. (2010).

Box 9.2: Best Practises (interviewer training) (AAPOR 2011b)

Train interviewers carefully on interviewing techniques and the subject matter of the survey.

Insisting on high standards in the recruiting and training of interviewers is also crucial to conducting a quality survey or poll. For high quality data to be collected, interviewers in telephone or in person surveys must be carefully trained to do their work properly through face-to-face ("classroom") or telephone training, self-study, or some combination of these. Good interviewer techniques should be stressed, such as how to make initial contacts, how to deal with reluctant respondents, how to conduct interviews in a professional manner, and how to avoid influencing or biasing responses. Training should also involve practice interviews to familiarize the interviewers with the variety of situations they are likely to encounter. Time should be spent going over survey concepts, definitions, and procedures, including a question-by-question approach to be sure that interviewers can deal with any misunderstandings that may arise.

Schaeffer et al. (2010) give a useful state-of-the-art overview of interviewers' effects on non-response error. With regard to observable characteristics (race, gender, age and voice) of interviewers on non-response she concludes that effects are small. Unobservable characteristics, such as experience (often related to age), personality, knowledge, attitudes and expectations are also mixed. Experienced interviewers seem better at overcoming negative survey attitudes and delaying tactics (too busy). Experience can however be confounded by selection effects (less successful interviewers drop out). It is also difficult to assess the level of experience, especially when interviewers work for several survey organisations. With regard to personality characteristics Schaeffer et al. (2010) confirm the conclusions of Groves and Couper (1998) and Hox and De Leeuw (2002), that "… interviewers who report more positive attitudes about persuasion strategies, express greater belief in confidentiality, ascribe to the importance of refusal conversion, and express 'willingness to proceed as usual in the face of obstacles' tend to obtain higher response rates" (Schaeffer et al. 2010, p. 445).

West and Olson (2010) tried to assess how much of interviewer variance is in fact non-response error variance, or in other words whether interviewers have an effect on the answers given by respondents or on the selection of respondents. They did this using information from the Wisconsin Divorce Study that comprised data for both respondents and non-respondents. Although they focused on overall non-response outcomes and not on refusal, one of their results is particularly interesting, namely that "… interviewers may be more successful at

recruiting respondents around the same age as themselves, consistent with liking theory (Durrant et al. 2010). Alternatively, certain interviewers may be more proficient with particular age groups (e.g. talking slowly for the elderly), regardless of the interviewers' age" (West and Olson 2010, p. 1022). These findings highlight the importance of training and tailoring (see Sect. 9.5.3).

Additionally, recent studies (Blom et al. 2010; Durrant et al. 2010; Jäckle et al. 2011) approached the relationship between interviewer characteristics and survey cooperation from a different angle: they linked results from interviewer questionnaires to response outcomes in a number of studies. Blom et al. (2010) analysed the impact of interviewers on non-response in the European Social Survey (ESS). Their analysis is based on an interviewer questionnaire (IQUEST) answered by 880 interviewers in eight countries during the first round of the ESS. Interviewers who were good at securing cooperation (after contact) were also good at making contact. However, it turned out that interviewer behaviour and attitudes explained variance in contact much better than in cooperation, possibly because of the strategy mentioned above of assigning difficult cases to more successful and experienced interviewers. Older interviewers achieved better cooperation, as did interviewers who could deal positively with reluctance and who tried to avoid and convert refusals on the doorstep by actively providing positive information.

Durrant et al. (2010) also used information from interviewer questionnaires linked to six major household surveys in the UK. First, they studied three types of interviewer influences on obtaining cooperation: (a) interviewer experience and pay grade; (b) interviewer confidence and attitudes; and (c) interviewer behaviour, strategies and approaches on the doorstep. After controlling for pay grade, they found a curvilinear effect of experience. "Skill level, reflected in pay grade, appears to be the real underlying mechanism driving cooperation rates, not the simple length of time employed" (Durrant et al. 2010, p. 19). Interviewer confidence and attitudes (b) play an important role. More persistent and confident interviewers receive fewer refusals. This self-confidence is also reflected in the support of successful interviewers for assigning unsuccessful cases to a different interviewer. Interestingly, interviewer confidence seems to be more important in complex surveys. Interviewer self-reported behaviour, strategies and doorstep approaches (c) had little effect on cooperation. The authors ascribe this to the need for tailoring (see Sect.9.5.3) which means that different strategies may be needed to persuade different types of respondents.

Second, Durrant et al. (2010) also investigated the effect of homogeneity between interviewers and target persons. Female householders were more likely to respond than men to female interviewers, whereas the converse was not the case. This could be because the women were afraid to let a male stranger into their house. The effect of homogeneity with regard to education was quite complicated. A lower refusal rate is related to having the same educational levels (no education, a low education or a college degree). When educational levels differ, refusal rates are higher. The highest refusal rates occurred in those cases where the interviewer had no education or only a low educational level and the householder a professional degree of some form. One could assume that in these cases heterogeneity between target person and interviewer reduces survey enjoyment, because the chances of having an interesting conversation are being perceived as smaller.

Jäckle et al. (2011), in common with Blom et al. (2010) and Durrant et al. (2010) linked information from 842 interviewers who answered an interviewer survey to the response outcomes of 28 different surveys. Interviewers provided data on traits, skills and attitudes and also on the 'Big Five' personality traits (extroversion, agreeableness, conscientiousness, neuroticism and openness to experience). In summary, their results confirmed that interviewer experience predicts success. Interviewer attitudes towards the legitimacy and usefulness of persuading reluctant respondents are weakly predictive of cooperation. Furthermore, assertive, extrovert and less open interviewers are more successful. The latter result

seems to contradict the findings by Durrant et al. (2010) presented above. Jäckle et al. (2011) ascribe some of the impact of experience to higher scores on personality traits, skills and attitudes of experienced interviewers. Higher response rates for female interviewers could also be due to women scoring higher on personality traits, skills and attitudes that are positively related to survey cooperation.

These new studies help create a better understanding of why some interviewers achieve more cooperation than others, and may help in recruiting and training interviewers.

9.5 Decision to Cooperate

9.5.1 Heuristics or Social Exchange

The decision to participate is often considered as a process within the social exchange framework, in which it is assumed that participating in a survey has certain disadvantages and certain benefits that need not be strictly material but may also be social (Brehm 1993; Dillman 2000). The net outcome should determine survey participation. Social exchange is based on the rational choice theory of decision-making, positing that an individual weighs the costs of participation in a survey against the perceived benefits, the cognitive burden and confidentiality risks (Groves and Couper 1998). However, there is not much evidence for a fully rational approach. Contrary to the rational choice approach, for instance, is the observation that the decision to cooperate can be taken within seconds (Sturgis and Campanelli 1998; Dijkstra and Smit 2002). In line with this, the decision to participate may depend on shortcuts to decision making, or cognitive heuristics (Goyder et al. 2006). Seen from this point of view, any factor, such as likeability of the interviewer or the sense that the interviewer is a scarce opportunity may aid the approach. In addition, behind the distinction between heuristics and rational choice lies the distinct view of target persons: either the behaviouristic approach towards sample units or the appreciation of respondents and refusals as responsible citizens (Stoop 2005; Goyder 1987).

Goyder et al. (2006) tried to integrate exchange and heuristic theories of survey non-response. They quote Brehm (1993), who argued almost 20 years ago that heuristics are best at explaining refusals while rational calculation of costs and benefits best accounts for participation. This could be a positive outcome, assuming that non-response bias can be ignored when it results from superficial or general factors rather than being related to the topic or purpose of the survey, for example.

9.5.2 Psychological Concepts

Groves et al. (1992) unravelled three social psychological concepts that could help explain why people decide to participate in a survey: compliance with request, helping tendencies and opinion change. Compliance with requests is partially related to the reasons for survey participation discussed earlier in this chapter, and partially to interviewer behaviour. Groves et al. (1992) distinguish between reciprocation (also related to social exchange), consistency, social validation, authority (and civic duty), scarcity and liking (as discussed previously in this chapter). Hox and De Leeuw (2002) studied among other things the use of scarcity arguments (*this* is the chance to give your opinion) and social validation arguments (most people participate, emphasise positive aspects of participation) across countries and found large differences in interviewer behaviour between countries. They also suggest, however, that social validation arguments might remind people too much of a sales pitch and thus might even be counterproductive. Blom et al. (2010, p. 14) found that social validation arguments did not increase the likelihood of cooperation, although surprisingly it did result in a higher contact rate.

As well as compliance, Groves et al. (1992) discuss helping tendencies and opinion change. Helping tendencies may partly explain why response in face-to-face surveys is usually higher than in other modes. “A wet and cold interviewer standing on the doorstep in winter may elicit greater compliance” (Groves et al. 1992, p. 484). Three emotional states are related

to helping decisions: anger (negative), happiness (positive) and sadness (negative for high cost-benefit ratio, positive for low cost-benefit ratio). These findings can have an impact on interviewer behaviour (retreat when householder looks angry) and non-response bias (overestimation of happiness in population). Opinion change is an important factor when a request to participate is perceived. When the topic of a communication (here survey request) has high personal relevance, people will carefully review the pros and cons of participation, whereas when it is of low personal importance they will decide on the basis of a heuristic, as discussed in the previous section.

9.5.3 Tailoring and Maintaining Interaction, Leverage-Saliency and Benefit-Costs

Earlier in this chapter it was observed that general interviewer behaviours may have little effect on survey cooperation because different target persons require a different approach. Appeals to civic duty, handing out cash incentives, highlighting the intrinsic interest of a survey or referring to legitimate sponsors may not be uniformly effective. Face-to-face approaches may inspire confidence in some respondents and put off others. Advance letters may be too complicated for people with a lower education level and too simple for the more highly educated. This suggests that flexible survey designs and fieldwork approaches could enhance response rates. This is usually called 'tailoring'. Maintaining interaction gives the interviewer more time to find and express the right arguments and may make it more difficult for the target person to refuse (consistency, see above).

Groves and McGonagle (2001) applied the constructs of tailoring and maintaining interaction to a training protocol design. In tailoring, the interviewer emphasises those survey characteristics that are salient to the respondent, persists when useful and withdraws when necessary. The longer interaction is maintained, the more cues the interviewer receives that can be used in tailoring, and the better the opportunity to select the most appropriate approach from the wider repertoire of persuasion techniques that good interviewers have built up. In maintaining the conversation, "… the expert interviewer does not maximise the likelihood of obtaining a 'yes' answer in any given context, but minimises the likelihood of a 'no' answer over repeated turn taking in the contact" (p. 251).

The leverage-saliency theory on survey participation (Groves et al. 2000; Groves et al. 2004) combines a social exchange approach with tailoring doorstep behaviour to respondent concerns. According to this theory different survey attributes, such as topic, type of research organisation and presence of a cash incentive, are weighed by respondents in their decision on whether or not to cooperate. The attributes have a different leverage for individual respondents and can be made more or less salient to the respondent prior to the participation decision. Interviewers implicitly use this theory when they emphasise, for instance, academic sponsorship to respondents with a higher education level or hand out a gift to less wealthy respondents. As mentioned above, the first test of the theory showed that the effects of incentives were smaller for those with high community involvement than for those with low community involvement. The second study, however, found no significant relationship between incentive and personal gain orientation. For further discussion on the role of incentives please refer to Chap. 13 by Toepoel.

Singer (2011) developed a benefit-cost theory of survey participation; this can be seen as a synthesis of principles from the leverage-saliency theory and the social exchange theory. The former emphasises the role of the interviewer in making positive factors salient and overcoming respondent objections.

> Benefit-cost theory, in contrast, emphasises the role of benefits in the decision-making process and the need for making these benefits salient to the respondent and achieving a favourable ratio of benefits to costs. By implication, it also calls attention to the need to broaden researchers' understanding of the intrinsic motives for taking part in surveys. Benefit-cost theory also resembles social exchange theory, which posits that the

> respondent must perceive some personal benefit in order to give something back to the researcher (e.g. Dillman 1978). But unlike benefit-cost theory, social exchange theory tends to downplay or overlook the inevitable costs that survey participation entails for the respondent. (Singer 2011, p. 388).

One major risk on which Singer (2011) focuses is confidentiality (see also Singer et al. 1993; Singer et al. 2003). Previous research showed that elaborate assurances of confidentiality were a cost to respondents, who expected greater risks and worse outcomes, but also that, for sensitive topics, confidentiality increased response rates. Further studies on which she reports show that (a) the survey topic significantly influences survey participation; (b) information about disclosure risks had no effect on participation but subjective perception of disclosure risk did have a significant effect; and (c) survey participation is mainly determined by the perception of benefits for self or others and not by perceptions of risk or harm. Therefore, trying to reduce the costs of responding will not induce target persons to cooperate. The primary motives are therefore positive. This also explains why she developed a benefit-cost theory rather than a cost-benefit theory.

Singer (2011, pp. 387–388) comes back to the issue of heuristics. Agreeing to participate may depend on whatever comes to or is brought to mind—from an advance letter to the interviewee to something one has just read in the newspaper—is worth the costs. The calculation of whether benefits outweigh the costs need not be carefully reasoned, but may be applied on the basis of heuristics. One of the reasons for the decline in response rates may be that this heuristic used to result in a positive default response decision (say yes, unless…) whereas nowadays the default is negative (say no, unless…).

9.5.4 Minimising Refusal

Given the foregoing, the guidelines for enhancing response rates can easily be derived. The first rule ought to be that respondents should be respected. This means they should receive the proper information, be interviewed by friendly and competent interviewers who will come back at a more suitable time if necessary and who keep their appointments, and should be asked well-tested questions that can be answered. Survey procedures should make survey cooperation as easy as possible and small incentives should be given in advance as a sign of appreciation. The benefits of survey participation should be highlighted (interesting topic, we would like to know YOUR opinion, this is important for society, science, your local community, etc.) and costs minimised. Good interviewers should tailor their introduction in order to present the most salient benefits to the proper respondents.

Despite these efforts, many people will refuse and, if Singer (2011) is right, non-participation might have become the default heuristic, which means that many people will only participate when the particular benefits of participating in a particular survey have been made very clear. So what can we do? First we can try to convert initial refusals, and second we can try to minimise the effect of refusal by adjusting for non-response bias.

9.6 After the Refusal

9.6.1 Refusal Conversion

One strategy to increase response rates is to attempt to convince potential respondents to cooperate after an initial refusal. This is called *refusal conversion*. Apart from leading to higher response rates per se, refusal conversion has also been used to study non-response bias (Groves and Couper 1998, p. 49) by treating converted respondents as proxies for final non-respondents (e.g. Smith 1984). This use of refusal conversion cases will be discussed in the next section on non-response bias. Here, we focus on the response enhancement consequences of refusal conversion and related issues. This section is largely based on Billiet et al. (2007) and Stoop et al. 2010, pp. 161–203). Billiet and Matsuo in

the following chapter (Chap. 10) discuss non-response and measurement error as result, and how one can control it.

The idea of reapproaching initial refusals in a renewed attempt to interview them might appear to be an impossible task. However, refusing to cooperate in a survey is not always seen as a permanent state. Someone may refuse to take part in a particular survey in certain circumstances, but in other circumstances the response may be positive (Loosveldt et al. 2004, p. 73). The person may be busy when the interviewer arrives, or may be feeling unwell or irritable, and therefore refuse. In many cases a later visit at a better time, or perhaps by a different interviewer, might generate a more positive response. This will be true especially if the decision to participate is not seen as the result of a great deal of cognitive effort in deciding whether or not to participate, but rather as a heuristic (see Groves et al. 1992 on opinion change).

Refusal conversion is probably the standard practise in telephone surveys or in Internet/postal surveys because repeat contact attempts to non-respondents are fairly easy and cheap (Triplett 2006). In face-to-face surveys, by contrast, refusal conversion is a relatively expensive technique for increasing response, and perhaps compares poorly with the use of respondent incentives or extended interviewer efforts to make contact (Burton et al. 2006, p. 459). Apart from the large additional survey costs related to refusal conversion attempts in face-to-face interviews in some countries, the process is complicated by privacy and data collection concerns.

If refusal conversion is allowed, different strategies can be applied. One option is that all refusals are reapproached. This could result in complaints from angry target persons who made it explicitly clear on the first occasion that they had no intention of ever taking part in the survey. Such an option is in any case likely to be expensive and time-consuming. A second, more common option is that specific types of refusals may receive another visit for instance, those refusals who live on the route of a second interviewer, or those who refused an interviewer who delivered low response rates in a particular survey. Yet another option is that interviewers are instructed only to revisit those sample units where the original interviewer feels there is an acceptable chance of them cooperating. This latter strategy is usually the cheapest and perhaps the most common, but may also result in a more imbalanced final sample, with increased rather than decreased error. Deploying a new interviewer is the customary practise in refusal conversion (Groves 1989, p. 218). Sometimes better trained, more experienced interviewers or supervisors are utilised. Sometimes a new interviewer is explicitly chosen because they are of a different sex, age or ethnic background from the first interviewer. Sometimes, a second approach may be effective simply because the target respondent is in a better mood, less busy, or likes the new interviewer better. Depending on the strategy followed, the meaning of being a converted respondent will differ across countries.

Studies of refusal conversion evaluate the success of this approach from different angles, e.g. the success rate in terms of additional respondents, the characteristics of the total sample before and after refusal conversion efforts, the impact of refusal conversion attempts on non-response bias by comparing survey estimates of subjective (attitudinal) variables with and without the converted refusals, factors that explain variation in the success of refusal conversion attempts and the relationship between refusal conversion and measurement error.

Refusal conversion generally increases response rates, but the percentage of refusals who can be converted and the percentage point increase in response rates varies widely. In the European Social Survey (Stoop et al. 2010), the success rate of refusal conversion differed dramatically across countries (see Fig. 9.2). This was partly a consequence of the initial refusal rate, partly of the percentage of initial refusals that were reapproached and partly of the selection of which refusals to reapproach. Some countries reissued almost all refusals, while others reissued none, with a range of 3–92%. The mean reissue rate of 30% therefore masks vast differences in the use of this technique. In the end, however, the impact of refusal

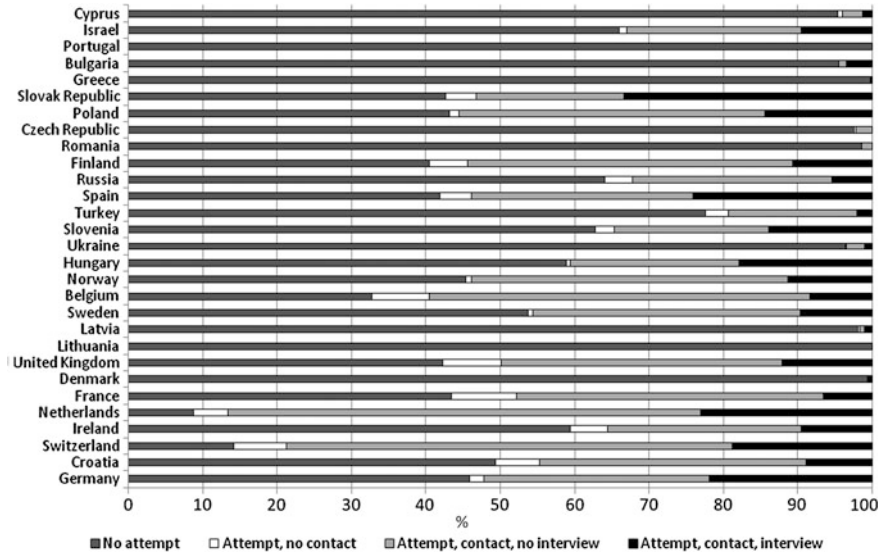


Fig. 9.2 Effects refusal conversion strategy in European Social Survey Round 4 (percentage of all initial refusals) Note: Based on Matsuo et al. (2010a, p. 65). Countries ordered according to final response rate (Germany: 42.7%, Cyprus: 81%)

conversion on final response rates is small in most ESS countries.

One key issue in refusal conversion that has been studied in some detail is the elapsed time between the initial refusal and the subsequent conversion attempt. Evidence from the US provides mixed evidence on the optimum waiting time before reapproaching initial refusals (Triplett 2002, p 26–27; Triplett et al. 2001; Edwards et al. 2004, p. 3440). Results from the ESS (Beullens et al. 2009) provide evidence that longer elapsed times are helpful in increasing the success of refusal conversion and provide further weight to the idea that in many instances refusal to participate in a survey is a time-dependent phenomenon. In addition, analysis of ESS contact form data clearly confirms that it is better to use a new and more experienced interviewer in the conversion attempt, since they have a higher likelihood of success.

As is often the case, previous research gives a mixed picture on the effect of refusal conversion on the representativeness of a survey. Brehm (1993) found that the effect of refusal conversion on the percentages of respondents who are married, working or male was both small and negligible when compared with the effect of increasing the number of calls, which had a far bigger impact. The only noticeable (though still small) effect of refusal conversion was on the age of the respondents; the mean age of the final sample obtained would have been a few months lower without refusal conversion, suggesting that disproportionately larger numbers of older respondents were converted (Brehm 1993, p. 162). In a study in which the socio-demographic characteristics of reluctant and cooperative respondents were compared, Stoop (2005) found that converting refusals actually worsened the sample structure according to the socio-demographic characteristics of the known population. In the *Index of Consumer Sentiment* survey of the period between 1979 and 1996 (Curtin et al. 2000), converted refusals were disproportionately respondents of lower socioeconomic status (whilst those who were more difficult to contact were disproportionately of higher socioeconomic status). As in the study by Stoop (2005), men were found to be less likely to be successfully converted after a refusal than women. Differences in the propensity to be converted were also found according to race; non-whites were less likely to end up being converted refusals.

With regard to measurement error, Cannell and Fowler (Cannel and Fowler 1963, p. 259–262)

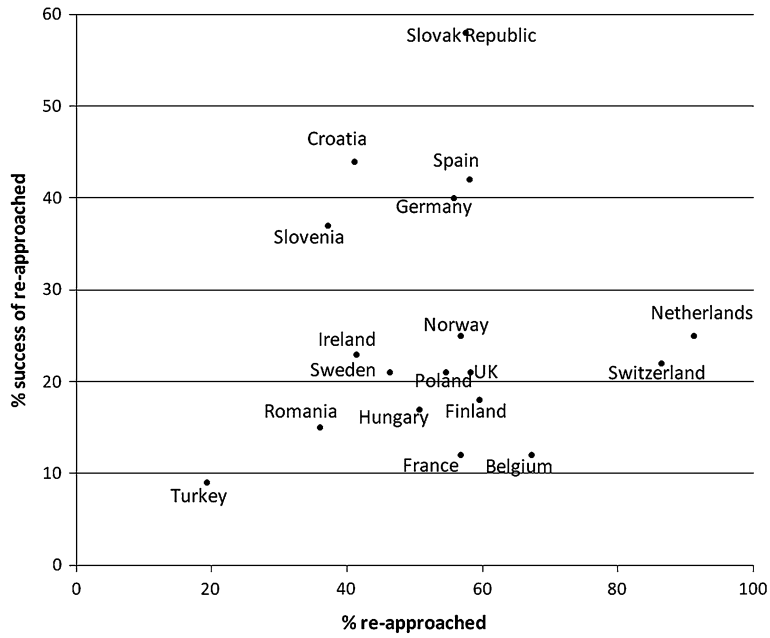


Fig. 9.3 Outcomes refusal conversion strategy in European Social Survey Round 4 Note: Based on Matsuo et al. (2010a, p. 66). Countries with fewer than 100 reapproached initial refusals are excluded

found in the early 1960s that reluctant respondents provided poorer quality data. They attributed this effect mainly to lower respondent motivation. The 'reluctant respondents' (late return) provides less accurate reports on past events than the early respondents. It was not possible to distinguish the effect of motivation and mode on accuracy among the respondents who did not respond to the second mailing in the self-enumerative condition, since these respondents were then approached by telephone or face-to-face.

Another decrease in data quality that might be expected from reluctant respondents is that they engage in *satisficing*, i.e. choosing the answer that requires least thought because they want to spend as little time on answering questions as possible (Krosnick et al. 1996). Stoop et al. (2010) gave an overview of evidence of satisficing behaviour from a number of studies. They also presented new evidence from the European Social Survey which showed virtually no evidence of satisficing (with the exception of giving 'don't know' answers, which by itself is not a convincing indicator of satisficing), and the data from more subjective indicators of response difficulty appear unreliable. The ESS results therefore suggests that converted refusals, at least those who are being converted at present, do not provide poorer data quality.

In summary, it may be concluded that there is clear evidence that refusal conversion can be used as a tool for increasing response rates and obtaining more cases during survey fieldwork. However, it is far from clear that this is the most effective way to achieve such an increase. The costs of refusal conversion in money and time are hard to assess. It will be clear that this strategy prolongs the duration of fieldwork, as initial refusals will have to be reassigned in a later stage of data collection. Refusal conversion will be expensive when hard or all refusals are reapproached and success rates are low. It will be inexpensive when only soft refusals are reapproached and success rates are high. Figure 9.3 shows the relationship between the percentage of initial refusals that were reapproached and the percentage of reapproached initial refusals who finally participated. From Fig. 9.2 it is clear that the Netherlands and Switzerland reapproached almost all initial refusals, with a moderate success rate but also a substantial number of converted refusals. The Slovak Republic has a very high success rate,

but reapproached slightly less than 60% of the initial refusals. Turkey, for its part, reapproached a small percentage with little success.

Data from the ESS suggest that refusal conversion—although it substantially increases response rates in a number of countries—makes little significant difference to the profile of the final achieved sample and that when it does there is variability in terms of whether it makes the samples more or less representative. The results from the ESS also show that refusal conversion is a strategy that is implemented very unevenly across countries.

It could be concluded that focusing on soft refusals to enhance response rates is probably not the best strategy, but that it is preferable to focus on hard refusals to minimise bias. Whether this is possible remains to be seen.

As noted earlier in this handbook, by Oldendick (Chap. 3) refusal conversion may also pose some ethical challenges which must be taken into consideration when attempting to convert refusals—soft or hard—as this may result in lower quality response as previously discussed.

9.6.2 Doorstep Questionnaires and Follow-Up Survey

Even someone who has refused and cannot be converted may be willing to answer a few survey questions on the doorstep or even to answer a (short) questionnaire later on. Two methods have been described in the literature, namely the Basic Question Approach (Bethlehem and Kersten 1985; Bethlehem 2009) and the Follow-up Survey among non-respondents (Hansen and Hurvitz 1946). Elliot (1991) compares the two methods. Both methods require a high response rate (Groves and Couper 1998) to minimise the possible effect of bias from non-response by the non-respondents.

Bethlehem and Kersten (1985) introduced the Basic Question Procedure, similar to the Pre-Emptive Doorstep Administration of Key Survey Items (PEDAKSI) method put forward by Lynn (2003). This boils down to putting a small number of basic or core or topical questions to all non-respondents. One reason that refusals may be willing to answer a small set of questions is because it is an example of the door in the face technique (see Mowen and Cialdini 1980; Hippler and Hippler 1986; Groves et al. 1992). This implies asking for a large favour (cooperation in a long interview) where there is a chance of having the door slammed in one's face, and if this request if rejected, asking for a smaller favour as a concession (cooperation in a short interview). This technique appears to be a good strategy for prompting refusers to at least give some information. Some drawbacks remain, however, notably that it may be difficult to decide which questions are the key survey items, especially in a multi-topic survey, and that single core questions when asked out of the context of the survey may measure something different from when the same question is posed as part of a battery of similar questions in a long questionnaire. Crucially, of course, some non-response remains.

A follow-up survey among non-respondents implies drawing a subsample from the non-respondents and asking them to answer the questionnaire. In a well-conducted survey it may be assumed that the non-respondents will be mainly refusals who could not be converted in the main survey. Obtaining their cooperation will require well-trained and highly motivated interviewers, and possibly larger incentives. In this way 70% of the most persistent refusers can be persuaded to cooperate after all. It has the drawbacks that the sample will generally be small (because of the costs of deploying the best interviewers, handing out larger incentives and other response-enhancing methods) and will usually involve some delay in completing the fieldwork.

Billiet et al. (2009), Matsuo et al. (2010b) and Stoop et al. (2010) present results from both approaches adopted in the European Social Survey. In Belgium, the doorstep procedure was used, in Norway, Poland and Switzerland a follow-up survey. The first approach yields data from refusers only, the second from both non-contacts and refusals (and respondents, for comparisons sake). As the non-contact rate is

usually low in the European Social Survey (see Fig. 9.1) former non-respondents who participate in the follow-up survey will generally have been former refusals. These efforts to obtain core information about respondents indicate that the response rates among former non-respondents is inversely related to the original response rates, that final non-respondents (who answered basic questions) differed from converted refusals and that the core information from refusers can be used to adjust for non-response bias.

9.7 Refusal and Bias

9.7.1 General and Particular Reasons for Non-Participation

Many researchers have tried to distinguish between general and particular reasons for survey refusal. Smith (1984, p. 481–485) distinguished between *propitiousness* and *inclination*. *Propitiousness* is a situational factor, determining whether the interview request is convenient at that particular occasion when the respondent has other activities (going out, napping, having a headache, having dinner, entertaining visitors). Coming back at a more suitable time or making an appointment to do this will then be a useful strategy. Independent of the level of propitiousness, respondents might differ in their *inclination* or willingness to be interviewed. This unwillingness can be temporary (having family problems or work-related pressures) or permanent. The latter, such as fear of inspection, unpleasant experiences with earlier surveys and general concerns about confidentiality and privacy, might be more difficult to overcome. Inclination may also be determined by personality traits, such as suspiciousness, misanthropy, misogyny, reclusion and paranoia. Refusal conversion may be possible when situational or temporary factors are in place, but not for permanent types of disinclination.

Lack of interest in the survey topic can result in lower cooperation or less meaningful data (more items missing), whereas busyness as a reason not to cooperate may lead to little systematic variation, which might mean that busyness is just a polite (and easy) way of saying 'no' or is not related to the topic of the survey (Couper 1997). Singer et al. (1998) found that respondents who said they would not do the survey again did consider it a waste of time or not useful, and that respondents who disagreed that responsible persons should cooperate provided data of poorer quality (more 'don't knows', and more refusals to answer a question). Laurie et al. (1999, p. 279) found that an initial refusal was more likely to be final when the reason for refusal was survey-related (confidentiality, too long, waste of time) rather than respondent-related (too busy, not able, stressful situation, refusal by proxy).

Rogelberg et al. (2003) distinguished between active and passive non-respondents in a follow-up survey. Failure to respond by the former, more or less accidental non-respondent was not based on a conscious and overt a priori decision. The active non-respondent, in their model had made a conscious decision not to respond to the survey as soon as the survey was received. In this study passive non-respondents turned out to be very similar to respondents. Active non-respondents, however, were less satisfied with the survey sponsor, less conscientious, more likely to leave the university and less agreeable.

Analysis of the success of refusal conversion in the European Social Survey Round 3 (Stoop et al. 2010) showed that in Germany and the Netherlands (both countries that delivered a fair number of converted refusals), interviewers felt that persons who gave general reasons for refusal such as 'bad timing' or 'not interested' were likely to cooperate in the future. German interviewers attributed a low likelihood of future cooperation to those who refused for privacy reasons (and cases with a low probability were most likely not reassigned to another interviewer in the German case).

These and other studies suggest that survey non-cooperation might be more or less 'at random' when situational factors are the reason for non-participation. These non-respondents are busy, have minor household crises to deal with, simply do not feel like doing it and have no

strong feelings on the value and enjoyment of surveys. They might well participate if the interviewer were to come back later or if they could have completed the interview at their own convenience. Their non-response might therefore be *ignorable*.

However, if non-participation is largely determined by the topic or the sponsor of the survey, or other serious concerns, non-response will be 'not at random', and cannot be ignored. In addition, persons who do not participate for survey-related reasons will be more difficult to convert than persons who do not cooperate for accidental reasons. This could mean that after refusal conversion efforts those non-respondents that remain, the hard refusals, will also be the group that causes the largest bias.

9.7.2 Detecting and Adjusting for Bias

Enhancing response rates is expected to minimise bias. As Groves and Peytcheva (2008) have convincingly shown, however, this is not necessarily the case. Indeed, as shown above, going for the easy cases might result in a non-participating group that may have a number of characteristics in common (privacy concerns, dislike of the topic or sponsor, dislike of surveys, distrust of government, no interest in politics). It is also not true that difficult cases (for instance converted refusals) can be used as a proxy for final refusals, as has been shown by Stinchcombe et al. (1981), Smith (1983), Smeets (1995), Voogt et al. (1998), Curtin et al. (2000), Borg (2000), Lynn and Clarke (2001), Keeter et al. (2000), Teitler et al. (2003), Stoop (2004, 2005), Abraham et al. (2006), Van Ingen et al. (2009), and Stoop et al. (2010).

If (hard) refusals are a specific group, this could result in substantial bias when their reason for refusal is related to the topic of the survey (Groves 2006). To adjust for this kind of bias, auxiliary variables have to be available to reflect these reasons behind non-participation that are related to the topic of the survey. Auxiliary variables that measure involvement in voluntary work, interest in politics, having a fixed landline telephone and having a listed number, etc., might help, but these variables are rarely available. A general strategy will be hard to devise, however, "part of the explanation for response behaviour rests on people's assessment about the merits and demerits of surveys in general, but as much, perhaps more, is unique to the particulars of each survey. It is this complexity that makes global remedies for non-response such as a universal weighting system to correct bias error-prone, even though there are various useful techniques for use on a survey-by survey basis." (Goyder 1987, p. 187).

9.8 Conclusion

Non-response due to refusal is a major problem in surveys. The extent of the problem is difficult to assess, however. First, refusal can only be measured in interview surveys, and even then imperfectly. Is refusal by a household member refusal or non-contact, for instance? Illustrative here is Fig. 9.1. In the European Social Survey the emphasis is on minimising the non-contact rate (with a fair amount of success in many countries) and minimising the refusal rate (which remains a problem in many cases). Stealthily, the category 'not able/other', usually a minor reason for non-response, has become a major factor in the ESS.

Second, the likeliness of a refusal, or the willingness to cooperate, is not a characteristic of specific, identifiable groups. People can cooperate for different reasons, depending on psychological factors, the topic of the survey, the sponsor, the mode, the presence of an incentive, the interviewer and their mood. Interviewers are an important factor in obtaining cooperation, and interviewer training is important. On the other hand, general characteristics of interviewers that predict success are also hard to identify. Even the effect of experience is difficult to isolate, if only because less successful interviewers may drop out. What is important is that interviewers tailor their behaviour to the person they meet on the doorstep. Having an adequate answer to reasons not to cooperate (I can come back later; it's not difficult, your answers will be kept confidential) certainly helps.

Third, in the field it is difficult to ascertain what the exact reason for refusal is. Interviewers can register doorstep comments, but this is difficult to standardise. In addition, 'I'm not interested' or 'I'm too busy' are easy ways out for unwilling potential respondents. There is some evidence that people who refuse for temporary reasons (being busy) are more willing to participate at a subsequent visit than people who refuse for more permanent reasons (don't like the topic, don't like surveys).

Finally, the decision to cooperate is often taken in a very short time and is more the result of a heuristic than a conscious decision. This makes it difficult to assess the particular reasons for reticence and to emphasise the attractive sides of survey participation.

Even so, there is one positive side to the absence of omnipresent overruling reasons for refusal. As there are so many factors behind survey refusal, the best solution would be to fully exploit all the different strategies in order to minimise survey refusal and non-response bias. In this way we could minimise the group that does not participate for one particular survey-related reason.

References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American time use survey: Who is missing from the data and how much does it matter? *Public Opinion Quarterly, 70*, 676–703.

American Association for Public Opinion Research (AAPOR). (2011a). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.).

American Association for Public Opinion Research (AAPOR). (2011b). *Best practices*: http://www.aapor.org/Best_Practices1.htm#best7. Downloaded on the July 13, 2011.

Bates, N., Dahlhamer, J., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics, 24*(4), 591–612.

Bethlehem, J. G. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey nonresponse* (pp. 265–287). Wiley: New York.

Bethlehem, J. (2009). *Applied survey methods: A statistical perspective*. Hoboken: Wiley.

Bethlehem, J. G., Cobben, F., & Schouten, B. (2011a). *Handbook of nonresponse in household surveys*. Hoboken: Wiley.

Bethlehem, J. G., & Kersten, H. M. P. (1985). On the treatment of nonresponse in sample surveys. *Journal of Official Statistics, 1*(3), 287–300.

Beullens, K., Billiet, J., & Loosveldt, G. (2009). The effect of the elapsed time between initial refusal and conversion contact on conversion success: evidence from the 2nd round of the European Social Survey. *Quality & Quantity, 44*(6), 1053–1065.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken: Wiley.

Billiet, J., Philippens, M., Fitzgerald, R., & Stoop, I. (2007). Estimation of nonresponse bias in the european social survey: Using information from reluctant respondents. *Journal of Official Statistics, 23*(2), 135–162.

Billiet, J., Matsuo, H., Beullens, K. & Vehovar, V. (2009). Non-response bias in cross-national surveys: Designs for detection and adjustment in the ESS. *ASK. Sociological Methods & Research, 18*, 3–43 (Published by the Polish Institute of Philosophy and Sociology. Polish Academy of Science).

Blom, A., de Leeuw E. D.,& J. Hox (2010). *Interviewer effects on nonresponse in the European social survey.* ISER Working Paper Series No. 2010-25, University of Essex.

Bogen, K. (1996). The *effect of questionnaire length on response rates—A review of the literature*. Proceedings of the Survey Research Methods Section, American Statistical Association.

Borg, I. (2000, November). Früh- versus Spätantworter. ZUMA-Nachrichten 47, Jg. 24, 7–19.

Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Burton, J., Laurie, H., & Lynn, P. (2006). The long term effectiveness of refusal conversion procedures on longitudinal surveys. *Journal of the Royal Statistical Society Series A, 169*, 459–478.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). *Can you hear me knocking: An investigation into the impact of interviewers on survey response rates*. London: The Survey Methods Centre at SCPR.

Cannel, C. F., & Fowler, F. J. (1963). Comparison of a self-enumerative procedure and a personal interview. *A validity study. Public opinion quarterly, 27*(2), 250–264.

Carton, A. (1999). *Selectie, training en evaluatie van interviewers binnen een interviewernetwerk*. Leuven: Garant.

Cohen, G., & Duffy, J. C. (2002). Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics, 18*(1), 13–23.

Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion Quarterly, 61*, 317–338.

Couper, M. P., & de Leeuw, E. D. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds), *Cross-cultural survey methods* (pp. 157–177). New York: Wiley.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64*, 413–428.

De Leeuw, E. D. (1999, April). How do successful and less successful interviewers differ in tactics for combating survey nonresponse? *Bulletin de Méthodologie Sociologique, 62*, 29–42.

De Leeuw, E. (2001). I am not selling anything: Experiments in telephone introductions. *Kwantitatieve Methoden*, *22*(68), 41–48

De Leeuw, E. D., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys. A meta-analysis. *Public Opinion Quarterly, 71*(3), 413–443.

De Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey Nonresponse* (pp. 41–54). New York: Wiley.

Dijkstra, W., & Smit, J.H. (2002). Persuading reluctant recipients in telephone surveys. In: R.M. Groves, D.A. Dillman, J.L. Eltinge & R.J.A. Little (Eds) (pp. 121–134), *Survey nonresponse*. New York: Wiley

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. New York: Wiley.

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly, 74*(1), 1–36.

Edwards, S., Martin, D., DiSogra, C. & Grant, D. (2004). *Altering the hold period for refusal conversion cases in an RDD survey*. Proceedings of the Survey Research methods Section (3440-3444). Alexandria: American Statistical Association.

Elliot, D. (1991) *Weighting for non-response. A survey researcher's guide*. London: Office of Population Censuses and Surveys.

European Social Survey. (2011). *Round 6 specification for participating countries*. London: Centre for Comparative Social Surveys, City University.

Goyder, J. (1986). Surveys on surveys: Limitations and potentialities. *Public Opinion Quarterly, 50*, 27–41.

Goyder, J. (1987). *The silent minority. Nonrespondents on sample surveys*. Cambridge: Polity Press.

Goyder, J., Boyer, L. & Martinelli, G. (2006, October) Integrating exchange and heuristic theories of survey nonresponse. *Bulletin de Méthodologie Sociologique, 92*, 28–44.

Goyder, J., Warriner, K., & Miller, S. (2002). Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics, 18*(1), 1–11.

Groves, R.M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons, Inc.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646–675.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475–495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., & McGonagle, K. A. (2001). A theory-guided interview training protocol regarding survey participation. *Journal of Official Statistics, 17*(2), 249–266.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. *Public Opinion Quarterly, 72*(2), 167–189.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*, 2–31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. Description and an illustration. *Public Opinion Quarterly, 64*, 299–308.

Häder, S. & Lynn, P. (2007). How representative can a multi-nation survey be? In R. Jowell, C. Roberts, R. Fitzgerald, G. Eva (Eds), *Measuring attitudes cross-nationally. Lessons from the European social survey* (pp. 33–52). London: Sage.

Hansen, M., & Hurvitz, W. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association, 41*, 517–529.

Hippler, H.-J., & Hippler, G. (1986). Reducing refusal rates in the case of threatening questions: The 'door-in-the-face' technique. *Journal of Official Statistics, 2*(1), 25–33.

Holbrook, A. L., Green, M., & Krosnick, J. A. (2003). Telephone vs. face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*, 79–125.

Hox, J., & de Leeuw, E. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey nonresponse* (pp. 103–120). New York: Wiley.

Hox, J., de Leeuw, E., & Vorst H. (1995, September). Survey participation as reasoned action: A behavioral paradigm for survey nonresponse? *Bulletin de Méthodologie Sociologique, 48*, 52–67.

Jäckle, A., Lynn, P., Sinibaldi J., & Tipping, S. (2011). *The effect of interviewer personality, skills and attitudes on respondent co-operation with face-to-face surveys*. ISER Working Paper Series No. 2011-14, University of Essex.

Johnson, T.P., O'Rourke, D., Burris, J., & Owens L. (2002). Culture and survey nonresponse In: R.M. Groves, D.A. Dillman, J.L. Eltinge & R.J.A. Little (Eds) (pp. 55–70), *Survey nonresponse*. New York: Wiley

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly, 64*, 125–148.

Koch, A., Fitzgerald, R., Stoop, I., & Widdop, S. (2010). *Field procedures in the european social survey round 5: Enhancing response rates*. Mannheim: European Social Survey, GESIS.

Krosnick, J. A., Narayan, S. S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds), *Advances in survey research* (pp. 29–44). San Francisco: Jossey-Bass.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal survey. *Journal of Official Statistics, 15*(2), 269–282.

Lesser, J., Eyerman, J., & Wang, K. (2008). Interviewer training. In E. de Leeuw, J. Hox, & D. Dillman (Eds), *International handbook of survey methodology* (pp. 442–460). New York: Lawrence Erlbaum Associates.

Loosveldt, G. (2008). Face-to-face interviews. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 201–220). New York: Lawrence Erlbaum Associates.

Loosveldt, G., & Carton, A. (2002). Utilitarian individualism and panel nonresponse. *International Journal of Public Opinion Research, 14*(4), 428–438.

Loosveldt, G., Carton, A., & Billiet, J. (2004). Assessment of survey data quality: a pragmatic approach focused on interviewer tasks. *International Journal of Market Research*, vol. 46 (1), 65–82

Loosveldt, G., & Storms, V. (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research, 20*(1), 74–89.

Lynn, P. (2003). PEDAKSI: Methodology for collecting data about survey non-respondents. *Quality & Quantity, 37*, 239–261.

Lynn, P., & Clarke, P. (2001). *Separating refusal bias and non-contact bias: evidence from uk national surveys*. Working Papers of the Institute for Social and Economic Research, paper 2001-24. Colchester: University of Essex.

Martin, E.A., Traugott, M.W., & Kennedy, C. (2005). A review and proposal for a new measure of poll accuracy. *Public Opinion Quarterly* 69(3), 342–369

Matsuo, H., Billiet, J., Loosveldt, G. & Malnar, B. (2010a) *Response-based quality assessment of ESS Round 4: Results for 30 countries based on contact files*. Onderzoeksverslag Centrum voor Sociologisch Onderzoek. CeSO/SM/2010-2.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010b). Measurement and adjustment of non-response bias based on non-response surveys: the case of Belgium and Norway in the European Social Survey Round 3. *Survey Research Methods, 4*(3): 165–178

Merkle, D. A., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 243–258). New York: Wiley.

Morton-Williams, J. (1993). *Interviewer approaches*. Aldershot: Dartmouth Publishing.

Mowen, J. C., & Cialdini, R. B. (1980). On implementing the door-in-the-face compliance technique in a business context. *Journal of Marketing Research, 17*, 253–258.

Luiten, A. (2011). Personalisation in advance letters does not always increase response rates. Demographic correlates in a large scale experiment. *Survey Research Methods*, *5*(1), 11–20.

Olson, K., Lepkowski, J., & Garabant, D. (2011). An experimental examination of the content of persuasion letters on nonresponse rates and survey estimates in a nonresponse follow-up study. *Survey Research Methods*, *5*(1), 21–26.

Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods, 4*(1), 3–25.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology, 88*(6), 1104–1114.

Schaeffer, N.C., Dykema, J. & Maynard, D.W. (2010). Interviewers and interviewing. In P. Marsden & J. Wright (eds.), *Handbook of survey research* (2nd ed., pp. 437–470). UK: Emerald Group Publishing Ltd.

Schräpler J. P., Schupp J., & Wagner, G. G. (2010). *Individual and neighborhood determinants of survey nonresponse: An analysis based on a new subsample of the german socio-economic panel (SOEP), microgeographic characteristics and survey-based interviewer characteristics*. SOEP papers 288, DIW Berlin, The German Socio-Economic Panel (SOEP).

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 163–177). New York: Wiley.

Singer, E. (2006). Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 637–645.

Singer, E. (2011). Towards a cost-benefit theory of survey participation: Evidence, further test, and implications. *Journal of Official Statistics, 27*(2), 379–392.

Singer, E., Groves, R. M., & Corning, A. D. (1999a). Differential incentives. Beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly, 63*, 251–260.

Singer, E., Mathiowetz, N. A., & Couper, M. P. (1993). The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 census. *Public Opinion Quarterly, 57*, 465–482.

Singer, E., Van Hoewyk, J., &Maher, M.P. (1998). Does the payment of incentives create expectation effects? *Public Opinion Quarterly*, 62, 152–164

Singer, E., Van Hoewyk, J., & Neugebauer, R. J. (2003). Attitudes and behavior. The impact of privacy and confidentiality concerns on participation in the 2000 census. *Public Opinion Quarterly, 67*, 368–384.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999b). The effects of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics, 15*(2), 199–216.

Smeets, I. (1995). Facing another gap: An exploration of the discrepancies between voting turnout in survey research and official statistics. *Acta Politica, 30*, 307–334.

Smith, T. W. (1983). The hidden 25 Percent: An analysis of nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly, 47*, 386–404.

Smith, T. W. (1984). Estimating nonresponse bias with temporary refusals. *Sociological Perspectives, 27*(4), 473–489.

Stinchcombe, A. L., Jones, C., & Sheatsley, P. (1981). Nonresponse bias for attitude questions. *Public Opinion Quarterly, 45*, 359–375.

Stocké, V., & Langfeldt, B. (2004). Effects of survey experience on respondents' attitude towards surveys. *Bulletin de Méthodologie Sociologique*, January 2004, 81, 5–32

Stoop, I. A. L. (2004). Surveying nonrespondents. *Field Methods, 16*(1), 23–54.

Stoop, I. A. L. (2005). *The hunt for the last respondent*. The Hague: Social and Cultural Planning Office.

Stoop, I. (2007). No time, too busy. Time strain and survey cooperation. In G. Loosveldt, M. Swyngedouw, & B. Cambré (Eds), *Measuring meaningful data in social research* (pp. 301–314). Leuven: Acco.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). improving survey response. Lessons learned from the European Social Survey. Chichester: Wiley.

Sturgis, P., & Campanelli, P. (1998). The scope for reducing refusals in household surveys: An investigation based on transcripts of tape-recorded doorstep interactions. *Journal of the Market Research Society 40*(2), 121–139

Te Riele, S. (2002) Vertekening door non-respons. Hoe nauwkeurig zijn de uitkomsten van persoons-enquêtes? Sociaal-economische Maandstatistiek, Jrg. 19, April 2002. Voorburg/Heerlen: Centraal Bureau voor de Statistiek, pp. 20–25.

Teitler, J. O., Reichman, N. E., & Sprachman, S. (2003). Costs and benefits of improving response rates for a hard-to-reach population. *Public Opinion Quarterly, 67*, 126–138.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

Triplett, T. (2002). *What is gained from additional call attempts & refusal conversion and what are the cost implications*? Research Report. Washington, DC: The Urban Institute.

Triplett, T. (2006). 2002 *NASF nonresponse analysis*. Methodology Reports, Report No. 7. Washington, DC: The Urban Institute.

Triplett, T., Scheib, J., & Blair, T. (2001). *How long should you wait before attempting to convert a telephone refusal? Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association.

Van Ingen, E., Stoop, I., & Breedveld, K. (2009). Nonresponse in the Dutch time use survey: Strategies for response enhancement and bias reduction. *Field Methods, 21*(1), 69–90.

Voogt, R. J. J., Saris, W. E., & Niemöller, B. (1998). Non-response, and the gulf between the public and the politicians. *Acta Politica, 33*, 250–280.

West, B., & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance. *Public Opinion Quarterly, 74*(5), 1004–1026.

Further Reading

Bethlehem, J. G., Cobben, F., & Schouten, B. (2011b). *Handbook of nonresponse in household surveys*. Hoboken: Wiley.

Blom, A., & Kreuter, F. (2011). Special issue on nonresponse. *Journal of Official Statistics, 27*(2), 151–414. (Guest editors).

Groves, R. M., & Couper, M. P. (1998b). *Nonresponse in household interview surveys*. New York: Wiley.

Singer, E. (2006). Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5).

Stoop, I. A. L. (2005). *The hunt for the last respondent*. The Hague: Social and Cultural Planning Office.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). Improving survey response. Lessons learned from the European Social Survey. Chichester: Wiley.

Non-Response and Measurement Error 10

Jaak Billiet and Hideko Matsuo

10.1 Introduction

This chapter deals with two broad issues that are crucial in survey research: non-response and measurement error. Both rely on different research traditions and developed largely separately, although both have a prominent place in famous handbooks or conference proceedings on survey methodology (Rossi et al. 1983; Groves 1989; Biemer et al. 1991), and in the programs of international conferences such as AAPOR, ESRA, and WAPOR. Moreover, these two aspects of *Total Survey Error* (Groves 1989, pp. 8–18; Groves and Lyberg 2010) are related. Non-response and measurement error have an effect on the response distributions, and sometimes on correlations, but higher response rates do not automatically lead to less measurement error (Groves 1989, p. 218). The opposite is possible when higher response rates are obtained by including less motivated respondents in the obtained samples (Kaminska et al. 2011; Stoop et al. 2010, pp. 188–203).

Central to the Groves conception of error is the distinction between variable error (variance) and bias. The total error of a survey is the mean square error (MSE), i.e., the sum of all variable errors (variance) and all bias. Variable error is associated with "precision", whereas, bias relates to "accuracy". Errors are specific to a single statistic (in a single survey) and can thus vary over the implementations of a survey design, while bias is a kind of constant error that affects a statistic in all implementations of a survey design (Groves 1989, pp. 8–14). Groves distinguishes between a survey statistics viewpoint, where (in principle) an observable true value is assumed, from the psychometric perspective wherein attributes are measured that cannot be observed by anyone (unless the respondents themselves). One can even discuss whether the attribute really exists.[1]

Bias has a clear meaning in situations where an observable true value of a statistic in the population exists and an observable true value for the respondents in the sample is assumed. Bias is then the difference between these two. In this conception there is no bias in the measurement of un-observable attributes as opinions, attitudes, preferences, and values, but only variable error over repeated measurements (variance) (Groves 1989, p. 18). If one wants to measure subjective attributes one needs

J. Billiet (✉) · H. Matsuo
Centrum Voor Sociologisch Onderzoek (CeSO),
K.U. Leuven, Parkstraat 45 Bus 3601,
3000 Leuven, Belgium
e-mail: Jaak.billiet@soc.kuleuven.be

H. Matsuo
e-mail: hideko.matsuo@soc.kuleuven.be

[1] See for example the discussion on so called 'non-attitudes' in Saris and Sniderman (2004).

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_10, © Springer Science+Business Media New York 2012

observed[2] indicators or sets of indicators to measure the underlying unobservable characteristic (constructs or latent variables) which can only be approximated with any applied measurement. Some refer to the conception of a "true value" in this situation as a "Platonic true score" (Bornstedt 1983). In the approaches to the study of error that will be discussed further, models are proposed in which it is possible to estimate unreliability (random error) and invalidity (systematic error).

Groves distinguishes in the TSE concept two main types of errors: errors of non-observation and observational error. Errors of non-observation are coverage, non-response, and sampling. The sources of observational error are the interviewer, respondent, the research instrument, and the mode (Groves 1989, pp. 14–24). The first part of this chapter is devoted to only one aspect of non-observation error, non-response, and the second part exclusively deals with measurement error related to the research instrument (the questionnaire). This means that this chapter covers a limited part of the quality of a survey. This becomes clear when we consider the recent developed "*Total Quality Management*" (TQM) framework (Loosveldt et al. 2004).

TQM is developed in the context of large interview surveys in which the management of a survey by a central coordination team plays a crucial role. This is for example the case in large-scale cross-national surveys as the *European Social Survey*. This framework is based on the distinction of both a process evaluation and an output evaluation of both the obtained sample and the registered responses (measures) to questions at the other. It integrates several existing approaches as '*Total Survey Error*' (Groves 1989; Groves et al. 2004), '*Continuous Quality Improvement*' (CQI) (Biemer and Caspar 1994), and the '*Total Design Method*' (TDM) (Dillman 1978, 2000). TQM builds on the strengths of each of these approaches to data quality assessment. Like TSE, it pays attention to the statistical error components of sample survey statistics, but places it in a much broader perspective of data quality in accordance with the concept of 'responsive design' (Groves and Heeringa 2006) which takes costs and time into account (Biemer 2010). TQM considers, like CQI, the process of survey production, and decomposes the survey into three phases: before, during and after the fieldwork (e.g., training and fieldwork preparation, contact procedure, follow-up, and feedback). Finally, common with TDM, the TQM framework considers factors that can increase response rates by sending out cover letters, paying attention to questionnaire design, and follow-up of non-respondents through response enhancement procedures.

Data quality thus covers all stages in the research process (Lyberg et al. 1997): the design and implementation of random samples; questionnaire construction; the translation process (and its documentation) in case of multi language samples; response enhancement during the fieldwork stage; the study of measurement error of the obtained measures; the collection of events and contextual information; transparency, documentation of the data files. In order to obtain equivalent samples and responses in cross-national surveys, the whole enterprise, which is performed by several task forces and research teams, can only succeed thanks to a firm coordination of the whole. The distinction between process evaluation and output evaluation of both the obtained sample and the registered responses is certainly useful in all stages of the process of a survey production/implementation but we evidently cannot deal with all these aspects within the context of this chapter.

In the first part of this chapter on error of non-observation, some elements of process evaluation are briefly discussed. The focus is further, however, mainly on output evaluation: the assessment of the obtained sample and the study of non-response bias. For illustration purposes examples of the *European Social Survey* (ESS) are used because the quality of this cross-nation survey is widely recognized[3] (Jowell et al. 2007).

[2] "Observed" in the meaning of "measured" in a broad sense by recording a response or by observing an event.

[3] The ESS Central Coordination Team received the 2005 *European Descartes Prize* for excellence in collaborative scientific research.

In the second part on measurement error, a state of the art is offered. It covers question wording effects that emerged from the early split ballot (field) experiments and the contribution of cognitive psychology to the understanding of survey response. After a short critical review of the early split ballot tradition, and some recent developments, attention is paid to the modeling of measurement error and response styles. The relation between both parts, non-response and measurement error, is briefly discussed in the conclusions.

10.2 Evaluation of Obtained Response

The first part of this chapter deals with the process and the output evaluation of the obtained sample, or response quality. It is largely based on our experience with ESS as it is reported in Stoop et al. (2010). The process evaluation covers all the planned steps of a survey production/implementation in order to optimize the response rate and to minimize non-response bias. Measures taken in view of response enhancement are crucial. The output evaluation is devoted to the study of the obtained sample by assessing response rates, the amount of non-response, and estimation of non-response bias (Billiet et al. 2007a, b; Stoop et al. 2010). Information about the characteristics of the non-respondents and non-response bias is also crucial.

10.2.1 Process Evaluation of Obtained Sample

With respect to the obtained sample, the process evaluation focuses on the fieldwork training and preparation, the contact procedure, and follow-up and feedback. This evaluation focuses specifically on aspects of interviewer training that deal with the planning of the visits, contact procedures, respondent persuasion, refusal conversion techniques, and the implementation of follow-up visits or quality control back-checks (Billiet et al. 2007a, b; Stoop et al. 2010, pp. 115–159).

Response enhancement measures usually target on measures occurring during the preparation stage and the data collection stage of a survey. However, in case of successive survey, as is the case in the biennial ESS, measures taken after the fieldwork may play a role in the subsequent surveys, for example, feedback to interviewers and information about response rates according to the time of the contact attempts (Stoop et al. 2010, pp. 121–142). Response enhancement measures in the preparation stage and during the data collection stage include properly selecting and training interviewers (in doorstep techniques); sending advance letters whenever possible; extending the fieldwork period to at least 30 days in order to undertake at least four contact attempts; the timing of the contact attempts (over the weekday and time of the day); contact mode of the first contact attempt (e.g. personal visit in ESS); at least one personal visit in case of a (first) refusal by telephone; reapproaching the respondents who refuse and "conversion" attempts instead of substitution of original sampling units; using incentives when response rates remain too low; developing and implementing detailed contact forms and call records which enable detailed analyses and documentation of the response process in each country; providing feedback to NCs based on the analysis of call record data in order to improve response rates in following rounds (Billiet et al. 2007a, b). The concrete implementation of these measures in ESS is documented in Stoop et al. (2010, pp. 62–63).

Not all measures are equally important in view of response enhancement. The most effective measures in ESS[4] concern the number and timing of contact attempts and the refusal conversion strategy. This is based on the analyses of the call record data of previous ESS rounds. Evaluation measures can therefore focus on these aspects. The first objective of the contact procedure is to establish direct contact with the sample unit. It is therefore important to analyze the number of contact attempts required to establish such direct contact. Computing (ratio) scores between mean number of contact

[4] Discussion of this can be found in the document of ESS Round 4 specification for participating countries.

attempts and non-contact rates, and analyzing the impact of the number of contact attempts on the final response rate, is an equally important tool in view of the assessment of the effect of the response enhancement measures. Usually, response is only possible after a prior contact. Detailed documentation of the refusal conversion process in terms as initial refusal rates, proportions of reapproached refusals, and success rates of refusal conversion attempts is crucial in the context of assessment of the cost and benefit balance (Stoop et al. 2010, pp. 171–176). The increase in final response rates can be substantial depending on the combined efforts in refusal conversion activities.

10.2.2 Output Evaluation of the Obtained Sample

It is useful to take a close look at response and non-response rates since high non-response rates may increase the *likelihood* of non-response bias, although this is not unconditionally the case (Groves 2006). Correct documentation of the response, and at least equally important, the non-response, is vital for assessment of the quality of a survey. It is also indispensable for the study of non-response bias.

10.2.2.1 Documentation of the (Non-)Response Outcomes

Anybody who wants to compare response rates should understand primarily how these rates are calculated since response rates are differently computed across surveys. A standard way of computing response rates in high quality surveys these days uses the AAPOR (2006) definitions. The starting point is the basic distinction at the sample unit level between 'completed interview' (I); 'partial interview' (P); 'refusal' (R); 'non-contact' (NC); 'other' (O); and 'ineligible' (IE). The ineligible cases are excluded from the denominator of the response rate formula. In ESS, which is a cross-country survey, the ineligible cases are sampled units that are passed off or that have moved outside the country. The response rate is calculated by dividing the number of achieved valid and complete interviews (I) by the number of eligible sample units (Stoop et al. 2010, pp. 61–67):

$$[I / (I + P + R + NC + O)]$$

Response and non-response rates can be assessed based on benchmarks, but these are mostly not the same for all single-country surveys. Effects of nonresponse in cross-nation research is even more challenging because non-response is affected by country-specific issues like, culture and past and current survey traditions (see Stoop and Harrison Chaps. 9, 15). Comparability of cross-nation data can be only achieved when data quality affected by non-response is assessed and improved (Couper and de Leeuw 2003, p. 157). Take for example that the target response was set to 70% (and a non-contact rate of 3%) in ESS. Call record data are the primary sources for the study of response rates and non-response. The files with call record data document the outcome of each contact attempt for every sample case (respondents as well as non-respondents) (Lynn et al. 2001; Blom et al. 2010). The contact forms (CF) are short and compact interviewer questionnaires applied to each sample (selected) person, and record information about the date, day, time, mode, and outcome of each contact attempt. Contact forms also contain information about the non-respondents (reason for refusal, estimation of future cooperation) and ineligible cases. Apart from the outcomes of each contact attempt, the CFs also contain observable data in the form of housing and neighborhood characteristics. In the case of address-based and household-based sample frames, the household and respondent selection procedure is noted in the contact form. All of these features in the contact form are useful for the analysis of non-response bias.

Contact record data also provide an extra opportunity for data quality checks. These data can be cross-checked against information from other sources including reports from fieldwork managers and the main file, which is the registered response file. Such cross-checks can focus on a number of indicators including the total number of sample units (response units, non-response units, and ineligible units), the total number of interviewed units by checking the consistency of information (e.g. occurrence and

time) across different data sets. In most cases, when the response rate is calculated on the basis of the contact files, the outcome of the last contact attempt is taken as the final response code. Refusals constitute an exception: even if the last contact attempt results in non-contact or another outcome, they are listed as refusals (Billiet et al. 2007a, b; Lynn et al. 2001).

10.2.2.2 The Assessment of Non-Response Bias

The problem of non-response bias has to be distinguished from the response enhancement question as such since there is no perfect relationship between the degree of non-response and the degree of bias (Groves 2006). Some scholars argue that better than aiming for high response rates one should try to minimize non-response bias by using a responsive design (Groves and Herringa 2006). However, minimizing bias during fieldwork is even more difficult to implement in practical fieldwork protocols than enhancing response rates. One should realize that non-response bias varies substantially across variables (Groves 2006; Vehovar 2007). Assessing non-response bias requires either population information with respect to the core variables of a survey, or similar information on non-respondents (Stoop et al. 2010, pp. 211–277). Both are rarely available, at least in surveys on opinions, attitudes, and values. There are several operational definitions of bias (Groves 2006) that are applied in ESS.

Bias as deviation between sample and population distributions. In this definition, a comparison is made between distributions of some variables in the sample and distributions of these variables in the population, as far as these are available. This method of bias estimation has been used in the ESS (Vehovar 2007). Bias is not only estimated for documented demographic variables in a population but also, under certain assumptions, for variables that are not documented in the sample. This is done by comparing the weighted sample (according to the known variables) with the unweighted one for a large number of variables. It is clear that the success of this way of evaluation of the sample depends strongly on the correlation between the documented variables (the so-called post-stratification variables) with the (often attitudinal) target variables in a survey. The correlations are mostly low (see Stoop et al. 2010, pp. 219–223).

Bias as difference between cooperative and reluctant respondent: This approach to study non-response bias is to compare converted refusals with cooperative respondents. Here, the underlying assumption is that with less field efforts these would have been final refusals, and with even more field efforts additional refusals could have been converted. This underlying assumption cannot always be tested and the support for it is ambiguous (Stoop et al. 2010, pp. 214–228).

Bias as difference between respondents and non-respondents. This way of studying bias more directly focuses on the characteristics of the non-respondents and on the differences between respondents and non-respondents in the gross samples. The call record data in the ESS data sets contain several possibilities. In principle, for all selected sampling units observational data has been observed about gender, age, and one can ask the interviewer to record some characteristics of the housing and neighborhood of the respondents (Stoop et al. 2010, pp. 229–241). Surveys among non-respondents some weeks after a survey is finished using a small (PEDAKSI) questionnaire (Lynn 2003), as is done after ESS round 3, is a rather expensive way of obtaining additional information about bias, but is another way to study bias (Stoop et al. 2010, pp. 252–276; Matsuo et al. 2010b). Equally, using contact sequences through optimal matching as a tool for non-response adjustment purposes is another kind of approach to study this (Kreuter and Kohler 2009).

Adjustment for bias in model-based approaches. These approaches are considered: model-based estimates of response probabilities; prediction and response propensity modeling; calibration estimation; the extended use of auxiliary data at the individual level and aggregate data at (small) context level; the introducing of instrumental variables and jack-knife methods in variance estimation with replicate weights computed in one step even though the calibrations weights themselves may be determined iteratively. The information needed for applying a model-based

approach can be obtained from the sample itself (respondents and non-respondents), from the geographical context, and from auxiliary variables in the population (Groves 2006).

Based on the described operational definitions of bias, research activities[5] are organized around four ways of bias detection and estimation in ESS. These are: (1) post-stratification weighting (Vehovar 2007); (2) comparison of cooperative with reluctant respondents (Billiet et al. 2007a, b); (3) using information from observable data (Stoop et al. 2010, pp. 243–252); and (4) obtaining information from additional surveys among non-respondents (Matsuo et al. 2010b).

Each of these approaches has its own advantages and disadvantages (Billiet et al. 2009): (1) post-stratification approach using population data/statistics is useful for adjustment but this information is limited to variables such as, age, gender, and education. However, this information is not always available and reliable in all countries and the relationship with target variables is rather weak; (2) many variables are available for reluctant respondents but non-response can be only adjusted for refusals (subgroup of non-respondents) and not for all non-response; moreover, it is questionable how far the reluctant respondents are representative for all refusers; (3) observable data (dwelling and neighborhood variables) is available for all sample units including both respondents and non-respondents but the information is limited to a small number of variables (type of house, physical condition of house, presence of vandalism, and presence of litter); missing values in these variables are high, the measurement is not very reliable, and the relationship with target variables is also weak; (4) a survey among non-respondents some months after the fieldwork of main ESS round contains small number of (relevant) questions (16) but is only available for four countries. The last approach applied propensity score stratification method (10 classes) on the basis of the information of co-operative respondents and non-respondents but the weighting was not very effective (ESSi workshop 2011 review).

10.2.3 Increasing Response or Reducing Bias?

Survey response rates are generally decreasing (de Heer 1999), and non-response rates increase including refusals and non-contacted sampled persons. This may affect bias in the estimates. However, response rates are insufficient indicators of survey quality since what matters most is the extent to which respondents differ from non-respondents on crucial (target) variables in a survey. Against this background, the 'Risq' project'[6] has developed the R-indicator or representativeness indicator. This indicator allows for comparing the quality of surveys between populations and over time. It also constitutes a valuable tool for monitoring the quality of the fieldwork in accordance with the concept of 'responsive design' (Groves and Herringa 2006).

The unique contribution of the R-indicator is situated in the sample-based models for individual response propensities.[7] R-indicators are random variables based on the standard deviation of estimated response propensities, and the modeling of response propensities has important implications. This means that the sample size and availability of auxiliary variables play an important role in the estimation of response propensities. When survey responses are representative, all response propensities have equal value, whereas, when the survey response is not representative, response propensities have large variations.

The kind of variables used for the R-indicator can differ between surveys in terms of focus and extent. Use is usually made of auxiliary information like the socio-demographic characteristics of the sample unit, sample unit dwelling and

[5] These research activities are ESSi (infrastructure) program funded project of 'Joint Research Activities (JRA2): improving representativeness of the samples' implemented in 2006–2011.

[6] This project is financed by the EU 7th Framework Programme and involving a number of European institutions.

[7] Schouten et al. (2009) discusses the mathematical features of R-indicators. Risq project (http://www.risq-project.eu/indicators.html)

neighborhood characteristics, or the results of every attempt to contact the sample unit. These are available from sample frame data, censuses, or administrative data, to contact files based on interviewer observation. Needless to say, such variables must be available for all sample units outside the survey questionnaire and are constant during the fieldwork period. Considered variables are: age, gender, ethnicity, degree of urbanization, type of household, region, province, existence of listed phone number at household level, average house value at zip code level, welfare status (receiving disability or social allowance), employment status (Schouten et al. 2009) as well as treatment variables (or so-called key process variables) such as number, timing, mode of contact attempts, and interviewer skills (Loosveldt and Beullens 2009).

10.2.4 Evaluation of the Obtained Sample in ESS: Illustrations

This section presents three applications of the evaluation of the obtained sample in a cross-national survey. In the first application, different response and non-contact rates are compared between four rounds of ESS. This is a typical example of output evaluation. The second application deals with one aspect of the response enhancement strategy, namely, the refusal conversion process, which is an example of process evaluation. The third application is an example of output evaluation assessing the level of non-response bias by comparing cooperative respondents with reluctant respondents. The evaluation of the obtained samples in ESS took place in the context of general ESS research activities and the aforementioned *Joint Research Activities* (JRA) on non-response bias (Billiet et al. 2009).

10.2.4.1 Illustration 1: Assessing Response and Non-Contact Rates

Response rates are most usefully assessed with regard to a benchmark, usually a minimum response rate target. The ESS minimum response rate target is 70% while the maximum non-contact rate is 3%. These targets are not always reached. In Round 1, the number of countries reaching (or almost reaching) the 70% minimum response rate target is small (four countries) yet significant cross-country variation exists as response rates range from 32.5% (CH) to 79.5% (GR). The mean response rate is 60.3% and the dispersion (standard deviation) is 12.6 points. Two rounds later, the average Round 3 response rate for 24 countries increases to 62.5% with a standard deviation of 7.9 points, whereas a decline in the average response rate among 30 countries (0.6% points) and an increase in the standard deviation of 2.2 points are observed between Rounds 3 and 4. The spread between the lowest (42%) and the highest (81.0%) response rates is the second highest (38.3% points) of the four rounds. Target non-contact rates, however, are not always reached either. It should be noted that low non-contact rates (maximum 3%) is more easily achieved than high response rates (minimum 70%). The number of countries reaching the maximum 3% non-contact rate increases by each rounds (from 5 in Round 1 to 13 in Round 4). In Round 1, the mean non-contact rate is 5.4% and the dispersion (standard deviation) is 3.5 points. In Round 3, the average Round 3 non-contact rate for 24 countries decreases to 4.6% with a standard deviation of 3.1 points. In Round 4, there is an increase in the average non-contact rate among 30 countries (0.6% points) and an increase in the standard deviation of 1.0 points (Matsuo et al. 2010a, p. 55) (Table 10.1).

10.2.4.2 Illustration 2: Evaluating Response Enhancement Strategy—Refusal Conversion Process

As explained before, refusal conversion consists of reapproaching initially reluctant respondents in order to persuade them to reconsider participating in the survey. Refusal conversion can boost the response rate since it can be assumed that many initial refusals are 'soft' refusals instigated by particular circumstances (e.g., bad timing, inappropriate interviewer tailoring technique, etc.) that may not apply at a later occasion.

In ESS, substantial cross-country variation existed with respect to both the proportion of

Table 10.1 ESS round 1 (2002)–round 4 (2008)

	Round 1	Round 2	Round 3[a]	Round 4
Response rates				
Mean%	60.3	62.6	62.5	61.9
STD	12.6	9.6	7.9	10.1
Lowest%	32.5	43.6	46.0	42.7
Highest%	79.5	78.8	73.2	81.0
Range	47	35.2	27.2	38.3
# more than 70%	4	6	4	7
Non-contact rates				
Mean%	5.4	5.4	4.6	5.2
STD	3.5	3.0	3.1	4.1
Lowest%	0.8	0.9	0.8	0.0
Highest%	14.7	10.9	13.1	14.1
Range	13.9	10.0	12.3	14.1
# less than 3%	5	7	10	13
N (countries)	20	23	24	30

Reproduction of table based on Matsuo et al. 2010a

Average response and non-contact rates (% of total eligible sample sizes) and characteristics of the response rate distributions at country level [Figures based on the analysis of the call record data and/or National Technical Summary (NTS) (see Matsuo et al. 2010a, p. 55)]

[a] Country with largest response rates is missing in Round 3

initial refusers contacted for refusal conversion and the refusal conversion success rate. Table 10.2 shows the effect of refusal conversion presenting figures on the number and proportion of initial refusers, the proportion reapproached, number of converted respondents, the success rate of the conversion attempts, and the effect on final response rates. Both the share of initial refusers reapproached for refusal conversion and the refusal conversion success rate depend to a large extent on the nature of national refusal conversion administrative and fieldwork specifications on the identification, selection, and contacting of reluctant respondents. Examples are given from the most recent round (Round 4) as it has the highest number of participating countries implementing refusal conversion attempts (see Table 10.2).

Substantial cross-country variation is first of all observed with regard to the coverage of refusal conversion activities (percentage of initial refusers that are reapproached). A large proportion (>80%) of initial refusers was reapproached in some countries (e.g. NL and CH) while very limited refusal conversion efforts (<10%) were made in the other 10 countries.[8] For example in The Netherlands, the initial refusal rate is high (more than 50%) and approximately 90% of the initial refusers are called on again. Hardly any initial refusers are reapproached in the aforementioned ten countries with an initial refusal rate ranging between 4.6 and 37.9%.

The refusal conversion success rate also differs substantially between countries. This is the share of reapproached initial refusals which were successfully converted. However, one must interpret this carefully. Some countries target their refusal conversion activities to a restricted number of initial refusers, whereas other countries reapproach substantial numbers of initial refusers (e.g. more than 50% of initial refusers).

[8] BG, CY, CZ, DK, GR, LV, LT, PT, RO and UK.

Table 10.2 Outcome of refusal conversion attempts based on the analysis of contact files, Round 4

Country	Initial refusers (*N*)	Initial refusers (%)	Reapproached (%)	Converted respondents (*N*)	Success rate (%)	Effect on response rate[a]
BE	844	28.3	67.3	70	12.3	2.3
BG	493	16.6	4.5	17	77.3	0.6
CH	1535	41.2	86.5	291	21.9	7.8
CY	151	10.0	4.6	2	28.6	0.1
CZ	778	26.8	2.4	0	0.00	0.0
DE	2709	42.0	55.9	611	40.4	9.5
DK	986	33.1	0.6	5	83.3	0.2
ES	820	21.2	58.2	198	41.5	5.1
FI	750	23.4	59.6	80	17.9	2.5
FR	1614	38.8	56.8	106	11.6	2.5
GB	1197	27.8	58.3	146	20.9	3.4
GR	502	18.0	0.2	1	(100)[b]	0.0
HU	758	30.1	50.7	67	17.4	2.7
HR	949	29.4	41.1	170	43.6	5.3
IE	963	26.8	41.4	93	23.3	2.6
IL	281	8.7	34.5	27	27.8	0.8
LT	1347	37.9	0.0	0	0.00	0.0
LV	406	11.6	2.0	4	50.0	0.1
NL	1882	52.7	91.2	435	25.3	12.2
NO	892	34.8	54.6	101	20.7	3.9
PL	479	21.0	56.8	69	25.4	3.0
PT	627	20.1	0.0	0	0.00	0.0
RO	144	4.6	1.4	0	0.00	0.0
RU	986	26.4	36.0	53	14.9	1.4
SE	822	28.0	46.4	79	20.7	2.7
SI	553	25.3	37.3	77	37.4	3.5
SK	481	19.3	57.6	161	58.1	6.5
TR	719	18.9	19.3	13	9.4	0.3
UA	753	25.1	3.6	8	29.6	0.3

Estonia was excluded from the analysis due to the lack of information about refusal conversion process. On the basis of Matsuo et al. (2010a), this table takes a similar format as reported in Stoop et al. (2010), p. 172

[a] Same calculation applied expressed as percentage point differences relative to initial refusal rate

[b] Note the very small proportion of reapproached

Countries reapproaching substantial number of initial refusers combined with conversion rate that exceeded 2% are observed in 8 countries.[9]

The final effect on response rates is shown in the last column of Table 10.2 as a percentage increase of response rate with and without refusal conversion attempts. The impact of refusal conversion attempts on final response rate also differs substantially between countries. For some countries, the impact of refusal conversion on final response is major. Substantial changes in response rates (3+ or more) are observed in ten

[9] CH, DE, ES, GB, NL, NO, PL and SK.

countries.[10] Most of these ten countries, except for Poland and Slovenia, obtained more than 100 reluctant respondents. In six other countries,[11] changes of 2–3% points are observed. For the remaining 13 countries, the impact is minor as changes of <1% are observed.

10.2.4.3 Illustration 3: Comparing Cooperative and Reluctant Respondents

One way of assessing non-response bias, consists of taking a closer look at the category of converted initial refusers or "reluctant respondents". Like a comparison between respondents and non-respondents, a comparison between "reluctant respondents" and "cooperative respondents" (respondents who agree to survey cooperation without refusal conversion activities) can provide an insight into the presence and extent of non-response bias.

The case study (Stoop et al. 2010, pp. 234–238) presented here pertains to ESS Round 2 and focuses on countries with comparatively high numbers (>100) of reluctant respondents: Switzerland, Germany, Estonia, The Netherlands, and the Slovak Republic.[12] Reluctant respondents were compared with cooperative respondents for a number of background, attitudinal, and media use variables and for responses to a number of survey questions that were marked by a rather large average absolute standardized bias according to the post-stratification approach (Vehovar 2007).

A multivariate logistic regression model was used to assess the differences between the two types of respondents (reluctant vs. cooperative) (Table 10.3). The dependent variable was the odds ratio between the two types of respondents. Target persons from larger Swiss households were initially more reluctant to participate but more often persuaded to cooperate at the refusal conversion stage. In Germany, converted refusals were significantly more likely to be women, to be older, to live in large cities, to engage in Internet surfing, and to have a history of unemployment, and less likely to participate in politics. In Estonia, reluctant respondents were more likely to be women and to have a paid job, while living in a village and having ever been unemployed were associated with a greater likelihood of being a cooperative respondent. In Slovakia, compared to cooperative respondents, converted refusals were more likely to be old, to have a medium level of education (upper and post-secondary education), to be more religious, and to feel comfortable with their household income.

On the other hand, they were less likely to have ever had a job and to feel safe walking in the neighborhood after dark. In The Netherlands, the likelihood of being a converted refusal increased when the respondent was female, had an average upper and post-secondary level of education, watched more television, and surfed more frequently on the Internet. Converted refusals were also more likely to have "ever had a job", to feel healthier, to see immigrants as a threat, and to be dissatisfied with their own life. However, unexpectedly given the literature, respondents who trusted political institutions, obeyed the law, participated in political organizations, and felt socially integrated were somewhat more likely to be converted refusals (Stoop et al. 2010, pp. 234–38).

10.3 Measurement Error

The conception of survey measurement error and the way to study it has been changed over time. After the early "split ballot" approach, the study of measurement error changed in the direction of mathematical modeling (Rossi et al. 1983, p. 17). Until the 1990s the conception of survey errors was statistical rather than (cognitive) psychological in character. The statistical models paid more attention to the variance component of the error as is the case in the Groves' (1989) standard work on survey errors and survey costs. Since the CASM[13] movement

[10] CH, DE, ES, GB, HR, NL, NO, PL, SI and SK.

[11] BE, FI, FR, HU, IE and SE.

[12] Slovenia was excluded from the analysis due to data quality issue although the number of reluctant respondents was high.

[13] CASM: cognitive aspects of survey methodology.

Table 10.3 Statistical significant (<0.05) Logistic regression estimates (odds ratios) for reluctant versus cooperative respondents, ESS 2

	Switzerland	Germany	Estonia	Slovakia	The Netherlands
Background variables					
Male (= Yes)		0.797[c]	0.706[b]		0.614[a]
Age		1.019[a]		1.019[c]	
Number of household members	1.177[b]				
Urbanicity					
Countryside and village		0.634	0.677[b]		
Town and small city		0.816[b]	1.293[c]		
Large cities and suburbs		1.312[a]	1.143[a]		
Level of education					
Lower and lower secondary				0.682	0.817
Upper and post-secondary				1.758[b]	1.243[b]
Second stage of tertiary				0.834	0.984
Labour market status					
Paid job (1 = yes)			1.863[a]		
Ever job (1 = yes)				0.425[c]	1.694[c]
Ever unemployed (1 = yes)		1.357[c]	0.584[c]		0.589[c]
Good health (1 = yes)					1.296[c]
Comfortable income (1 = yes)				1.887[c]	
Religious involvement (0–10)				1.334[a]	
Attitudes					
Perceived ethnic threat (0–10)					1.136[a]
Trust political inst. (0–10)					1.101[c]
Political participation (0–10)		0.911[c]			1.096[c]
Civil obedience (0–10)					1.067[c]
Satisfaction and integration					
Attitudes					0.862[a]
Social isolation (0–10)					1.097[c]
Safe after dark (= yes)				0.464[b]	
Media use					
TV watching (minutes/day)					1.001[a]
Internet use (no access to every day = 0–7)		1.065[b]			1.051[c]

[a] $p < 0.001$
[b] $p < 0.01$
[c] $p < 0.05$
Only statistically significant parameters are reported ($p < 0.05$) for all variables, except for categorical variables where all parameters are reported when one of the categories has a significant effect on the response variable

(mid 1980s) researchers have increasingly traced measurement errors to the psychological processes of the respondent. The new developed cognitive models do not so much contradict the earlier models of error but supplement them and pay much more attention to the bias component. (Tourangeau et al. 2000, pp. 320–321).

Central to measurement error is the idea that there is no perfect relationship between the observed response and the underlying unobserved

variable that serves to generate the response (Borhnstedt 1983, p. 70). The obtained measure is a (linear) combination of the true value, the effect of the method used, and random error. A very general distinction in the literature of measurement error is between *reducers* and *measurers* (Groves 1989, p. 4). Applied to the survey instrument (the questionnaire), the *reducers* try to reduce error as much as possible by applying rules for appropriate phrasing of survey items. Examples of this approach are in the classical volumes on the art of asking questions and questionnaire design of Payne (1951), Sudman and Bradburn (1974, 1982), Converse and Presser (1986), Foddy (1993), and more recently in Saris and Gallhofer (2007, pp. 81–153). This work applied to questions in standardized questionnaires and guidelines are offered for different kinds of questions. Many of the suggestions and rules are based on research findings of the measurers on the effect of variations in question wording. Reducers and measurers are thus not necessarily different scholars, but different ways of dealing with survey questions and responses. The measurers try to measure measurement error in order to understand it, to identity the causes of error, or to obtain estimates (e.g. correlations) that are adjusted for measurement error.

Among the measurers one can distinguish several approaches that are discussed in the remaining part of this chapter. It starts with the findings, advantages, and disadvantages of measuring the effects of the wording and order of questions and response scales in the so-called "split ballot designs". The contribution of the CASM movement to the understanding of some of the major findings of split ballots is then discussed. After a critical appraisal of the split ballot approach we move to the modeling approach in which much attention is paid to the "*multitrait-multimethod*" design used to assess the reliability and validity of questions, and to the modeling of *response styles*.[14]

[14] The challenge of the cross-cultural equivalence of measurement models in not considered here. For this, we can refer to "*Cross Cultural Analysis. Methods and Applications*" (Davidov et al. 2011).

10.3.1 The Split Ballot Approach to Measurement Error

In the classical field experiments of the measurers, the so-called "split ballots", questions are formulated in different ways (Rugg 1941; Cantril 1944). The variants are administered to equivalent random samples of respondents and the response distributions are compared. Sometimes panel data is used, i.e., the same respondents are interviewed on two or more occasions with different question forms. The experiments are mostly carried out in the context of general social surveys in order to combine the testing of causal hypotheses through experimental manipulation (internal validity) with the ability to generalize to the national population (external validity) (Schuman and Presser 1981).[15] Not all findings about what can go wrong with questions come from split ballots. Some researchers like Belson (1981) systematically investigated the meaning of the words used in standardized questionnaire by asking respondents how they understood the questions. This method was adopted in later cognitive interviewing (Forsyth and Lessler 1991).

The most debated response effects deal with the following formal characteristics of questions: the length of questions, open and closed questions, yes/no-questions versus choice questions, contrasting questions, the use of arguments, questions with or without the middle-alternative, the "no opinion" option and the use of filters, the number of the answer categories, the order of the alternatives, the overall position of the question in the questionnaire, and the sequence of the questions (Molenaar 1986; Kalton and Schuman 1982). It seems that nearly every type of variation in question form and wording has an effect on response distributions.

10.3.1.1 Major Findings from Split Ballot Experiments

Let us start with a general view on the major findings from split ballot field experiments with

[15] A revised edition appeared in 1996, Thousand Oaks Ca.: Sage Pub, Inc.

variations in the wording and order of questions and response categories. For a short overview see the introductory chapter of Foddy (1993, pp. 2–10).

(1) Slight changes in the tone of wording produce major changes in the distribution of responses. This is in general the case with opinion questions, but answers to questions about the respondents' characteristics or about past behavior are also not completely protected from this (Foddy 1993, pp. 4–7). A well-known example of the effect of small variation is the difference in acceptance of policy measure depending on the verb "forbid" or not "allow" in a "yes or no" format (Schuman and Presser 1981, pp. 276–286). This kind of response effect finds a favorite application in the context of referenda with pairs of verbs as 'revoke/maintain', 'revoke/uphold', or 'facilitate/prevent' (Beckers and Billiet 2010). Holleman (2000) offered in her doctoral study a plausible (cognitive) explanation of this kind of response effects: there is a mental latent continuous scale behind the yes/no dichotomy and the respondent is moving on this scale depending on the terms used.

(2) The emotional and social pressure on the respondents to choose an answer that is in accordance with the suggested direction of persuasive questions can emerge from different formal attributes of a question. One-sided arguments, suggestive words, or unbalanced response categories are likely to change the response distribution in the direction of the persuasion (Schuman and Presser 1981, pp. 179–199; Molenaar 1986, p. 166).

(3) Respondents' answers to questions are sometimes affected by question format per se. Open questions without pre-listed response alternatives produce quite different results from closed questions, even when an exactly identical question text is used (Foddy 1993, pp. 7–8). Not seldom, less than 50% of the responses to offered categories in a closed question are mentioned on the closed form (Schuman and Presser 1979, p. 696). In a closed format, response alternatives may remind respondents of opinions that they may otherwise not consider in a free-recall task. Another explanation points to the possibility that the set of response categories affects the interpretation of the question's content. Presenting respondents a list of activities reduces the likelihood that respondents report activities that are not on it. Respondents may assume that the researcher is not interested in information that comes spontaneously in their mind but that is not included in the list (Schwarz and Hippler 1991, pp. 43–45).

(4) Responses to closed questions with the request to indicate for each listed option in a long list whether it applied or not, differs considerably from closed questions with the request to indicate from a list all options that apply. Response distributions over time (or between countries) are not comparable when the format has been changed. We found a striking example of this response effect in the 2008 *European Value Study* in Belgium. The shift in format happened in the field organization at the moment that the paper version of the questionnaire was transformed into an electronic version in view of CAPI. This shift in format resulted in serious underestimation of responses to questions containing lists of characteristics of a job, preferred qualities of children, and groups in the population the respondents do not want as neighbors. The results are no longer comparable with previous waves and with other countries that participated in EVS 2008 because of this shift in format (Billiet 2011).

(5) Even small changes in the number and wording of the response categories had serious effects on the response distributions (Schuman and Presser 1981, pp. 161–178). This is for example the case when a logical (and realistic) middle alternative is omitted (see later Table 10.2). This sensitivity of

response distributions for changes in the wording and form of response scales makes it indispensable to screen always the response categories when response distributions of (seemingly) identical questions are compared between groups over time.

(6) Response scales can provide not intended information, and may have an effect on the respondents' choice of a response category. If questions about the frequency of engagement in activities (by themselves or others) are asked (*how often did you…*) and a response scale of specific length is offered, then the respondents seem to use the range of the numeric values given on the rating scale when they select a (frequency) category. (Schwarz and Hippler 1991). This response effect was found in split ballot experiments with high and low range response scales, for example about the daily time of watching television. Apart from the meaning of "watching" television, the range of the scales informs the respondents about the likely answer. This results in much higher frequencies when a high range is offered. When respondents are asked to check the appropriate response category they may infer from the range of response alternatives that a particular form of behavior is typical or socially desirable (Schwarz et al. 1988; Schwarz and Hippler 1991, p. 48). Respondents assume probably that the researcher offers a fair scale and that the middle of the scale reflects the behavior of modal people while the extremes represent the "outliers". Verbal anchors of numeric response scales also may carry unintended information and affect the respondents' choice (O'Muircheartaigh et al. 1995).

(7) Respondents often answer questions even when it appears that they know very little about the topic (Foddy 1993, pp. 8–9). It has been found that virtually all the respondents chose substantive response options when the "Don't know" (DK) option is not offered, but a substantive number of respondents (up to 1/4) select that option when it is offered. This finding stimulated a lengthiness discussion on 'non-opinions' or 'non-attitudes'. Does a large portion of the respondents really have weak or even no opinions at all on most issues (Converse 1964)? Is it an indication of privacy protection? Or is the un-stability related to the fact that respondents have several opinions depending on the considerations they have in mind at the moment of their choice (Zaller 1992)? This response effect and its explanations relate to the finding that answers to attitude questions are very unstable over time.

(8) Respondents' reported opinions, beliefs, choices, habits, and interest often seem to be very unstable. A study of the reliability of survey measurement estimated the test–retest reliability of 41 attitude items from the *US National Elections Study* panels (1950–1970) to be 0.55 on average (Smith 1984). Abstract attitude questions are less stable than more concrete ones, and the lower educated and respondents with weak attitudes are most unstable. Latent multiple indicator variables are more stable than single items (Billiet et al. 2004). Several explanations are offered for the observed instability of attitude measurements. As was mentioned, Converse suggested that many respondents "simply do not have meaningful beliefs even on issues that have formed the basis for intense political controversy among the political elite for substantial periods of time" (Converse 1964, p. 245). Others contend that instrument unreliability and measurement error, rather than non-attitudes, were the cause of low correlations (Alwin 1992). It is argued that the underlying "true" attitudes are more stable (see Smith 1988 for an overview), and that the observed changes in responses to opinion questions are due to changing considerations (Zaller 1992, 42–51).

(9) Answers to earlier questions can affect respondents' answers to later questions.

These are so-called "context effects". The context of a question, that is the preceding question, and response scale, plays a crucial role in determining the meaning of that question. Preceding questions and responses can modify the meaning of the later questions, and responses to later question(s) are then consistent or in contrast with prior questions. A well-known example of a contrast effect is the change in acceptance of abortion in a more general context (less acceptable situation) after a very specific question dealing with a more acceptable situation. When the item with the acceptable reason for abortion is asked first, respondents may assume that the specific (acceptable) argument is not included in the subsequent general question and it then becomes easier to oppose the general item (Schuman and Presser 1981, pp. 36–39; Billiet et al. 1992, pp. 134–135).[16] The scope of the term "abortion" has changed because the most acceptable reason is mentally left out (mental subtraction). Previous questions not only evoke a change in the meaning of later questions, but they can also increase the saliency of a topic (Schuman and Presser 1981, pp. 44–47).

(10) Changing the order in which response options are presented affects respondents' answers. Primacy (choosing first option) or recency (choosing last option) effects appear depending on the mode of administration (face-to-face, mail, telephone) (McClendon 1991) and the use of response cards where visual display of the categories is offered (Sudman et al. 1996, pp. 138–142). The alternatives to a closed question are mostly asked in the same order for all respondents and the effect is only detected when for one or another reason, the order has been changed in comparable samples. Researchers then realize that the choices made by the respondents are artifacts of response order. It is therefore suggested to randomize the order of response categories in order to have at least an idea about the size of the effect.

(11) The cultural context in which a question is presented often has an impact on the way respondents interpret and answer questions. This context might be mediated by the main subject of the questionnaire, by the presence of the interviewer, and even by the ongoing debates in the media. Independent of the question context, different meanings may be differentially accessible to different respondents because of the frequency with which they employ them in daily life (Sudman et al. 1996, p.61).[17]

10.3.1.2 Initial Understanding of Question Wording Effects

In an attempt to understand the effects of question form on response marginal, Schuman and Presser (1981, pp. 298–301) suggested two concepts that are each applied to several classes of response effects. These are *question constraint* and *response persuasion*.

Question constraint applied to several of the findings that are mentioned before: omitting a donot know or a middle alternative (see Table 10.2), changing the number response categories, open versus closed questions, etc. In the context of survey interviewing with standardized questionnaires, respondents seem to accept the "rules of the game" as they are conveyed by the form of the questions, and they earnestly try to work within that framework once they have agreed to be interviewed. The

[16] The standard interview situation, where it is assumed that the interviewer does not ask the same question twice (the '*given new contract*' rule) is a favorable context for this kind of context effects.

[17] In a survey in the 1980s, we have observed that some respondents offered numbers between 20 and 30 in response to the question "*How many children do you have*?". These respondents were school teachers who interpreted the question to refer to the children in their classes, the meaning that was most accessible in their memories (cited in Bradburn, 1992, p. 317). That question was in the last part on the household characteristics of teachers in a self-administered questionnaire about their attitude toward mixed (qua gender) classes.

Table 10.4 Response distributions of two questions on in-group favoritism in two forms (split ballot: with and without a middle alternative) in two random samples of the Flemish population in 1996 (in percentages)

Form A (no MA included)		Form B (explicit MA included)	
"Suppose that there are two employees, one is Flemish, the other belongs to an ethnic minority. Both are equal on all characteristics. When one of them has to be dismissed because the company's financial situation, who should be dismissed according to you, the Flemish or the employee belonging to an ethnic minority?"		"Suppose that there are two employees, one is Flemish, the other belongs to an ethnic minority. Both are equal on all characteristics. When one of them has to be dismissed because the company's financial situation, who should be dismissed according to you, the Flemish or the employee belonging to an ethnic minority? *Or should this made no difference?*"	
The employee of ethnic minority	45.1	The employee of ethnic minority	25.9
The Flemish employee	7.4	The Flemish employee	6.4
Should not made any difference*	39.5	Should not made any difference	65.1
No opinion*	7.9	No opinion*	2.6
N (100%)	532	N (100%)	455
$\chi^2 = 70.087$; df = 3; $p = 0.001$			
"Suppose that there are two employees, one is Flemish, the other belongs to an ethnic minority. Both are equal on all characteristics. When one of them have promotion prospects who should then according to you get this promotion, the Flemish or the employee belonging to an ethnic minority?"		"Suppose that there are two employees, one is Flemish, the other belongs to an ethnic minority. Both are equal on all characteristics. When one of them have promotion prospects who should then according to you get this promotion, the Flemish or the employee belonging to an ethnic minority? *Or should this made no difference?*"	
The employee of ethnic minority	4.0	The employee of ethnic minority	0.8
The Flemish employee	47.1	The Flemish employee	31.7
Should not made any difference*	43.0	Should not made any difference	65.5
No opinion*	6.0	No opinion	2.0
N (100%)	532	*N* (100%)	455
$\chi^2 = 57.195$; df = 3; $p = 0.001$			

[a] Not included in the question text but spontaneously answered

constraint imposed by the question is very powerful, and the researchers' decision to omit an alternative, or to modify the wording, has considerable influence in shaping the resulting response distributions (Schuman 1984, p. 34). At first glance survey research practice may find a rule is this understanding of response effects: remove the constraint in questions. It is however not that simple as was obvious in the discussion of Table 10.4. There is even not one simple rule for the DK filter; If one wants to measure clearly informed opinions, then one may include an implicit DK filter, but if one wants to obtain the responses to a set of indicators (items) that is intended to measure an underlying disposition (or a latent attitude variable), then the DK filter is not recommended (see Schuman and Presser 1981, pp. 147–160).

Response persuasion deals with the persuasiveness of the question text or with response categories or scale. Emotional and social pressure on the respondents to choose an answer that is in accordance with the suggested direction can emerge from different attributes of a question: questions with one-sided arguments; clauses and emotionally loaded terms in a question; reference to opinions of others; answer categories that express positive judgments; unbalanced questions (i.e. no explicit mention of a negative alternative of a statement). A general rule for survey practitioners is more evident here: never deliberately load questions. But there are even exceptions in this case. It is sometimes suggested to use leading questions in case of expected underreporting of social undesirable responses to factual questions, albeit with care

(Sudman and Bradburn 1982). Experimental variation of (counter-) arguments in repeated measurements (in web surveys) may be used for measuring the stability and crystallization of opinions, and should be distinguished from deception (see Sniderman and Grob 1996).

Several response effects that were listed but are not covered by these two concepts received full attention in the CASM movement. This is the case of *unintended information* provided by the researcher through the response scales. Respondents use the information in the response scales to interpret the meaning of a question, and to find out what kind of information the interviewer wants, or what is the 'correct' answer.

Context effects became in the 1980s a favorite subject for theory-based cognitive experiments with survey questions in which interpretations of these effects were tested (Schwarz and Sudman 1992, pp. 28–44; Sudman et al. 1960). These are considered as shifts of the respondents' *frame of reference*. The context plays a crucial role in determining the meaning of a question. Preceding questions can modify the scope of subsequent questions (e.g. the abortion questions). Previous questions may also make some responses to later questions more accessible (by priming), or may increase the saliency of questions (Tourangeau 1992).

10.3.1.3 Social Cognition and Question Wording Effects

Schuman and Presser reacted initially very skeptically to the recommendation that research on question wording should draw its theory from cognitive and social psychology. They suggested that a substantial contribution to survey research would be made only if those disciplines become directly involved in studying survey data, and formulate problems that arise directly from the data, methods, and ideas of survey researchers (Schuman and Presser 1981, pp. 313–314). A number of social and cognitive psychologists accepted the challenge (Jabine et al. 1984).

Considerable progress in understanding the cognitive and communicative processes underlying survey responses has been made over the past 25 years (Schwarz et al. 2008). Scholars in social cognition tried to "find the substantive meat in what others may well regard as methodological poison" (Tourangeau and Rasinski 1988, p. 301). The standardized question situations of surveys offer excellent databases to psychologists who studied the cognitive determinants of responses to questions. In such a restricted conversational interaction that excludes negotiation about the intended meaning of questions, the respondents rely to a large degree on previous questions and answers (the context), on the formal characteristics of the target question, and—in case of face-to-face interviewing—on the instructions and cues provided by the interviewer as sources of information (Strack et al. 1991, pp. 112–113).

Several theoretical models of the cognitive and communicative processes underlying responses to survey questions have been proposed in collaborative research programs of (cognitive) psychologists and survey methodologists (Biderman 1980; Hippler et al. 1987; Schwarz and Sudman 1992). Several volumes are published in which the results of question wording experiments are theoretically grounded in the social cognition tradition. "*Thinking about Answers*" (Sudman et al. 1996) and "*The Psychology or Survey Response*" (Tourangeau et al. 2000) offer comprehensive syntheses of the contribution of (cognitive) psychology to the understanding of survey response. This work mainly relates to questionnaires in face-to-face or telephone interviews. More recent developments apply to web surveys (Couper 2001; Dillman 2001; Heerwegh 2003), and even surveys by mobile phones (Lynn and Kaminska 2011).

According to the model of information processing in a survey interview (Strack and Martin 1987), respondents need first (1) to *understand* the question and determine which information they are asked to provide. Further, they need to (2) *recall* the relevant information from memory. In case of an opinion question, they will rarely find a ready-for-use answer stored in memory. They need to (3) form a *judgment* on the spot based on the information that comes to mind at that time and that is considered relevant and useful for answering the question. In case of factual questions on events or on past behavior,

respondents need to retrieve relevant episodes. Once a judgment is formed or relevant episode is retrieved, they rarely report it in their own words but they need (4) to *format* it in the response alternatives offered. Respondents may however hesitate to (5) communicate (*editing*) their 'internal' response because of social desirability and self-presentation (Schwarz et al. 2008, pp. 18–19). The response effects that we discussed are related to each of these five tasks and can be explained by several cognitive psychological processes in interaction with the instrument (questionnaire), and the communication context of the interview (e.g. priming, information accessibility, the inclusion/exclusion of information,). "*Satisficing*" occurs when respondents "shortcut the cognitive processes necessary for generating optimal answers" (Krosnick et al. 1996, pp. 29–31). When a respondent works through each of the five mental steps with a maximum level of effort, they are optimizing their answers.

10.3.1.4 Response Effects as Useful Data: An Appealing Example

A pair of questions on in-group favoritism was included in a survey on political knowledge organized by the *Institute for Political Opinion Research* (ISPO), in a random sample of the general population in Flanders in 1996 ($N = 987$). It was decided to split up the sample into two random parts according to the absence (Form A) or the presence (Form B) of the middle alternative (MA) in questions and response cards. The response distributions of the discrimination questions in the two subsamples are in Table 10.4.

Comparison of the response distributions in the two forms suggests that not all respondents who want to dismiss (45.1%) the employer of an ethnic minority or offer promotion to the Flemish employee (47.1%) have a fixed opinion. We are not sure since these are split samples and not repeated measurement, but this conclusion is very likely. The percentage of respondents who are ready to give preferential treatment to a person of their own ethnic group is 15–20% points lower when the option "*should not made any difference*" is offered.

The conclusion about the degree of favoritism is very different depending on the response scale used. What is preferred in case of these questions? With the concept "*question constraint*" in mind the middle alternative (Form B) might be preferred since it is a reasonable and 'real' option. Unfortunately, we miss then the insight that a substantial amount of respondents has no fixed opinions on the issue. Actually, there is no general rule about the preferred question in this case. It depends on what one wants to measure. If we want to find out what fraction of the population firmly rejects by themselves favoritism toward their own ethnic group, then we should use Form A. If one however wants to detect what fraction in the population holds fixed opinions in favor of their own ethnic group, then we should use Form B. Actually, in substantive research one should use both in split samples.

The lesson learned by this example is that one should not consider this kind of response effects as error but as additional data that can be used for interpretation of the responses, and for better understanding of the measurements (see Schuman 1982; Schuman and Ludwig 1983). This example also illustrates the possibility of using split ballots in substantive research as a way of providing the researchers with useful data for interpreting the meaning of responses to questions that are crucial in their research (Schuman and Presser 1981, p. 313; Schuman 1984, p. 130).

One can even improve the spit ballot design by combining it with repeated measurements. This enables identification of the 'movers' (unstable respondents) according to other variables in the questionnaire (education, political position, age, etc.) if one may (reasonable)[18] assume that the two questions are congeneric[19]

[18] Because of a strong correlation between the responses to the question on dismissal and the question on promotion.

[19] Congeneric measures are defined as measures of the same underlying construct in possibly different units of measurement or possibly different degrees of precision (Lord and Novick 1968, pp. 47–50). The congeneric measures are indicators of different characteristics that are all simple linear functions of a single construct (Groves 1989, p. 18).

measures of the same underlying latent variable. It is possible to cross the forms over the two subsamples. Each sample contains then the A and the B forms, but it is important to have the B form with MA second in the two samples, in order to avoid priming of the middle alternative and creating that way a consistency effect.

10.3.1.5 Critical Reflections on the Split Ballot Approach

The split ballot experiment is a strong design for the assessment of a question wording effect. However, this approach had some shortcomings. A finding holds only for a specific question involved in the experiment. To be able to generalize a particular wording effect to other questions, many similar experiments on the same wording variable would be required, the results of which could then be summarized and tested for significance. The number of investigations is not large enough to warrant general statements about the effects of particular wordings (Molenaar 1986, p. 161). Moreover, it is not always clear what the most valid wording is. We have seen in our example of in-group favoritism that it depends on the research question in a substantive research context.

As an alternative, Molenaar (1986) worked out a *non-experimental design* that consists of a meta-analysis of hundreds of questions and distributions according to formal characteristics of questions. He also paid serious attention to linguistic characteristics of questions. One should realize that the critical view of Molenaar on the split ballot experiments apply to these in the period before the fertilization of these experiments by the insights from cognitive psychologists. The findings about particular question wording effects became more general in the 1980s by enriching these by their theoretical insights in cognitive processes, and by setting up theory-driven experiments (see Sudman et al. 1996).

An important development that changed the nature of split ballot experiments, and that overcomes the ad hoc and a theoretical character of the early split ballots was the introduction of computer-assisted interviewing and the possibility of web surveys. Sniderman and Grob (1996) argued that the introduction of computer-assisted interviewing (in whatever mode) has changed the opportunities of the classical split ballots considerably. A new wave of experimentation in survey research was made possible by the introduction of computer-assisted interviewing and surveys via the Web. This new wave of experimentation is characterized by a double shift: from fixed and predominantly methodological to variable forms and predominantly substantive focus (Sniderman and Grob 1996, p. 380). Randomization of samples and several versions of questionnaires became much easier (Sniderman and Grob 1996, p. 379). The computer-assisted experimental design has been applied in large-scale general population samples. Questions as "how stable are these opinions for which subgroups in a population?" or "how likely are these to change when some considerations are primed?" can now be answered. The effects of several relevant considerations (and combinations) can be studied by introducing variation in the framing of the choice presented to respondents, and variation in the context of questions (Sniderman and Grob 1996, p. 393). Researchers who apply this strategy should be very open-eyed not to cross the fine line between experimentation and manipulation, and to avoid deception of respondents.

10.3.2 Modeling Measurement Error

The initial models used for quality assessment of measurement instruments are the *Quasi Simplex* (QS) approach and the *Repeated Multimethod Model* (RMM) approach (Saris and Andrews 1991, pp. 575–548). The QS approach originated from Heise's (1969) idea that test–retest studies could be the most appropriate approach for evaluating survey questions in panel research in which traits were *repeatedly* measured with the *same method*, and was used by Alwin (1989) within the framework of error sources in survey data. The QS procedure was, under certain assumptions, appropriate to assess the unreliability of measures, but not for detecting invalidity in measurement instruments because only

one method was used (Saris and Andrews 1991). Jöreskog (1971) proposed the RMM approach and suggested that measurement instruments could be evaluated with respect to validity by a specific *congeneric*[20] test model.

10.3.2.1 How to Disentangle Measurement Validity and Invalidity

Central to the RMM approach is the notion of an unobserved theoretical construct that the researcher attempts to measure by means of observed indicators (questions) characterized by specific response scales. Validity refers thus explicitly to the extent to which an observed measure reflects the underlying theoretical construct that the investigator intends to measure (Cronbach and Meehl 1955, p. 282). Measurement instruments that reflect only weakly the underlying (and intended) constructs because they are systematically affected by other sources of variation are invalid. Reliability is associated with the stability of measures (given that there is no real change). It is perfectly possible to obtain stable measures of measures that are invalid.[21] For that reason it is recommended not to use mathematical equations that express validity completely in function of reliability. This recommendation is applied in Saris version of the "true score" MTMM model that is shown in the Appendix (Saris 1995) in which it is, contrary to the classic "true score" model (see Lord and Novick 1968), possible to estimate reliability independently of validity when this model is used in a *multitrait-multimethod* context (MTMM).

The MTMM design was at the end of the 1950s developed by Campbell and Fiske (1959) who focused mainly on the assessment of validity. They suggested measuring each of a number of traits with a number of *different* methods in order to build a MTMM correlation matrix. The analysis technique they developed was largely qualitative and provided a raw indication of the '*convergent*' and the '*discriminant*' validity of the MTMM measures in view of potential method effects. In reaction to problems that emerged from Campbell and Fiske's strategy of analysis of MTMM correlation matrices, a path-analytic approach to the MTMM matrix introduced the idea of decomposition of the correlations (or covariance) between the measurements (Kallenberg and Kluegel 1975, pp. 2–4) in terms of validity, invalidity, and unreliability parameters. The development of *Structural Equation Modeling* (SEM) software facilitated the application of these models, and Confirmatory Factor Analysis (CFA) became the main method for analyzing MTMM data (Schmitt and Stults 1986). This procedure was most appropriate for evaluating the validity and reliability of measurement instruments; it provides numerical estimates of these measurement quality parameters (Saris and Andrews 1991). Andrews and his colleagues of the *Survey Research Center* provided a convincing account of the advantages of MTMM models for survey research (Andrews and Withey 1976; Andrews and Crandall 1976). The proposed repeated measurements of several traits with different methods in order to obtain estimates of the measurement quality of survey data, and to arrive at 'rules' about how survey questions should best be worded (Andrews 1984).

In the classical MTMM matrix, an observed score y_{ij} is linear decomposed into three unobserved components[22]:

$$y_{ij} = v_i F_i + m_j M_j + e_{ij},$$

where y_{ij} is a measure of ith trait by jth method, F_i is the ith trait, M_j the jth method factor, and e_{ij} is random (residual) error. The slope parameter v_i can be considered the validity coefficients, and the slope parameter m_j is the invalidity coefficients, This formulation makes explicit that

[20] See footnote n°7.

[21] Take, for example, an IQ test in Dutch intended to measure the IQ of French-speaking students. The results may be very stable but it measures very probably the students' knowledge of Dutch, and not their IQ.

[22] The symbols τ, η, and ε are altered in F, M, and e, and the slope parameters λ are replaced by v and m in order to be consistent with Saris' true score model which is shown in the appendix. (Saris 1990b, p. 119).

some of the reliable variance is actually invalid since it is induced by the measurement instrument (or scale) used (Alwin 1997, p. 328). The observed (co)variances in an observed MTMM matrix are thus conceived as the result of two kinds of underlying latent variables: traits and method factors. The residual variance is not explained by the common trait and method factors.

One should notice that in this linear regression approach[23] to measurement the variances in the observed indicators are conceived as 'caused' by the variance in a latent variable, which is conceived as the common source. The strength of a relation between an observed indicator and the latent variable behind indicates "measurement validity" of this indicator. This does not coincide with "theoretical validity" (or construct validity) that relates to the not directly measurable relation between the latent variable and the intended concept.

This MTMM approach was further developed by Saris[24] and his research team in an MTMM project that intended to develop strategies for evaluating the quality of survey measures (Saris 1990a, pp. 52–80). An MTMM 'true score' model with measurements of at least three traits each measured by at least three different methods was proposed, and problems with the design were discussed (Saris 1990b, pp. 118–129; De Wit 1994; Scherpenzeel 1995; De Wit and Billiet 1995). The modified MTMM 'true score' was developed in order to be able to estimate the reliability parameters independently from the residual variance (see Saris and Andrews 1991).

A 2 × 2 MTMM true score model is shown in the Appendix. An additional latent variable T is introduced in the true score model. This allows an independent estimation of the reliability parameter r. There are nine true scores estimated in a full MTMM model (one true score for each trait-method combination). The validity coefficients v are, as is the case in a classical MTMM model, no longer given by the direct effects of the trait factors on the observed variables, but by the effects of the trait factors on the true scores. Similarly, the in-validity coefficients m (method effects) are the effects of the method factors on the true scores and not on the observed scores (Scherpenzeel and Saris 1997, p. 346; Saris 1990b, pp. 118–119).

10.3.2.2 How to Adjust Measures for Invalidity

There are two ways for adjusting the correlations between traits for measurement invalidity and unreliability, depending on the question whether the quality estimates are directly available in the collected data or not. In the latter, one must derive the estimates from other sources (Andrews 1984, 1990; Scherpenzeel and Saris 1997).

Researchers who repeatedly measured several traits with different response scales, and who analyzed this MTMM data as is the case in ESS, directly obtain correlations between the traits $\rho(F_{ij}, F_{ij})$ that are adjusted for unreliability and for effects of the methods used. They can apply the most reliable and valid measurement of the traits (see Appendix). In ESS the MTMM design is combined with a (randomized) split ballot design in order to diminish the burden on interviewers and respondents of repeated measurements. Each respondent only received two versions of the experimental questions. This design nevertheless allows the estimation of a full MTMM matrix (Saris et al. 2004). The main problem is the very limited applicability of this direct method because repeated measurements are only available for a small number of measured traits (or variables) in a survey.

Researchers who cannot rely on direct MTMM-based validity and reliability estimates

[23] Contrary to standard linear regression, the parameter on the left side of the equation is a measured quantity and is expressed in function of un-measured variables on the right side of the equation. For that reason, a set of equations need to be solved in order to obtain values of the regression parameters in covariance structure analysis. The number of unknowns is reduced by the introduction of constraints between the parameters (Bollen 1989 pp. 16–23). These constraints are based on theoretical ideas, e.g., equality of slopes between indicators and a common method factor.

[24] Saris established in 1989 the *International Research Group on Measurement Error in Comparative Survey research* (IRMCS).

are advised to publish their correction coefficients, which are based on a meta-analysis of a large number of MTMM studies (Scherpenzeel and Saris 1997). According to these scholars, the average reliability and validity of a measurement should be adjusted depending on the effects of a number of characteristics of the actually used measurement instrument (i.e., subject or domain of the study, response scale, data collection mode, position in the questionnaire, language group, etc.). For development of the core questionnaire and the rotating models, ESS applied the *Survey Quality Predictor* (SQP) program[25] (Saris and Gallhofer 2007) to estimate the reliability and validity of the proposed question wordings and response scales. SQP is based on the database of quality estimates that are derived from the growing number of MTMM studies it reports the expected validities and reliabilities of the planned questions.

10.3.3 Measurement of Response Styles

Response style refers to a systematic tendency to respond to a range of survey items on some basis other than what the items were specifically designed to measure (Paulhus 1991). It is a kind of measurement error typical for sets of items characterized by the same format, which is typical for multiple indicators (*congeneric* measures) that measure a latent variable. Four kinds of response styles are distinguished: (1) Agreeing (ARS): the tendency to endorse assertions independently from their content called "acquiescence"; (2) Denying (DRS): the tendency to deny assertions independently from their content (DRS) which is the opposite of yes-saying; (3) Extreme responding (ERS): the tendency to choose extreme response categories of response scales independently from the content of items (ERS); (4) Midpoint responding (MRS): the tendency to choose the middle of a response scale independently from their content. Response styles might be related to satisficing.

The main challenge in assessing response style is to measure it *independently* from content. Weijters (2006) distinguished two dimensions in the classification of response style measures: the status of the items on which the response measures are based, and the treatment of the content. According to the *status*, one can distinguish multi-functional measures that measure both content and style by a common set of measures, and specific measures that only measure style. There are four ways of *treating* the content in response style measures: (a) the content is not deliberatively planned or selected before data collection (no specific ex ante control); (b) the content has been eliminated with the aim to measure style; (c) the content is manipulated (e.g. opposite meanings in sets of items) in order to cancel out the effect of content; (d) content is randomized in such a way that systematic influence of content on response is hindered. The classification of *status* by *treatment* does not lead to eight realistic kinds of response style measures since some combinations are per definition impossible (Weijters 2006, pp. 33–43).[26]

There are attempts to develop content free items and to measure ARS by specific measures, but these were not very effective (Rorer 1965). Random selection of items in view of specific style measures is often used in marketing research. Some scholars (Baumgartner and Steenkamp 2001; Weijters 2006), follow the method proposed by Greenleaf (1992a, b), and make use of a set of items that is maximally heterogeneous in content. The basic idea behind this approach is to reduce the effect of content in the set of items to random noise: if all the items represent different constructs that are unrelated,

[25] See http://www.sqp.nl/media/sqp-1.0/.

[26] E.g. the elimination of content in a multi-functional measure, or the randomization of content in multi-functional items that are intended to measure content and style. If one eliminates the content then there is no longer a multi-functional measure, and the selection of items to measure content is per definition not random.

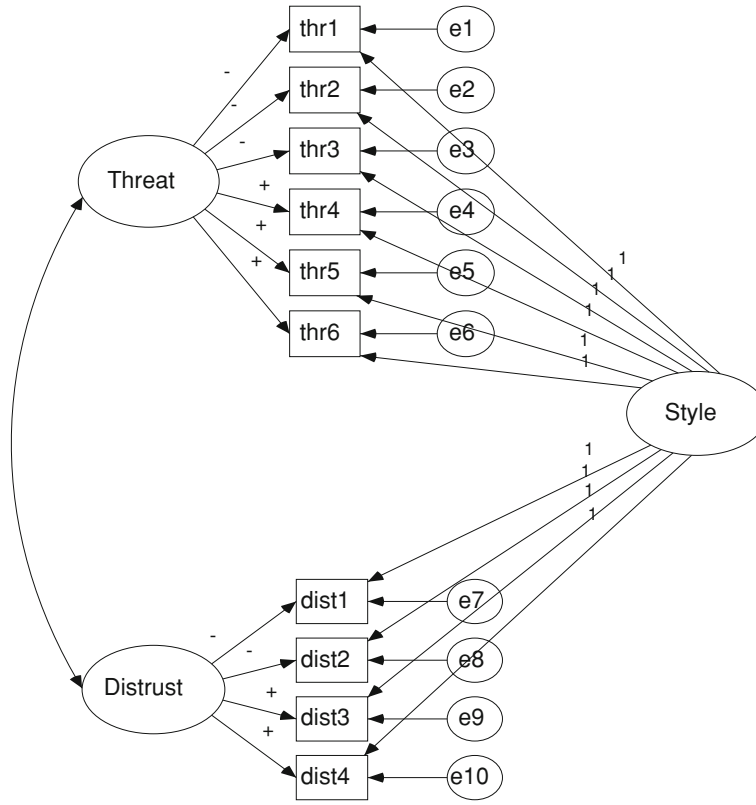


Fig. 10.1 Measurement model for two interrelated content factors and a style factor

it can be expected that there is no consistency in responses other than that induced by response styles (Weijters 2006, p. 42). Response styles are then measured by the sum of items characterized by specific answers (agreeing, disagreeing, extreme responses, selecting the middle of response scales). The disadvantage of this approach is that a large extra set of unrelated items is needed in order to account for a style effect. The advantage is that it applies to the four response styles.

Multi-functional items combined with experimental control are possible for measuring ARS and DRS when balanced sets are used to measure content. Following a study by Mirowsky and Ross (1991) and a proposal by McClendon (1992), Billiet and McClendon (2000) measured acquiescence for two balanced sets of items. They applied a model with two content factors (ethnic threat and distrust in politics) each measured by a balanced set of items, and an additional style factor (acquiescence) which loaded equally on all items.

This model yielded a significantly better fit than a model with only two content factors (see Fig. 10.1[27] for a formal visualization) or other model specifications.[28]

As was expected, the two content factors did not correlate with the style factor. This confirms the theoretical postulation about a zero relation between style and content. The study provided a strong support for the existence of acquiescence since the latent variable correlated very strongly

[27] The model contains two content factors, Threat and Distrust, with balanced sets of six and four observed indicators, respectively, and with identical slopes for the style factors (denoted 1). For actual parameter values, see Billiet and McClendon 2000, pp. 622–625).

[28] Models with two positive and two negative factors, or models with a separate style factor per content.

with the sum of agreements,[29] and was replicated in different populations (Billiet et al. 2002).

This approach does not need a separate measurement of style and, contrary to MTMM, no repeated measurements with identical items with different response scales. Content and style are clearly separated in the modeling. The conditions are that at least one more or less congeneric set of indicators is used to measure a content variable, and that the items have a specific response scale (Likert scales, or any "agree–disagree" scales). This method is however only applicable to measure ARS (or eventually DRS) and not applicable to the other styles. The method is also applicable to response scales that are considered categorical (or ordinal) in the context of latent class analysis (Moors 2004).[30]

10.4 Concluding Discussion

This chapter focused on some aspects of two sources of systematic error in surveys: non-response error and measurement error. The structure of this chapter in two parts reflects the separate development of two research traditions, the one on errors of non-observation, the other on observation error. In the first part, only the quality of the obtained response was discussed with focus on non-response reduction and bias estimation. Another error source of non-observation, non-coverage of part of the population in the sampling plan, was not touched. In this part, both the reducers approach (response enhancement) and the measurers (estimation of non-response and bias) were discussed. The response enhancement section was mainly, but not exclusively, covered by process evaluation, while the non-response and bias section was largely an aspect of the output evaluation of survey quality. The scope of the second part on measurement error was in a way much smaller: it dealt only with systematic measurement error as far as the source of it was the instrument (questions in a standardized questionnaire). Errors induced by the respondent were not directly discussed in this part, e.g., the observation that a number of systematic errors relate to education level because of the cognitive sophistication required for providing answers to survey questions. The same holds for the bulk of findings on interviewer effects and interviewer variance. The focus was thus nearly exclusively on question wording, order, and response scale effects. In the part on measurement error the reducers or 'process' approach (i.e. *how to prepare* '*good*' *questions?*) was not touched. This topic, however, was covered in detail by previous chapters in this handbook (Chaps. 7, 8).

This concluding discussion throws light upon the most actual developments in the domain of response and measurement error and on the relation between response and measurement quality.

Concerning measurement error, the development of measurement models that include both theoretical relevant substantive concepts and method or style effects is still in progress (see Davidov et al. 2011). The main challenge is the inclusion of the recording of method and style effects in substantive questionnaires without placing too much burden on the respondents. At this point, the use of repeated measurements and split samples are now more realistic options in the context of web surveys on substantive issues (see Sniderman and Grob 1996). For an application of this approach in the domain of ethnic prejudice, see Sniderman and Hagendoorn (2009).

In the domain of non-response error and bias, several developments need to be mentioned. The first development has been mentioned and deals with the shift in focus from response enhancement into bias reduction. It means that in the context of survey quality assessment one should not only publish correct response rates but also indications of bias (e.g. the Representativity indicator or R-indicator). That is on the output side. It means

[29] In order to obtain validation, Billiet and McClendon (2000, pp. 623–626) investigated the relation between the style factor and a variable that measures the sum of agreements (named "scoring for acquiescence"), across a balanced set of 14 positively and negatively worded items. This procedure led to a strong correlation of 0.90 ($t = 22.26$) between "scoring for acquiescence" and the style factor in a random sample of Flemish voters in Belgium. This finding was replicated several times in other samples.

[30] One can model MRS and ERS with latent class analysis by specifying a latent class of respondents that are very likely to choose the middle of scales or extreme response categories.

during the process of data collection that, in line with a responsive design (Groves and Herringa 2006), the completion of the planned sample is permanently fed by information about response expected propensities according to demographic and territorial variables (Schouten et al. 2009).

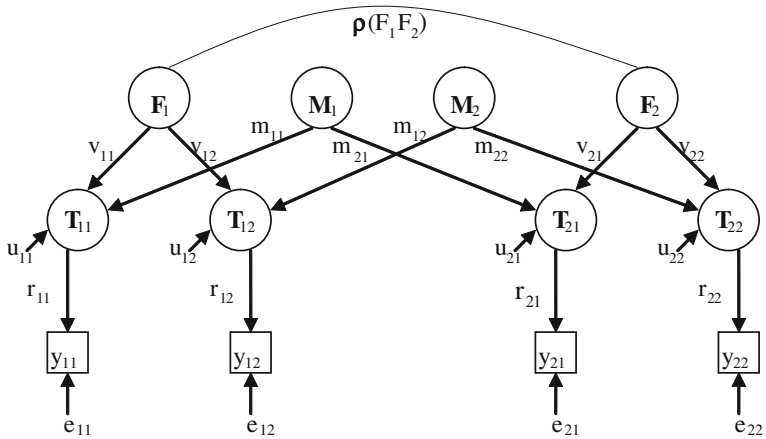
Adjustment for bias by means of propensity score weightings is an alternative approach. The success of both approaches (R-indicator and adjustment by weighting) depends strongly on the availability of information about variables that are related with the target variables. This information might be at both the respondent level, or at the context level. The likelihood that such information is available depends on the accessibility of databases, which is more likely at group (context) level than at individual level. If relevant context data is available, hierarchical models might be useful. Another future development is the collection of paradata in view of the adjustment of biased sample statistics (Kreuter et al. 2010).

Future research should also focus on sequential mixed-mode designs that are developed to reduce both non-response and survey costs (Dillman et al. 2009). Researchers start with the cheapest mode (e.g. mail or Internet) and then switch to telephone and finally to the most expensive mode (face-to-face) (see Chap.21 by Shine and Dulisse).

Contrary to the impression one may have at bases of separate treatment of response and measurement error in this chapter, both are actually related since both error components have an effect on the response distributions. This leads mostly to under-, or overestimation of means and percentages of specific categories of variables, and it may in principle effect correlations or regression parameters. Our experience with ESS learns that measurement error mostly has stronger effects than non-response bias, in particular in case of 'true' scores of subjective variables (platonic true values) as opinions, attitudes, values, and choices. Furthermore, non-response and measurement error are related. Practices to reduce survey non-response may bring into the respondent pool more respondents that provide data filled with measurement error (Olson 2006). This is presumably because reluctant respondents are more likely to satisfice and due to their lower cognitive ability (Kaminska et al. 2011). Moreover, in the actual discussion on the implementation of sequential mixed mode surveys the likelihood of mode effects, a kind of measurement error, is real (Dillman et al. 2009). Researchers who are concerned with response enhancement in surveys should also pay attention to error-sensitive characteristics of additional respondents.

10.5 Appendix

The MTMM true-score model for two correlated variables measured with two different methods (response scales), and with unique components and random error



y_{i1} and y_{i2}: observed indicator of i measured with methods 1 and 2;
F_1 and F_2: state of latent variables 1 and 2 which are correlated;
M_1 and M_2: latent method factors 1 and 2 that are assumed un-correlated;
T_{i1} and T_{i2}: true score of variable Fi repeatedly measured with methods 1 and 2;
e_{i1} and e_{i2}: random measurement error for item i measured with methods 1 and 2;
r_{i1} and r_{i2}: the reliability coefficients for item i measured with methods 1 and 2; the square of these are estimates of the test–retest reliability in case of repeated identical methods)
v_{i1} and v_{i2}: the true score validity coefficients for item i measured with methods 1 and 2; ${v_{i1}}^2$ (or ${v_{i2}}^2$) is the variance in T_{i1} (or the T_{i2}) explained by the variable F_i that one intends to measure;
m_{i1} and m_{i2}: the method effect on the true score of methods 1 and 2; ${m_{i1}}^2$ (or ${m_{i21}}^2$) is together with u_{i1} (or u_{i2}) the part in T_{i1} (or T_{i2}) not explained by the intended variable, and thus invalidity;
u_{i1} and u_{i2}: the residual variance of the true score of item i measured with methods 1 and 2;
$\rho(F_1F_2)$: the correlation between the two latent (intended) variables

$$y_{ij} = r_{ij}T_{ij} + e_{ij} \tag{10.1}$$

$$T_{ij} = v_{ij}F_i + m_{ij}M_j + u_{ij} \tag{10.2}$$

This model is a simplified version of the true score MTMM model used. It is in reality not possible to estimate a '2(traits) × 2(methods)' model. One needs at least three traits and three methods. But even then is it not possible to estimate the parameters, unless certain assumptions are made. Most simple (and acceptable) is to assume that the unique residual variances of the true scores (u_{ij}) are zero (Saris 1990b, p. 119; Scherpenzeel and Saris 1997, pp. 344–347).

In line with Eqs (10.1) and (10.2), and assuming zero unique variance $u_{ij,}$ the true score model is written as:

$$y_{ij} = r_{ij} \times v_{ij}F_i + r_{ij} \times m_{ij}M_j + e_i \tag{10.3}$$

Path analysis suggests that the observed correlation between the measures of two traits, both measured with method M_1, is (Scherpenzeel and Saris 1997, pp. 372–373):

$$\mathrm{Corr}(y_{11}, y_{2,1}) = v_{11} \times r_{11} \times \rho(F_1, F_2) \times v_2 \times r_{21} + r_{11} \times m_{11} \times m_{21} \times r_{21} \tag{10.4}$$

From which one can derive:

$$\mathrm{r}(F_1F_2) = [\mathrm{cor}(y_{11}y_{21}) - (r_{11} \times m_{11} \times m_{21} \times r_{21})]/(v_{11} \times r_{11} \times v_2 \times r_{21}) \tag{10.5}$$

This expression is useful for adjusting observed correlations between variables when one can rely on validity, reliability, and invalidity estimates (e.g. based on meta-analysis of MTMM data).

References

AAPOR. (2006). *Standard Definitions. Final Dispositions of case Codes and Outcome rates for Surveys* (pp. 46). American Association for Public Opinion Research: http://www.aapor.org/pdfs/standarddefs_4.pdf)

Alwin, D. F. (1989). Problems in the estimation and interpretation of the reliability of survey data. *Quality & Quantity, 23*(3–4), 277–331.

Alwin, D. F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. In P. V. Marsden (Ed.), *Sociological Methodology 1992* (pp. 83–118). San Francisco: Jossey Bass.

Alwin, D. F. (1997). Feeling thermometers versus 7-point scales. Which are better? *Sociological Methods & Research, 25*(3), 318–340.

Andrews, F. M. (1990). Some observations on meta-analysis of MTMM studies. In W. E. Saris & A. Van Meurs (Eds.), *Evaluation of measurement instruments by meta-analysis of multitrait-multimethod studies* (pp. 15–51). North Holland: Amsterdam.

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly, 48*(2), 409–442.

Andrews, F. M., & Crandall, R. (1976). The validity of measures of self-reported well-being. *Social Indicators Research, 3*(1), 1–19.

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being. Americans' perceptions of life quality*. New York: Plenum Press.

Baumgartner, H., & Steenkamp, J. -B. E. M. (2001). Response styles in marketing research: A cross national investigation. *Journal of Marketing Research, 38*(2), 143–156.

Belson, W. A. (1981). *Design and understanding of survey questions*. Aldershot: Gower.

Beckers, M., & Billiet, J. (2010). Uphold or revoke? A study of question wording in twelve municipal plebiscites in Flanders. *World Political Science Review*, 6(1), article 5, p. 29: http://www.bepress.com/wpsr/vol6/iss1/art5/

Biderman, A. (1980). *Report of a workshop on applying cognitive psychology to recall problems of the national crime survey*. Washington, DC: Bureau of Social Research.

Biemer, P., & Caspar, R. (1994). Continuous quality improvement for survey operations: Some general principles and applications. *Journal of official statistics, 10*(3), 307–326.

Biemer, P. P. (2010). Total survey error, design, implementation and evaluation. *Public Opinion Quarterly, 74*(5), 817–848.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in survey*. New York: Wiley.

Billiet, J. (2011). Godsdienstige betrokkenheid en de houding tegenover vreemden: Een verdwijnend verband? In K. Abts, K. Dobbelaere & L. Voyé (Eds.), *Nieuwe tijden, nieuwe mensen. Belgen over arbeid, gezin, ethiek, religie en politiek*, Chapt. X (pp. 215–243). Leuven: Lannoo campus & Koning Boudewijnstichting.

Billiet, J., Waterplas, L., & Loosveldt, G. (1992). Context effects as substantive data in social surveys. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 131–147). New York: Springer.

Billiet, J., & McClendon, J. M. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608–628.

Billiet, J., Cambré, B., & Welkenhuysen-Gybels, J. (2002). Equivalence of Measurement Instruments for Attitude Variables in Comparative Surveys taking Method Effects into Account: the Case of Ethnocentrism. In A. Ferligoj & A. Mrvar (Eds.), *Developments in Social Science Methodology* (pp. 73–96). Ljubljana: FDV.

Billiet, J., Swyngedouw, M., & Waege, H. (2004). Attitude Strength and Response Stability of a Quasi-Balanced Political Alienation Scale in a Panel Study. In W.E. Saris & P. Sniderman (Eds.), *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error and Change* (pp. 268–292). Princeton: Princeton University.

Billiet, J., Koch, A., & Philippens, M. (2007a). Understanding and improving response rates. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 113–137). London: Sage.

Billiet, J., Philippens, M., Fitzgerald, R., & Stoop, I. (2007b). Estimation of nonresponse bias in the European social survey: Using information from reluctant respondents. *Journal of Official Statistics, 23*(2), 135–162.

Billiet, J., Matsuo, H., Beullens, K., & Vehovar, V. (2009). Non-response bias in cross-national surveys: Designs for detection and adjustment in the ESS. *ASK. Research and Methods, 18*, 3–43.

Blom, A., Jäckle, A., & Lynn, P. (2010). The use of contact data in understanding cross-national differences in unit-nonresponse. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, T. L. E. Lyberg, P. P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (pp. 335–354). Hoboken: Wiley

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bradburn, N. M. (1992). What have we learned? In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 315–324). New York: Springer.

Borhnstedt, G. W. (1983). Measurement. In P. H. Rossie, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research: Quantitative studies in social relations* (pp. 70–122). New York: Academic Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105.

Cantril, H. (Ed.). (1944). *Gauging public opinion*. Princeton: Princeton University Press.

Converse, J. M. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (pp. 206–261). New York: The Free Press.

Converse, J. M., & Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*, Sage University, Series on Quantitative Applications in the Social Sciences. Beverly Hills: Sage Publications.

Couper, M. P., & De Leeuw, E. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 157–178). New Jersey: Wiley.

Couper, M. P. (2001). Measurement error in web surveys. *Methodology & Statistics*.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Davidov, E., Schmidt, P., & Billiet, J. (Eds.). (2011). *Cross-cultural analysis: Methods and applications* (p. 503). New York: Francis & Taylor.

De Wit, H. (1994). *Cijfers en hun achterliggende realiteit. De MTMM-kwaliteitsparameters op hun kwaliteit onderzocht.* Leuven: K.U.Leuven, Departement Sociologie.

De Wit, H., & Billiet, J. (1995). The MTMM design: Back to the founding fathers. In W. E. Saris & A. Münnich (Eds.), *The multitrait-mutlimethod approach to evaluate measurement instruments* (pp. 39–60). Budapest: Eötvös University Press.

de Heer, W. (1999). International response trends: Results of an international survey. *Journal of Official Statistics, 15*(2), 129–142.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. New York: Wiley.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K, Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research, 38*(1), 1–18.

Foddy, W. (1993). *Constructing questions for interviews and questionnaires. Theory and practice in social research*. Cambridge: Cambridge University Press.

Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. P. Biemer, et al. (Eds.), *Measurement errors in surveys* (pp. 393–418). New York: Wiley.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*(2), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56*(3), 328–351.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M. (2006). Non response rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646–675.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New York: Wiley.

Groves, R., & Herringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A, 169*(3), 439–457.

Groves, R., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849–879.

Heise, D. R. (1969). Separating reliability and stability in test-retest-correlation. *American Sociological Review, 34*(1), 93–101.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review, 21*(3), 360–373.

Hippler, H.-J., Schwarz, N., & Sudman, S. (Eds.). (1987). *Social information processing and survey methodology*. New York: Springer.

Holleman, B. (2000). *The forbid/allow assymetrie. On the cognitive mechanisms underlying wording effects in surveys*. Amsterdam: Ed. Rodopi.

Jabine, T. B, Straf, M. L, Tanur, J. M., & Rourangeau, R. T. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (Eds.). (2007). *Measuring attitudes cross-nationally: Lessons from the European social survey*. London: Sage.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrica, 36*(2), 109–133.

Kallenberg, A.L., & Kluegel, J.R. (1975). Analysis of the multitrait multimethod matrix: some limitations and an alternative. *Journal of Applied Psychology, 60*, 1–9.

Kaminska, O., McCutcheon, A., & Billiet, J. (2011). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly, 74*(5), 956–984.

Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A Review. *Journal of the Royal Statistical Society, 145*(1), 42–73.

Kreuter, F., & Kohler, U., (2009) Analysing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustment in the European Social Survey. *Journal of Official Statistics, 25*, 203–226.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., et al. (2010). Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples from multiple surveys. *Journal of the Royal Statistical Society Series A, 173*(2), 389–407.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in survey research* (pp. 29–44). San Francisco: Josey-Bass.

Loosveldt, G., Carton, A., & Billiet, J. (2004). Assessment of survey data quality: A pragmatic approach focused on interviewer tasks. *International Journal of Market Research, 46*(1), 65–82.

Loosveldt, G., & Beullens, K. (2009). *Fieldwork monitoring. Work package 6, Deliverable 5*. (Version 2): http://www.risq-project.eu

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., & Trewin, D. (1997) (Eds.), *Survey measurement and process quality*. New York: Wiley.

Lynn, P. J., Beerten, R., Laiho, J., & Martin, J. (2001) *Recommended standard final outcome categories and standard definitions of response rate for social surveys*. Working Paper 2001, 23 October.

Lynn, P. (2003). PEDAKSI: Methodology for collecting data about survey non-respondents. *Quality & Quantity, 37*, 239–261.

Lynn, P. & Kaminska, O. (2011). *The impact of mobile phones on survey measurement error* (p. 45). Working Paper Series. Essex: Institute for Social & Economic Research, 2011–2007

Matsuo, H., Billiet, J., Loosveldt, G., & Malnar, B. (2010a). *Response-based quality assessment of ESS Round 4: Results for 30 countries based on contact*

files (pp. 80). CeSO: Working paper, CeSO/SM/2010-2.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010b). Measurement and adjustment of non-response bias based on non-response surveys: The case of Belgium and Norway in European Social Survey Round 3. *Survey Research Methods, 4*(3): 165–178.

McClendon, M. J. (1991). Acquiescence and recency response-order effect in interview surveys. *Sociological Methods and Research, 20*(1), 60–103.

McClendon, M.J. (1992). On the measurement and control of acquiescence in latent variable models. Paper presented at the annual meeting of the American Sociological Association. (pp. 24–29). Pittsburg: Augustus.

Mirowsky, J., & Ross, C. E. (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2×2 Index. *Social Psychology Quarterly, 54*(2), 127–145.

Molenaar, N. J. (1986). *Formuleringseffecten in survey-interviews: een nonexperimenteel onderzoek*. Amsterdam: Vrije Universiteit.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review, 20*(4), 303–320.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly, 70*(5), 737–758.

O'Muircheartaigh, C., Gaskell, G., & Wright, D. B. (1995). Weighing anchors: Verbal and numeric labels for response scales. *Journal of Official Statistics, 11*(3), 295–307.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Payne, S. L. (1951). *The art of asking questions*. Princeton: Princeton University Press.

Rorer, L. G. (1965). The great response style myth. *Psychological Bulletin, 63*(3), 129–156.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). *Handbook of survey research: Quantitative Studies in Social Relations.* New York: Academic Press.

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly, 5*(1), 91–92.

Saris, W.E. (1990a). Models or evaluation of measurement instruments. In W. E. Saris. & A. van Meurs. (Eds.), *Evaluation of measurement instruments by meta-analysis of multitrait-multimethod studies* (pp. 52–80). Amsterdam: North Holland.

Saris, W. E. (1990b). The choice of a model fore evaluation of measurement instruments. In W. E. Saris & A. van Meurs (Eds.), *Evaluation of measurement instruments by meta-analysis of multitrait-multimethod studies* (pp. 118–132). Amsterdam: North Holland.

Saris, W. (1995). Designs and models for quality assessment of survey measures. In W. Saris & A. Munnich (Eds.), *The multitrait-Multimethod Approach to Evaluate measuremnt Instruments*. Eötvös University press: Budapest.

Saris, W. E., & Andrews, F. M. (1991). Evaluation of measurement instruments using a structural modeling approach. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 575–598). New York: Wiley.

Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach to evaluating quality of measurement instruments: The split-ballot MTMM Design. *Sociological Methodology, 34*(1), 311–347.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken: Wiley.

Saris, W. E. & Sniderman, P. M. (Eds.) (2004). *Studies in public opinion: Attitudes, nonattitudes, measurement error and change*. Princeton: Princeton University.

Scherpenzeel, A. (1995). Misspecification effects. In W. E. Saris & A. Münnich (Eds.), *The multitrait-mutlimethod approach to evaluate measurement instruments* (pp. 61–70). Budapest: Eötvös University Press.

Scherpenzeel, A. C., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods & Research, 25*(3), 341–383.

Schmitt, N., & Stults, D. N. (1986). Methodology review. Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1–22.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, *35*(1), 101–113.

Schuman, H., & Presser, S. (1979). The open and closed question. *American Sociological Review, 44*, 692–712.

Schuman, H. (1982). Artifacts are in the mind of the beholder. *The American Sociologist, 17*, 21–28.

Schuman, H. (1984). Response effects with subjective survey questions. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. Vol. I, pp. 129–147). New York: Russell Sage Foundation.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York: Academic Press.

Schuman, H., & Ludwig, J. (1983). The norm of even handedness in surveys as in life. *American Sociological Review, 48*(1), 112–120.

Schwarz, N., & Hippler, H. J. (1987). What response scales may tell your respondents. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 163–178). New York: Springer.

Schwarz, N., Strack, F., Mueller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of questions: Further evidence

on informative functions of response alternatives. *Social Cognition, 6*(2), 107–117.

Schwarz, N., & Hippler, H. J. (1991). Response alternatives: The impact of their choice and presentation order. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 163–178). New York: Wiley.

Schwarz, N., & Sudman, S. (Eds.) (1992). *Context effects in social and psychological research*. New York: Springer.

Schwarz, N., Knäper, B., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. de Leeuw, J.J. Hox & D.A. Dillman (Eds.), *International Handbook of Survey Methodology*. (pp. 18-34). New York: Lawrence Erlbaum Associates.

Smith, T. W. (1984). Nonattitudes: A review and evaluation. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 215–255). New York: Russell Sage Foundation.

Sniderman, P. M., & Grob, D. B. (1996). Innovations in experimental design in attitude surveys. *Annual Review of Sociology, 22*, 377–399.

Sniderman, P. M., & Hagendoorn, L. (2009). *When ways of life collide: Multiculturalism and its discontents in The Netherlands*. Princeton: Princeton University Press.

Strack, F., & Martin, L. L. (1987). Thinking, judging, and communicating: A processs account of context effects in attitude surveys. In H. Hippler, N. Schwarz & S. Sudman (Eds.) *Social Information Processing and Survey Methodology* (pp. 123–148). New York: Springer.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition, 9*(1), 111–125.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions. A practical guide to questionnaire design*. San Francisco: Jossey Bass.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass Publishers.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). Improving survey response: Lessons learned from the European Social Survey. Chichester: Wiley.

Tourangeau, R. (1992). Context effects on responses to attitude questions: attitudes as memory structures. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 35–48). New York: Springer.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299–314.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Vehovar, V. (2007). Non-response bias in the European Social Survey. In G. Loosveldt, M. Swyngedouw & B. Cambré (Eds.), *Measuring meaningful data in social research* (pp. 335–356). Leuven: Acco publisher.

Weijters, B. (2006). *Response styles in consumer research*. Belgium: Vlerick Leuven Gent Management School.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge: Cambridge University Press.

Why People Agree to Participate in Surveys

11

Gerald Albaum and Scott M. Smith

11.1 Introduction

Why do some people participate in survey research as respondents while others fail to participate? There is very little known about the reasons for or against participation and among some respondents why the quality of data varies. A great number of empirical studies have been conducted on response inducement techniques and other methodological artifacts that affect response to surveys, particularly mail surveys. Of concern in these studies have been the effects of devices such as preliminary notification, foot-in-the-door technique, follow-ups, questionnaire format and length, survey source or sponsorship, personalization, cover letters (or e-mails), deadline date, and premiums and rewards. Nature of return envelopes and typed postage are relevant only for mail surveys. Reviews by Peterson and Kerin (1981) and by Yu and Cooper (1983) included all major methods of survey data being used at the time (mail, telephone, personal interview), as did the quantitative review of response effects to interviews by Sudman and Bradburn (1974). The consensus of all this research is that there is no strong empirical evidence that any technique or aspect of a technique is always best at inducing response, except, perhaps, for pre-notification, follow-ups, and the use of a monetary incentive, (particularly when the incentive that is sent with the request for participation). Typically survey response studies use one or more of the following as a dependent variable or measure of effect: response rate, item omissions, response speed, cost, and in a much more limited manner, response content itself.

Since the late 1990s, online survey data collection has been growing in importance and popularity. Summaries of the research on methodological characteristics of this mode of survey data collection are presented by Miller (2006), Smith et al. (2006), and Roster et al. (2007). Internet-based survey data collection is fast becoming the preferred mode for conducting survey-based research in all areas of business and society.

The lack of a generally accepted framework or theory of survey response has been a hindrance to the development of effective survey techniques. A theory of survey response could explain behavior such that a researcher could predict which techniques would do what under differing conditions. A theory of survey response will guide researchers and help determine how to encourage response and accurate and truthful reporting of data.

This chapter addresses these issues by presenting a framework for explaining survey response that is applicable to all modes of data

G. Albaum (✉)
Robert O. Anderson Schools of Management, University of New Mexico, Albuquerque, NM, USA
e-mail: albaum@mgt.unm.edu

S. M. Smith
Brigham Young University, Provo, UT, USA
Qualtrics, PO Box 970340, Orem, UT 84097-0340, USA

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_11, © Springer Science+Business Media New York 2012

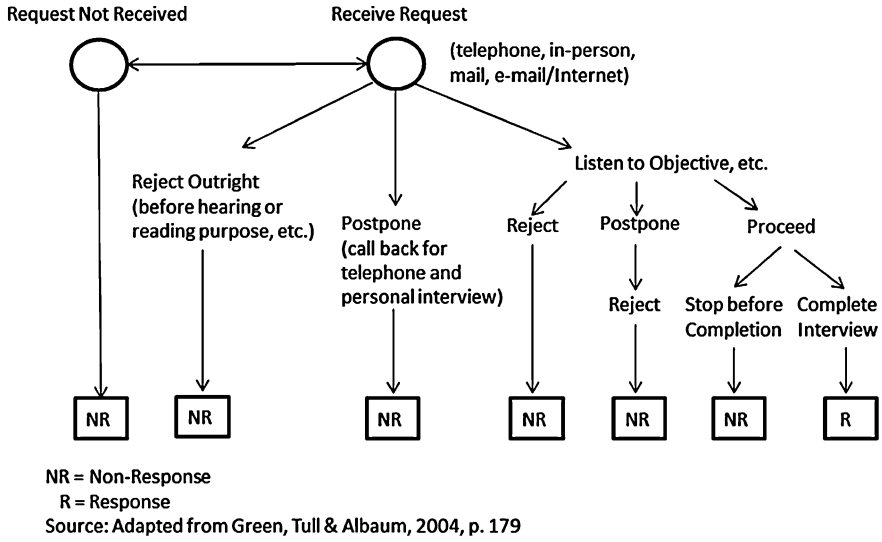


Fig. 11.1 A decision-stage model of response to surveys

collection. We then discuss potential underlying theories of survey response behavior. Researchers can use such theories to guide them in developing appeals and techniques that encourage and facilitate completion and submission. Following this, specific data collection techniques that can leave an effect on response (quantity and quality) are discussed briefly. Finally, we discuss how the interest and sensitivity of the topic can affect response rate and data quality.

11.2 A Framework for Understanding Survey Response

11.2.1 A Decision-Stage Model

Decision models are one form of framework that describes prospective survey participants' decision process from initial contact through completion of a survey. A decision-stage model describes a process that carries prospective respondents from initial contact to submission, or failure to submit, a survey, regardless of the modes of data collection used. Green et al. (2004) proposed one such model to explain response to telephone surveys as shown in Fig. 11.1. The model begins with the respondent being exposed (or not) to a solicitation to participate. The last stage refers to whether the respondent completes the questionnaire, and returns it to the researcher. At intermediary stages, the model proposes that respondents may take one of three actions: reject the survey, postpone response, or go on to the next stage. This simple model can be used to illustrate the process of response (or non-response) to all types of surveys.

Internet surveys are fast becoming a favored mode of survey data collection, and provide some opportunities that are not available with telephone, personal interviewing, and mail surveying. For example, the response rate for web surveys may be decomposed into click rate, submission rate (Han et al. 2009) and completion rate. Click rate is the proportion of those who click on the survey link to those who receive an invitation to participate in a survey. Completion rate, also known as submission rate, is the proportion of those who click on the survey link and who submit a questionnaire. Completion rate is the proportion of those who click on the survey link and complete and submit each of the sections of the

questionnaire. The potential differential impacts of techniques for increasing response rates on these two components expand opportunities for web-based methodological research on response rates. The proposed decision-stage model provides a conceptual framework for decomposing response rate into separate click and complete/submission components. The decision-stage model suggests that survey techniques may selectively influence response.

Three types of factors influence prospective respondents' rate of progressing through the stages and the likelihood of ultimately completing and returning the survey:

1. External factors, such as access to the Internet, whether the recipients check the e-mail account, speed of the connection, and spam filtering software. These are exogenous variables and while influenced by the survey distribution software, or the "cleanliness" of a panel list, are not influenced by a specific study.
2. Survey-related factors, such as whether the invitation to participate in the survey is compelling, and the effectiveness of techniques used to increase survey response.
3. Internal factors, such as whether the respondent has an interest in the topic, their motivational status, and the recipients' comfort level.

The five stages of the proposed decision model are defined as follows:

Stage 1 Respondents are (or are not) exposed to the solicitation to participate in the study.

Stage 2 Respondents choose to accept, postpone, or reject the solicitation to participate in the survey. If the decision is to postpone, the person may subsequently reject, or decide to continue.

Stage 3 The respondent accesses the survey (clicks on the link for Internet surveys) or listens to the introduction (for telephone and personal interview) for the survey. The opening will typically include information about the study and an appeal to complete the survey. The prospective respondent may reject the survey, postpone, or open the questionnaire and proceed. An unpersuasive appeal will reduce the number of prospective respondents who complete the questionnaire.

Stage 4 Prospective respondents begin responding to the questionnaire. They may then complete the questionnaire, or stop before completion. Item omission rates are increased and questionnaire completion rates are reduced by including questions that people do not understand, or do not want to answer. Some questions may be viewed as being sensitive in which they are intrusive, threatening or give risk to potential respondents (Tourangeau and Yan 2007; Albaum et al. 2011). In addition, using forced answering, which is an increasing phenomenon particularly for Internet-based research, will also reduce completion rates (Dillman 2000, p. 394; Albaum et al. 2010). Depending upon where in the questionnaire termination occurs, there still may be a partial completion submitted. Whether such a response can be included in analysis depends upon what is deemed an acceptable item omission rate, and which specific questions have not been answered. If key questions have not been answered then the submission cannot be really used in analysis.

Stage 5 The survey has been completed and the remaining task is for the respondent to return it to the researcher. Any procedure or requirement that complicates returning the questionnaire will operate at this stage.

Describing the survey response process as a series of stages and participation decisions opens for consideration *techniques* for increasing survey research quality that may be described in terms of their differential impacts on participation/postponement decisions at the respective stages. An example of such a technique is using some type of incentive. We discuss such technique later in this chapter.

11.2.2 Modes of Data Collection

Survey information is obtained from respondents through communication in several alternative media modes. Respondents may be interviewed in person, by telephone, or they may be sent a questionnaire. Mail, fax, and the Internet surveys are similar in which non-personal self-reporting is involved.

11.2.2.1 Personal Interview

As the name implies, the personal interview consists of an interviewer asking questions to one or more respondents in a face-to-face situation. The interviewer's role is to get in touch with the respondent(s), ask the desired questions, and record the answers obtained.

While it is significantly more expensive on a per-completed-interview basis, the personal interview, as a collection medium, has several advantages relative to other modes of data collection. It provides the opportunity to obtain a better sample, since virtually all the sample units can be reached and, with proper controls and well-trained interviewers, non-response to the survey or to individual questions can be held to a minimum. It also gives the opportunity to obtain more information, as a personal interview can be of substantially greater length than other modes. Finally, it permits greater flexibility. More freedom is provided for adapting and interpreting questions as the situation requires, especially in the case of unstructured personal interviews where concepts or treatments are introduced using visual, auditory, or olfactory aids.

The limitations of the personal interview include time, cost, and the response bias that may be induced by poorly trained or improperly selected interviewers. Problems with personal interviews arise from its very nature in that it is a social interaction between strangers, often on the respondent's territory, initiated by an interviewer who may have little in common with the respondent.

In addition to the home and workplace, many studies conduct consumer interviews in malls, where the so-called mall-intercept method is used. This method avoids the logistic, financial, and time costs of travel to meet with respondents. The mall-intercept method involves having interviewers stationed at selected places in a mall who request interviews from people passing by. Presumably the people are chosen on the basis of a predetermined sampling plan. A similar approach is to "interview on the street".

A variation of the mall intercept that is often used in business and other types of organizations is to interview at conferences, sales meetings, or other gatherings representing the population of interest.

11.2.2.2 Telephone Interview

Telephone interviews are often used in lieu of personal interviews, especially when personal contact is desired, when the information must be collected quickly and inexpensively, and when the amount of information required is relatively limited. Compared to e-mail or mail surveys, telephone interviews are often more costly in terms of total costs of data collection. However, when cost is figured on a per-completed-questionnaire basis, telephone interviews may be less costly than mail, but more costly than e-mail. In addition, telephone surveys offer the opportunity to probe for clarification or further information.

It is generally recognized that for business-to-business and consumer research, telephone interviewing is as effective as personal interviewing for scope and depth of information obtained. In addition, when a telephone survey is conducted from a call center, they can be better supervised than personal interviews.

Virtually all telephone interviews are structured direct interviews. The likelihood of the potential respondent refusing to be interviewed is always present when starting a telephone interview. Telephone surveys are unique in which they allow the interviewer to respond to the potential respondent and attempt to turn a refusal into a completed interview.

The telephone survey may be a good approach to reach specific population segments, particularly when door-to-door interviews are not possible or might lead to serious distortions in response. It is obvious that there must be sufficiently high telephone penetration in the segment for this mode of data collection to be advantageous.

The basic limitations of telephone interviews are the relatively limited amounts of information that can be obtained (at least compared with alternative methods) and the bias that exists in any sample of home telephone subscribers. In the United States, for example, more than 25%

nationally and more than 50% in large cities are not listed in a published directory, either because they have an unlisted number or as a result of moving. A technique for including unlisted telephone numbers in the sample frame is called random digit dialing (RDD).

Another additional problem for telephone researchers is that home telephone subscribers are disappearing. Currently in the United States about one in five homes do not have a "landline" telephone, but rely instead on cell phones or computer-based phone services. This problem is even more pronounced for specific groups such as students or those in ages 18–30 who rely on cell phones rather than home telephones (see Vehovar et al. Chap.16).

Additional problems associated with telephone interviewing are those of sample control and interviewer performance. Often this is manifested by inadequate efforts to complete interviews with some of the harder-to-reach respondents. Adding another sample is no substitute for dealing properly with the original sample.

11.2.2.3 Mail Interview

Mail interviews have in the past been widely used. Mail questionnaires provide great versatility at relatively low cost and are particularly cost-effective when included as a part of scheduled mailing, such as a monthly correspondence or billing. A questionnaire may be prepared and mailed to people in any location at the same cost per person: the cost of preparing the questionnaire, addressing the letter or card sent, and the postage involved. Respondents remain anonymous unless a name is requested, the questionnaire is openly coded, or some ethically questionable practice is employed.

Timeliness of responses is critical in mail surveys. If the time given is reasonable, say one or two weeks, stating a deadline should not adversely affect the response rate. Stating such a deadline may encourage the potential respondent to not postpone the task indefinitely.

Perhaps the most serious problem with mail surveys is that of non-response. Typically, people indifferent to the topic being researched will not respond. It is usually necessary to send additional mailings (i.e., follow-ups) to increase response. But even with added mailings, response to mail questionnaires is generally a small percentage of those sent.

Since people responding to a mail questionnaire tend to do so because they have stronger feelings about the subject than the non-respondents, biased results are to be expected. To measure this bias, it is necessary to contact a sample of the non-respondents by other means, usually telephone interviews. This is a type of non-response validation. The non-response validation may indicate that population subgroups have not been omitted and that results may not be biased.

Many variations of the mail interview are frequently used. These include the warranty card, hand-delivered surveys, newspaper/magazine surveys, the fax back survey, and survey on the back of checks.

11.2.2.4 Internet and E-Mail Interviews

As computer coverage in home markets increase, the use of electronic surveys has increased. Internet surveys delivered by e-mail, pop-ups, or as part of website content, are fulfilling their promise to be a driving force in business and other types of research.

The Internet has experienced a growth rate that has exceeded any other modern technology, including the telephone, VCR, and even TV. The Internet has diffused from a highly educated, white-collar, upper income, male dominated core, to include most of the teen and adult population in the more economically developed countries of the world.

Electronics-driven lifestyles that include online social networks, massive use of texting, and pervasive Internet connections are no doubt responsible in part for, or seemingly responsible for, attitude and behavioral changes in the way we view our increasingly virtual world. Strong upward trends are observed in the percentage of Internet purchases for airline tickets, CDs, DVDs, books, computer software, hardware, and systems. These online customers provide excellent access for research purposes.

Table 11.1 Advantages and disadvantages of different survey methods

Survey Method	Advantages	Disadvantages
Personal surveys	Respondent is at ease and secure at home; face-to-face contact; can observe respondent's home or working place; interviewer can show, explain, probe, etc. Often much in-depth information per interview is gathered Multi-method data collection, including observation, visual clues, and self-administered sections, are feasible Rapport and confidence-building are possible (including any written reassurances that may be needed for reporting sensitive material)	Greater chance of interviewer bias; potential sampling problems Field staff required Cost per interview can be high; interviewer must travel to respondent's place The data collection period is likely to be longer than for most procedures
Intercept surveys	Interviewer can show, explain, probe as in door-to-door Fast and convenient data collection Same as personal surveys	There may be many distractions in a street environment; respondent may be in a hurry or not in a proper frame of mind; more chance for interviewer bias; non-probability sampling problems Intercept company often has exclusive interview rights for a shopping mall (especially in the US)
Telephone surveys	Lower costs per interview than personal surveys Interviewer starting and management easier than personal interviews; Random digit dialing sampling of general population Likely better response rate from a list sample than from mail	Sampling limitations, especially as a result of omitting those without telephone Non-response associated with random dialing sampling is higher than with interviews Possibly less appropriate for personal or sensitive questions if no prior contact Questionnaires or measurement constraints, including limits on response alternatives, use of visual aids, and interviewer observations
Postal surveys	Elimination of interviewer and associated biases; respondent can complete the questionnaire when convenient; respondent also can look up information and work at own pace Respondents have time to give thoughtful answers, look up records, or consult with others Ease of presenting questions requiring visual aids (in contrast to telephone interviews) Cost of interviewer eliminated	Especially careful questionnaire design needed Low response rates Open questions usually are not useful The interviewer is not present to exercise quality control with respect to answering all questions, meeting questions objectives, or the quality of answers provided
Online surveys	Ease of creating and posting Inexpensive to administer; respondent can be shown stimuli and possible rewards Data can be quickly gathered Low costs per respondent Very flexible and fast online statistical analysis available	Respondents must have access to the Internet Population may be skewed toward young, educated, above-average income respondents Unrestricted sample is a simple convenience sample Buying an e-mail list can be expensive and not up-to-date. People change e-mail address more often than they change phone numbers (Ray and Tabor 2003) Loss of anonymity

Source Schmidt and Hollensen (2006), p. 145

Advocates of online surveying quickly point to the elimination of mailing and interviewing costs, elimination of data transcription costs, and reduced turnaround time as the answer to client demand for lower cost, timelier, and more efficient surveys. Online research is growing and

Table 11.2 Comparative evaluation of alternative survey methods of data collection

Criteria	Telephone CATI	In-home interviews	Mall-intercept interviews	CAPI	Mail surveys	Mail panels	Internet/ e-mail
Flexibility of data collection	Moderate to high	High	High	Moderate to high	Low	Low	Moderate to high
Diversity of questions	Low	High	High	High	Moderate	Moderate	Moderate to high
Use of physical stimuli	Low	Moderate to high	High	High	Moderate	Moderate	Moderate
Sample Control	Moderate to high	Potentially high	Moderate	Moderate	Low	Moderate to high	Low to moderate
Control of data collection environment	Moderate	Moderate to high	High	High	Low	Low	Low
Control of field force	Moderate	Low	Moderate	Moderate	High	High	High
Quantity of data	Low	High	Moderate	Moderate	Moderate	High	Moderate
Response rate	Moderate	High	High	High	Low	Moderate	Very low
Respondent anonymity	Moderate	Low	Low	Low	High	High	High
Social desirability	Moderate	High	High	Moderate to high	Low	Low	Low
Obtaining sensitive information	High	Low	Low	Low to moderate	High	Moderate to high	High
Potential interviewer bias	Moderate	High	High	Low	None	None	None
Speed	High	Moderate	Moderate to high	Moderate to high	Low	Low to moderate	Very high
Cost	Moderate	High	Moderate to high	Moderate to high	Low	Low to moderate	Low

Source Malhotra (2004)

researchers are operating in a much faster paced environment than ever before. This pace will continue to increase due to increased use of wireless PDAs, Internet-capable mobile phones, Internet TVs, and other Internet-based appliances yet to be announced. Each is an acceptable venue for interacting with the population and conducting online research.

Substantial benefits accrue from the various approaches to computer-assisted data collection in surveys as shown below:

- Respondents need few computer-related skills.
- Respondents choose their own schedule for completing the survey.
- Complex branching within the survey can easily be incorporated.
- Text answers and the text of selected and not selected question items can be easily piped into questions throughout the survey. Respondent response times can accurately be measured to key questions or the survey.
- A variety of graphics can easily be displayed and directly related to questions.
- Eliminates need to encode data from paper surveys.
- Errors in data are less likely, compared to equivalent manual methods.
- Speedier data collection and encoding, when compared to equivalent manual method.

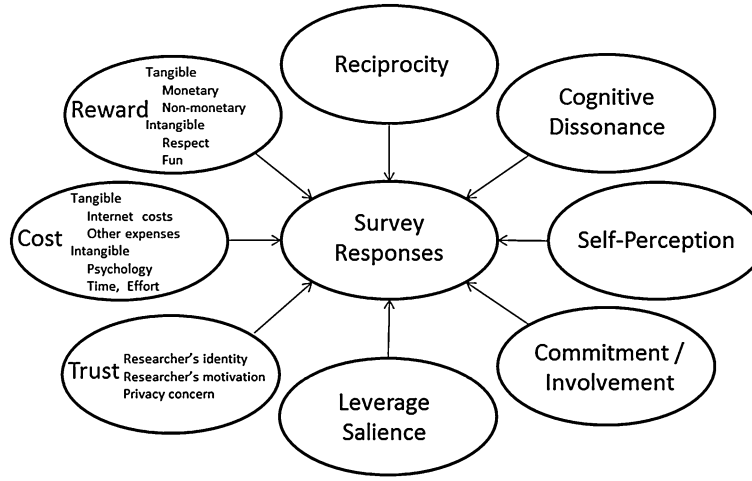


Fig. 11.2 Alternative theories of response source: adapted from Han et al. 2009, p. 430

11.2.2.5 Comparing Modes of Data Collection

The advantages and disadvantages of each of the modes of data collection are summarized by Schmidt and Hollensen (2006) and shown in Table 11.1. A more direct comparative evaluation of the alternative data collection modes is shown in Table 11.2.

11.3 Underlying Theories

A number of motivation theories have been suggested as explanations for the survey response decision (Linsky 1975; Dillman et al. 2009; Yammarino et al. 1991; Albaum et al. 1996, 1998; Poon et al. 2003; Evangelista et al. 2008). Figure 11.2 shows eight theories that have potential relevance for understanding responses (or lack thereof) to surveys. The four most cited theories for explaining survey response behavior are social exchange rewards, costs, trust (Dillman 2000; Dillman et al. 2009); cognitive dissonance (Furse and Stewart 1984; Hackler and Bourgette 1973); self-perception (Allen 1982; Tybout and Yalch 1980); and commitment/involvement (Albaum 1987). These theories are further explained in the various chapters of this handbook (e.g. Toepoel, Chap.13).

Another approach to theory is suggested by Groves et al. (1992). These authors suggest that critical elements in the development of a theory of survey participation are an understanding of the interaction between respondent and interviewer, and of the behaviors, attitudes and expectations each brings to the interaction. It is contended that understanding of survey participation can benefit from looking at the social psychological factors underlying compliance, helping tendencies, and opinion change. This is what a study by Groves and Couper (1998) has done. These researchers developed and empirically studied a basic theoretical framework of influences on survey participation. Specifically, Groves and Couper (1998) looked at attributes of design, interviewers' behavior, prior experiences of the respondent, interaction between respondent and interviewer, and the social environment in which the request for survey participation is made, as influencing the participation decision. Their research is limited primarily to face-to-face surveys and to studies of household population in the United States conducted or sponsored by US government agencies. One major concluding deduction is that "… many survey design features have their effect because they make favorable consequences of the interview more salient to the respondent prior to his/her decision; however, any stimulus presented during the survey request can interact with individual characteristics of the respondent" (Groves and Couper 1998, p. 296).

Yet another view of theory underlying survey response behavior is provided by Cavusgil and Elvey-Kirk (1998) who denote that six inherent motivators of human behavior can also be applied to explain survey response behavior. The motivators can be: net individual benefit, societal outcome, commitment, novelty, convenience, and expertise. For the most part, these motivators appear to fit within one of the four theories mentioned above.

We now discuss briefly each of the "major" theories underlying survey response behavior.

11.3.1 Social Exchange

The theory of social exchange, developed by Homans (1961), Blau (1964), and Thibaut and Kelly (1959), proposes that the actions of individuals are motivated by the anticipated relationship between the rewards and costs associated with these actions. As it was first applied to survey methodology (Dillman 1978), social exchange focused on appealing to potential respondents in multiple ways (Dillman et al. 2009, p. 22). People participate in a social exchange with others (e.g. responding to a survey) when the perceived rewards are greater than the expected costs. Applying social exchange theory to survey research, people are most likely to respond when they expect and trust that the rewards of responding to the questionnaire outweigh the costs associated with responding. Thus, there are three key factors in social exchange theory: rewards, costs, and trust.

Rewards are defined as what respondents receive from participating in a survey, and may be tangible (e.g., money) or intangible (e.g. emphasizing the respondent's importance and pointing out the usefulness of the reward). In contrast, cost is what respondents "pay" for participating in a survey, and may also be tangible (e.g., an Internet connection fee for a web-based survey, or time expended to participate) or intangible (mental effort required, sensitivity of questions, etc.). Finally, a social exchange relationship cannot take place without trust. In the end, any reward means nothing if potential participants do not believe the reward will be received.

Dillman (2000) and Dillman et al. (2009) assert that the guidance offered by exchange theory is only general. Many of the techniques used to induce compliance with the request for response, including the nature of the request itself, can be viewed as being designed to provide rewards, lower costs, and establish trust. For example, if completing a questionnaire or responding to an interviewer's questions can be made to be a rewarding act, the process itself may provide the motivation to participate in the mail or telephone survey. An implication for the researcher, therefore, is to make the questionnaire as interesting as possible (Dillman 1978). Another example is that of token financial incentives. Their effectiveness is probably not so much due to their monetary value as it is to their being a symbol of trust.

11.3.2 Cognitive Dissonance

Those using cognitive dissonance theory (Festinger 1957) propose that the motive for responding to a survey is avoidance, or reduction, of unpleasant feelings associated with nonresponse (Hackler and Bourgette 1973; Furse and Stewart 1984). Thus, it is an important component of the "respond/not respond" decision of potential survey respondents. Furse and Stewart (1984) postulate that cognitive dissonance theory provides a single model for integrating much of the empirical research that had been done up to that time.

For all types of surveys, the process is "triggered" by the request for participation, regardless of how this request is made–mail, personal, telephone, e-mail/Internet. Failure to respond to a survey may be inconsistent with a person's self-belief of being helpful, or at least a person who honors reasonable requests. The inconsistency created by not responding will produce a state of dissonance (i.e., state of anxiety and internal disharmony) which the potential respondent seeks to reduce by participating in the survey.

11.3.3 Self-Perception

Self-perception theory asserts that persons infer attitudes and knowledge of themselves through interpretations made about the causes of their

behavior (Bem 1972). Interpretations are made on the basis of self-observation. To the extent that a person's behavior is attributed to internal causes and is not perceived as due to circumstantial pressures, a positive attitude toward the behavior develops. These attitudes (i.e., self-perception) then affect subsequent behavior. Allen (1982) extended the self-perception paradigm to the broad issue of mail survey response. To enhance self-perception effects on response, self-observation labels should be created. Labeling involves classifying people on the basis of their behavior such that they will later act in a manner consistent with that characterization. Self-perception would predict that self-labeling of one's behavior would cause that person to view herself or himself as the kind of person who engages in such behavior, and therefore the likelihood of later label-consistent behavior is increased (Tybout and Yalch 1980).

The theory has also been applied in survey research to explain the effects, or lack of effects, associated with foot-in-the-door technique (DeJong 1979). This technique involves first gaining compliance to a small request such as answering a couple of questions in a telephone pre-call. Self-perception theory might be applicable to an Internet survey to explain internal—versus external—sources of motivations for responding to a survey. People respond to keep their behavior consistent with their favorable self-perception. Techniques that remind or promote the respondent's self-perception of being responsive and co-operative should increase survey response.

11.3.4 Commitment/Involvement

The concept of commitment is used to explain consistent behavior. It can be defined as "a variable, which encompasses the ranges of allegiance an individual may be said to have for the system of which he is a member" (Hornback 1971, p. 65). A person who is highly committed to an activity, such as responding to survey requests, is less likely to terminate the activity than one who is uncommitted (Ford 1973). Commitment may be attached to many different aspects of a survey, such as the source or the sponsor, the researcher, the research organization, the topic and issues in the survey, and even the process itself.

Consistent behavior is a central theme in this theory, including the following characteristics:

1. Persists over some period of time.
2. Leads to the pursuit of at least one common goal.
3. Rejects other acts of behavior.

Consequently, the major elements of commitment are viewed as including the following:

1. The individual is in a position in which his or her decision regarding a particular behavior has consequences for other interests and activities not necessarily related to it.
2. The person is in that position by merit of his or her own prior behavior.
3. The committed person must recognize the interest created by one's prior action, and realize it as being necessary.

The theory of commitment does not suggest that there will be consistent behavior regarding response to surveys in general, although this is not precluded. Rather, the theory suggests that there would be consistent response behavior for people facing the same type of survey under the same situation and conditions. Albaum et al. (1998) denote that research practitioners reported a relatively high level of awareness of the theory and its major aspects and a relatively high use of techniques derived from the theory as compared to other better-known theories. In brief, the theory depicts that people are more likely to respond to surveys if the topic, sponsor, or researcher is relevant to them. Their level of involvement will determine their degree of commitment to survey response.

11.3.5 Reciprocity

In simple terms, reciprocity refers to responding to a positive action with another positive action, and responding to a negative action with another negative one. Very broadly, reciprocity fits within the domain of duties and obligations because it does not necessarily involve leaving

special feelings of love and benevolence. It requires that a person give an in-kind response to another. This means that people will respond to each other in similar ways.

As a survey research theory explaining survey responses behavior, or as a basis for designing a survey in the first place, the theory of reciprocity fits within the domain of social exchange theory. Thus, much of what was discussed about social exchange theory applies also to reciprocity. The most obvious link is when an incentive, monetary or otherwise, is offered to the prospective respondent. A fair response to the incentive would be to participate in the survey.

Reciprocity is not limited to incentive-based surveys (Cavusgil and Elvey-Kirk 1998, p. 1170). The felt obligation to reciprocate and comply with the expectations of others by completing and returning the questionnaire may originate on the receipt of the request to participate since the recipient recognizes that time, effort, and resources were invested by the researcher. Submitting the completed questionnaire is the only way to reciprocate for the time, effort, and resources, and behave in a consistent manner with societal expectations. Thus, reciprocity also may be related to self-perception.

11.3.6 Leverage-Salience

This theory suggests that interest in a topic is a key factor in prospective respondents' willingness to participate in surveys. *Leverage* refers to the magnitude and direction (positive and negative) of the influence of various psychological predispositions toward participating in surveys in general and toward various design features in specific (e.g., incentive, topic, sponsor). Potential respondents vary in the degree to which leverage influences them. *Salience* refers to the prominence potential respondents assign to each component of the survey request protocol and design features (Groves et al. 2004). Thus, leverage-salience theory proposed that potential survey respondents vary in the importance they assign to different aspects of a survey request (Groves et al. 2004a). Research has shown that people responded at higher rates to surveys on topics of interest to them (Groves et al. 2004a; Schwartz et al. 2006).

11.3.7 Summary

This section has described six theories to explain survey response behaviors. The most widely studied have been social exchange, cognitive dissonance, self-perception, and commitment/involvement. Survey response researchers know the least regarding the effects of leverage-salience and reciprocity, although the latter has been studied within the context of social exchange.

Albaum and his associates have conducted a series of studies across a number of countries exploring the relevance of motivational theories to survey response (Albaum 1987; Albaum et al. 1996, 1998; Evangelista et al. 1999, 2008; Poon et al. 2003). These studies have been of research practitioners and non-practitioners. In general, the most prominent theory that has been, and is being, used is social exchange. In general among practitioners the least used is cognitive dissonance, although in some countries commitment or self-perception is least. Such mixed results may be a reflection of cultural influence. In many ways, we could not expect respondents in Hong Kong and Western Europe to behave in the same way or think similarly. What this means to researchers is that the use of theory of response behavior in solicitation for participation and survey design must be viewed as a cultural emic (culture-bound) and not a cultural etic (culture-free). Most of this research has been conducted within the context of the non-Internet-based modes of data collection. However, there is nothing to suggest that studying Internet-based survey responses would yield results any different from those shown above.

We deliberately have not attempted to assess which one is best to use when designing a survey. The best we can say is "it all depends!" Moreover, a study can be designed using more than one theory. For example, the request for participation can be based on self-perception theory while the techniques used can be based on social exchange.

11.4 Data Collection Techniques

As previously mentioned, theory of response behavior is inherent as an underlying influence in all techniques used to obtain a response from potential survey respondents. The overall process of data collection from an "ideal" mail survey is summarized as follows (Smith and Albaum 2005, p. 206):

1. Preliminary notification—letter or postcard, telephone call, e-mail.
2. Initial request for answering questions—mailing of questionnaire.
3. Reminder—letter or postcard, telephone call, e-mail.
4. Follow-up contact(s)/requests—questionnaire sent.
5. Non-response validation—telephone call, personal interview, e-mail.

With minor modifications this general sequence of contact activities is applicable to personal interview, telephone, and e-mail/Internet surveys.

A survey consists of many specific techniques that can affect response rate and data quality. Included are the following:

- Preliminary notification.
- Time required of respondent.
- Use of inducements.
- Open coding.
- Identification of sponsor and source.
- Follow-up policy.
- Questionnaire (measurement instrument) design, including visual designs.
- When to contact respondents.
- Type of appeal to use.
- Potential respondents' interest and commitment to the topic and/or study.
- Others unique to a specific technique of data collection (e.g., type of postage for a mail survey).

These techniques are discussed in detail in other chapters of this book and in other sources (Smith and Albaum 2005; Schmidt and Hollensen 2006; Dillman et al. 2009).

To illustrate the application of a theory of response behavior, we consider the "type of appeal to use."

Exchange theory
Incentive of a $2.00 scratch lottery ticket (Australia). McDonald's coupon worth HK $17.90

Self-perception:
responsible and concerned householders like yourself would like their opinions to be heardBeing a helpful person…

Commitment/involvement:
topic of survey is very important

Cognitive dissonance:
non-participation will cause us great difficulties

Similarly, the "use of inducements" is based on exchange theory, and "follow-up policy" could utilize any of the types of appeals illustrated above.

11.5 Topic

11.5.1 Interest

As mentioned above within the context of the leverage-salience theory of survey participation, the extent of interest in a topic will affect survey participation decisions (Groves et al. 2004). The greater the interest, the more likely one is to participate. Thus, survey response rates will be higher, the more the interest.

Interest in a topic, therefore, potentially affects non-response error. People more interested in the topic tend to have attributes or the key survey variables different from those of people less interested in the topic (Hanson and Smith 2011). Therefore, statistics computed on major variables of the topic are likely to be most susceptible to non-response error, especially when the topic is made salient in the appeal for participation (Groves et al. 2004, p. 2).

11.5.2 Sensitivity

There does not appear to be precise agreement on the definition of a sensitive topic. For example Lee and Renzetti (1993, p. 5) propose the following definition:

> …a sensitive topic is one that potentially poses for those involved a substantial threat, the emergence of which renders problematic for the researcher and/or the researched the collection, holding, and/or dissemination of research data.

Yet another view is that questions can be considered sensitive if respondents perceive them as intrusive, if the questions raise fears about the potential repercussion of disclosing information, or if they trigger social desirability concerns (Tourangeau and Yan 2007). A recent study by Kreuter et al. (2008) looks at the effects of mode of data collection—CATI, IVR, and the web—on response of sensitive information and reporting accuracy relative to the other modes of data collection.

Albaum et al. (2011) define sensitivity of topic as a topic that possesses a substantial threat to those involved as it may be perceived as intrusive and could raise fears about potential repercussions/consequences of disclosing the information request. There may be potential costs (or threats) to the respondent. Viewed this way, topic sensitivity is an individual matter, although there may be consequences within a defined population on the extent of sensitivity of any topic. Moreover, the level of perceived sensitivity of a topic to respondents may vary across cultures. With the increased reporting of cross-national/cultural research in marketing, topic sensitivity takes on greater importance. In a recent study, Albaum et al. (2011) report that topic sensitivity for 11 topics varied significantly between six countries (Australia, China PRC, Hong Kong SAR, The Netherlands, New Zealand, United States). Using a 5-point scale, where 1 = "not sensitive at all" and 5 = "extremely sensitive" the range of sensitivity scores for representative topics are:

- Family income/personal finances: 2.67 (United States) to 4.08 (Hong Kong SAR).
- Personal computer security behaviors: 2.17 (The Netherlands) to 3.88 (China PRC).
- Assessment of your employer/supervisor: 2.18 (United States) to 3.83 (Hong Kong SAR).
- Purchasing goods and services over the Internet: 1.56 (The Netherlands) to 3.38 (China PRC).
- Attitudes toward green marketing: 1.31 (New Zealand) to 3.00 (China PRC)
- Ethical attitudes about behaviors like cheating: 1.82 (United States) to 4.13 (China PRC).

Thus, sensitivity of topic appears to be a cultural emic (culture-bound) rather than cultural etic (culture-free).

One concern about asking questions on a sensitive topic is offered by Tourangeau et al. (2010). These researchers propose that when a survey includes potentially embarrassing questions (i.e., sensitive to the respondent), those in the socially undesirable category—i.e., those who are at risk of being embarrassed by the questions—are both more likely to refuse to be a respondent (assuming they know the topic, in advance) and more likely to misreport if they do participate.

11.6 Conclusion

This chapter has addressed a major question regarding survey research—why do people agree to participate. We first presented a framework for explaining survey response. This framework has various stages, and is applicable to all modes of data collection—mail, personal interview, telephone, and Internet. Next we discussed six theories that have been discussed—to a greater or lesser extent—in the literature to explain survey response, or lack thereof. Such theories have been applied to the appeals to participate and to overall research design and technique that would encourage and facilitate respondent completion and submissions. Specific data collection techniques were discussed briefly as these are discussed elsewhere in this book. Finally, the matter of effect of topic was presented. Both interest in topic and sensitivity of topic have potential impacts on response rate and data quality.

References

Albaum, G. (1987). Do source and anonymity affect mail survey results? *Journal of the Academy of Marketing Science, 15*(3), 74–81.

Albaum, G., Evangelista, F., & Medina, N. (1996). Australasian research practitioners' views of the role of response behavior theory in survey research strategy. *Australasian Journal of Market Research, 4*(2), 1–16.

Albaum, G., Evangelista, F., & Medina, N. (1998). Role of response behavior theory in survey research, a cross-national study. *Journal of Business Research, 42*, 115–125.

Albaum, G., Roster C. A., & Smith, S. M. (2011, May 24–27). *Topic sensitivity; implications for Web-based surveys*. Paper presented at the annual conference of the Academy of Marketing Science, Coral Gables.

Albaum, G., Roster, C. A., Wiley, J., Rossiter, J., & Smith, S.M. (2010). Designing Web surveys in marketing research: Does use of forced answering affect completion rates? *Journal of Marketing Theory and Practice, 18*(3), 305–313.

Allen, C. T. (1982). *Perspectives on mail survey response rates: the self-perception paradigm and beyond*. Paper presented at the American Marketing Association Conference on Marketing Theory.

Bem, D. J. (1972). Self-Perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press.

Blau, P. M. (1964). *Exchange power in social life*. New York: Wiley.

Cavusgil, S. T., & Elvey-Kirk, L. A. (1998). Mail survey response behavior: A conceptualization of motivating factors and an empirical study. *European Journal of Marketing, 32*(11/12), 1165–1192.

De Gong, W. (1979) An examination of self-perception mediation of the foot-in-the-door effect. *Journal of Personality and Social Psychology, 37*, 2221–2239.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dillman, D. A., Smyth, J. D., & Christian, L. H. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York: Wiley.

Evangelista, F., Albaum, G., & Poon, P. (1999). An empirical test of alternative theories of survey response behavior. *Journal of the Market Research Society, 41*(2), 227–244.

Evangelista, F., Poon, P., & Albaum, G. (2008). *Enhancing survey response rates: Lessons from a field experiment*. *Proceedings, Australia and New Zealand Marketing Academy Conference*. Sydney.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press, Stanford

Ford, D. (1973). Commitment: A mathematical model. *Quality & Quantity, 7*, 1–41.

Furse, D. H., & Stewart, D. (1984). Manipulating dissonance to improve mail survey response. *Psychology & Marketing, 1*(Summer), 79–94.

Green, P. E., Tull, D. S., & Albaum, G. (2004). *Research for marketing decisions* (5th ed.). New Delhi: Prentice Hall of India Private Ltd.

Groves, R. H., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475–495.

Groves, R. M., & Couper, M. P. (1998). *Non-Response in household interview surveys*. New York: Wiley.

Groves, R. M., Presser, S., & Dipko, S. (2004a). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2–31.

Groves, R. M., Singer, E., & Corning, A. (2004b). Leverage-salience theory of survey participation. *Public Opinion Quarterly, 64*, 299–308.

Hackler, J., & Bourgette, P. (1973). Dollars, dissonance, and survey returns. *Public Opinion Quarterly, 37*(Summer), 276–281.

Han, V., Albaum, G., Wiley, J. B., & Thirkell, P. (2009). Applying theory to structure respondents' stated motivation for participating in web surveys. *Qualitative Market Research, 12*(4), 428–442.

Hanson, J. M. & Smith, S. M. (2011). The impact of highly-interesting questions on completion rates and data quality in online marketing research. *International Journal of Marketing Research* (forthcoming).

Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, brace, and world.

Hornback, K. (1971). Toward a theory of involvement propensity for collective behaviour. *Sociological Forces, 4*(Summer), 61–77.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web-surveys: The effects of mode and question sensitivity. *Public Opinions Quarterly, 72*(5), 847–865.

Lee, R. M., & Renzetti, C. M. (1993). The problems of researcher sensitive topics: An overview. In C. M. Rengetti & R. M. Lee (Eds.), *Researching sensitive topics* (pp. 3–13). Thousand Oaks: Sage Publications.

Linsky, A. S. (1975). Stimulating responses of mailed questionnaires: A review. *Public Opinion Quarterly, 39*, 82–101.

Malhotra, N. (2004). *Marketing research: An applied orientation* (4th ed.). Upper Saddle River: Pearson Education, Inc.

Miller, J. (2006). Online marketing research. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research* (pp. 110–131). Thousand Oaks: Sage Publications.

Peterson, R. A., & Kerin, R. (1981). The quality of self-report data: Review and synthesis. In B. Emis & K. Roering (Eds.), *Annual review of marketing* (pp. 5–20). Chicago: American Marketing Association.

Poon, P. S., Albaum, G., & Evangelista, F. U. (2003). Why people respond to surveys: A theory-based study of Hong Kong respondents. *Journal of International Consumer Marketing, 16*(2), 75–92.

Ray, N. M., & Tabor, S. W. (2003). Cyber-surveys come of age. Marketing Research, Spring, pp. 32–37.

Roster, C. A., Rogers, R. D., Hozier, G. C., Jr, Baker, K. G., & Albaum, G. (2007). Management of marketing research projects: Does delivery method matter anymore in survey research? *Journal of Marketing Theory and Practice, 15*(2), 127–145.

Schmidt, M. J., & Hollensen, S. (2006). *Marketing research: An international approach*. Harlow: Pearson Education Ltd.

Schwartz, L. K., Goble, L., & English, E. H. (2006, May). *Counterbalancing topic interest with cell quotas and incentives: Examining leverage-salience theory in the context of the poetry in America survey*. Paper presented at the meetings of the American Association of Public Opinion Research.

Smith, S. M., & Albaum, G. S. (2005). *Fundamentals of marketing research*. Thousand Oaks: Sage Publications.

Smith, S. M., Smith, J., & Allred, C. R. (2006). Advanced techniques and technologies in online research. In R. Grover & M. Virens (Eds.), *The handbook of marketing research* (pp. 132–158). Thousand Oaks: Sage Publications.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago: Aldine Publishing Company.

Thibaut, J. W., & Kelly, H. H. (1959). *The social psychology of groups*. New York: Wiley.

Tourangeau, R., Groves, R. M., & Redlime, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly, 74*(3), 413–432.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 858–883.

Tybout, A., & Yalch, R. (1980). The effect of experience: A matter of salience? *Journal of Consumer Research, 6*(March), 406–413.

Yammarino, F. J., Skimmer, S. J., & Childers, T. L. (1991). Understanding mail survey response behaviour: A meta-analysis. *Public Opinion Quarterly, 55*, 613–639.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20*(February), 36–44.

Respondents Cooperation: Demographic Profile of Survey Respondents and Its Implication

12

Patrick Glaser

Research data collection that involves human subjects is dependent upon either voluntary or involuntary participation of persons. In the opinion research field, when utilizing conventional survey and opinion research methods, the degree and completeness to which individuals willingly engage in the data collection process is known as "respondent cooperation." Respondent cooperation has been a preeminent concern of survey and opinion researchers since at least the latter part of the twentieth century both because of its implications for data quality, as well as its reflection upon research methods and the resulting ethical and regulatory considerations.

In this chapter, respondent cooperation is defined and explained in detail. Specific attention is given to the importance of respondent cooperation in research, theories, and evidence of *how* and *why* respondent cooperation has changed over time, and the general profile of survey respondents. Also, this chapter concludes with a discussion of respondent cooperation in the future along with a list of suggested readings that are provided for a more detailed discussion of the topic. Particular attention is given to respondent cooperation in surveys due to its importance to the survey method.

Respondent Cooperation Defined The precise definition of "respondent cooperation" varies depending on the context within which it is presented. In its broadest sense, respondent cooperation refers to the *general willingness of the public at-large to participate in research*. From this perspective, the term relates to societal trends of participation in research and whether or not "the public" is prone to survey participation. To date, there have been numerous, widespread efforts aimed at improving the general level of respondent cooperation. Perhaps the most far-reaching of these are the ethical responsibilities and principles prescribed and enforced by many professional opinion research trade associations within the United States and abroad that specifically safeguard respondent interests in research (Marketing Research Association 2007a, b; ESOMAR 2011).

In a narrower sense of the word, the term "respondent cooperation" is used to describe the *level and comprehensiveness of participation achieved on a particular survey data collection effort*. With respect to cooperation in particular research efforts, this chapter focuses mainly on the individual accepting the research invitation and participating in the research study. However, cooperation issues may also come into play within a particular study, where a respondent may participate in the research project as a whole, but forgo participation with respect to particular questions, portions, or activities within the research.

In surveys, individuals who refuse to participate in a survey effort are said to create "unit non-response." Refusals to particular survey

P. Glaser (✉)
Washington, DC, USA
e-mail: patrick.glaser@mra-net.org

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_12, © Springer Science+Business Media New York 2012

questions are known as "item non-response." The concept of survey "non-response" refers to; and reflects, respondent cooperation with respect to individual survey research study efforts.

Conventions in Research Respondent cooperation levels are important in all forms of research that involve the active participation of individuals. However, cooperation is most critical in research efforts that seek to project findings from a statistically valid sample to population of interest. For opinion researchers, the focus of cooperation issues thus is most relevant to sample surveys.

Researchers consider anything less than complete (i.e., 100%) respondent cooperation potentially problematic. In other words, a researcher has not obtained full respondent cooperation unless every single individual selected and invited to a survey has participated. Importantly, probability and sampling theory assume the collection of data from all eligible sampled units. In practice, very nearly all surveys result in some amount of respondent cooperation deficiency. Moreover, the lack of respondent cooperation in surveys is typically substantial.

Respondent cooperation levels for surveys are linked to the conventions of, and typical methods employed by, the type of research being undertaken and what level of nonresponse is considered tolerable to that purpose and type of research. For example, a survey project designed for collecting data to be reported in an academic journal may employ more stout techniques to achieve high cooperation levels than an effort commissioned for a marketing research project. This is because basic, or pure, academic research requires a higher level of rigor and confidence in results than an applied marketing research effort.

Conventions such as these have been established over time as norms and are influenced by such considerations as the potential consequences of inaccurate survey results, the willingness of the survey sponsor to fund high-quality research efforts, and sensitivity to utilizing aggressive recruitment protocols towards the survey population. Notably, there is some amount of controversy as to the appropriateness of equating high cooperation levels with degree of quality. This is because nonresponders to surveys do not necessarily bias results. Moreover, it is also true that surveys with high cooperation levels may suffer from bias on some or all measures, while surveys with relatively low cooperation do not suffer from systematic "non-response bias" at all. These issues are discussed in more detail later in this chapter.

Generally speaking, researchers view nonresponse in surveys as a limitation and source of survey error. As such, many researchers have a sense of the level of nonresponse they are willing to tolerate on a given survey effort prior to data collection. Similarly, researchers are typically able to anticipate cooperation levels given their research design and population of interest. Prior professional experiences and a wealth of literature exist to inform individual survey practitioners of the likely response rates for their efforts.

Researchers use a variety of approaches to promote cooperation in surveys, including monetary and non-monetary incentives, limits to survey length and burden, well-planned survey invitations and participation requests, the use of highly trained interviewers capable of securing cooperation, employing numerous follow-up attempts with survey respondents in order to convince them to participate, and other techniques. In the most well-planned survey efforts, the particular techniques utilized to encourage cooperation in a survey effort are carefully selected according to theory. As discussed later in the chapter, frameworks exist for understanding the participation decision. Thus, researchers are able to draw upon these models of understanding when crafting an appeal to the survey respondent (see Stalans, Chap. 6).

Survey Nonresponse Metrics Respondent cooperation for individual surveys is typically measured by a variety of metrics that, for probability-based sample surveys, is sometimes collectively referred to as "outcome rates." Examples of widely-used and cited outcome rates are "response rates" which reflect the proportion of all eligible sampled respondents

who have completed the survey and the "refusal rate" which reflects the proportion of all eligible sampled respondents who refused to complete the survey. Other outcome rates include the "cooperation rate" and the "contact rate" (American Association for Public Opinion Research 2011).

Notably, surveys that rely on nonprobability selection methods (e.g., self-selected, or "opt-in," access panels) may report alternative nonresponse metrics (Association Collaborative Effort 2009). Researchers generally rely on several, rather than a single, respondent cooperation metric(s) in order to provide a greater breadth of information in understanding the performance of a data collection effort. Together, the evaluation of these metrics can lend clues to both the degree of and reasons for cooperation levels.

Investigating Survey Nonresponse There have been many individual published and nonpublished studies that have investigated techniques to improve respondent cooperation, as well as understand and mitigate the potential biases that may occur due to a lack of respondent cooperation. Notably, earlier efforts to understand respondent cooperation focused on measurement of the problem and the calculation of nonresponse metrics as well as understanding how to improve respondent cooperation for individual surveys. More recently, some attention has shifted to understanding the underlying drivers of biases that can result from respondent cooperation issues (Singer 2006).

Importantly, "Respondent Cooperation" has been a principle concern of researchers for many years. For example, scholarly journals in the social science and statistical fields have included articles on problems relating to respondent cooperation since the 1940s, calls for standardization regarding the reporting of respondent cooperation levels for individual surveys started as early as 1965, when Williams Edward Deming proposed reporting nonresponse outcomes in "Principles of Professional Statistical Practice," and the US Federal Government has published standards concerning the minimum level of respondent cooperation acceptable in surveys since at least 1979 (Smith 1999).

12.1 Importance of Respondent Cooperation

Cooperation is considered important in a research context because it is believed to relate to a fundamental source of error: nonresponse error (Dillman 2009). Surveys that are designed to project findings from a selected sample to a population of inference rely on statistical theory and methods. The application of this theory assumes that data is collected from each sampled unit, or each selected survey respondent. In practice, this is rarely, if ever achieved.

Survey researchers have recognized that error may ensue from "survey non-response." When some selected cases, or individuals, cannot or do not participate in a survey they have been selected for, the potential representativeness of that sample may be affected. To understand why this is so, sampling itself must be considered.

The process of sampling is intended to provide a group that is drawn from, and reflective of, a larger population to a known degree of sampling error (survey sampling techniques were discussed earlier in Chap. 5, by Hibberts, Burke, and Hudson). Without random selection, cases, or individuals, may be selected for a survey that has a particular bias. In turn, this could create data that is not suitable for inference to a population.

For example, a survey intended to represent an entire university student body should consider each student. As such, a biased sample might be drawn if only students from a particular department within the university were selected. In such a case, survey data resulting from the biased sample might reflect the qualities of students in that particular department as opposed to those throughout the university at-large. It might be reasonable to assume, for example, that a survey of students selected solely from the Fine Arts department might be more interested in art than the average university student.

In a similar fashion, respondent cooperation can also affect survey error. If cases, or individuals, whom are selected to participate in a survey cannot or do not participate, their

nonresponse might affect the accuracy and reliability of survey results. Researchers speak of this phenomenon as "non-response bias" or "non-ignorable, non-response." It occurs when those who are selected but do not participate are different from those who do participate in important and meaningful ways—ways that correspond to what the researcher is measuring.

Nonresponse Bias Survey nonresponse does not necessarily create "non-response bias." It is also worth noting that it is not necessarily the case that the more survey nonresponse, the greater the degree of nonresponse bias. In fact, it is plausible that a survey would suffer from substantial amounts of survey nonresponse yet suffer from no amount of systematic bias. Bias occurs if and when the particular variables a researcher is measuring are related to the causes of the survey participation.

For example, a survey aimed at understanding the mix and popularity of jobs and job growth might be impacted if nonresponders did not participate for reasons that are related to their employment. In this hypothetical example, one imaginary circumstance that might cause such a situation is a case where members of professions that work many long hours (e.g., medical, legal) might be less available for surveying and thus underrepresented compared to their actual numbers in society. In such a case, it is also conceivable that those specific measures that relate to jobs be biased, but that those survey questions unrelated to an individual's work remain unbiased.

Cooperation Issues for Other Research Thus far, this chapter has focused on respondent cooperation issues in surveys. This is because cooperation is vital for surveys, which are most often conducted for the purpose of making population inferences from a sample of cases or individuals. However, other, noninferential forms of research may be utilized that also rely on the participation of individuals. For example, focus group facilities and recruiters expend significant effort in getting individuals to participate in focus groups and other forms of qualitative research.

Noninferential forms of research may not be subject to the same degree of concern over cooperation as inferential surveys and other similar efforts. However, cooperation may still be considered as important for two reasons. First, the more effort required to recruit participants to research, the greater the burden. This often translates into lengthier requirements for recruiting time, the need for more highly skilled recruiters, greater incentives and other substantial, and robust recruitment efforts. Generally speaking, when cooperation levels are high and people are willing to participate in research, it is easier to conduct research regardless of whether the method is inferential or noninferential.

Second, the wide use of monetary research incentives has resulted in attracting willing research participants that actively seek-out research opportunities for the sake of cash benefits. Notably, some of these participants have been found to be disingenuous and dishonest when reporting their characteristics, behaviors, and other qualities. They typically do so in order to bypass screening criteria aimed at filtering out participants that may not fit the needs of the study. This problem has grown so great for focus groups that there have been several companies and organizations throughout North America that offer validation services specifically tailored to focus groups and their databases of respondents.

In this sense, researchers are broadly concerned with societal cooperation levels in order to ensure that there is a significant pool of participants and to avoid being limited to individuals who may be willing or likely to be dishonest to gain acceptance into the study. Interestingly, there is a tension between offering sufficient incentive to participate in research versus overmotivating individuals that may seek to recklessly gain entrance into research studies. This tension can be observed in the manner in which many focus group and survey companies attempt to attract respondents for database and panels by offering monetary rewards, but at the same time utilize validation and fraud prevention measures in order to weed-out undesirable individuals.

Cooperation as a Reflection of Research Many researchers have viewed respondent cooperation as a reflection of research methods. Whether looked at from an individual study or project or, from the general trend of society's

willingness to participate in research, cooperation may reflect how researchers are seen, and how popular research methods are understood. Importantly, these trends in opinion may have greater implications than simply the functional effect they have on bias, the prestige of research and researchers, or on the burden placed on participant recruiters. Public opinion of research may also lead to government, or even private corporation, limits on research methods.

Federal and state governments have passed legislation regulating information privacy and communications. For example, while the Federal government has exempted research from "Do Not Call" list provisions, they do subject research communications to limits placed on automatic dialer technologies. Due to Federal regulation, researchers may not contact cellular telephone numbers using an auto-dialer method unless they have prior consent.

Likewise, telecommunications, Internet service providers (ISPs), social media networks, and other mediums researchers use to connect with respondents have also established policies that may impact researchers. For instance, wireless carriers have established policies for appropriate methods of "text" communications to subscribers, and social media networks establish policies for data access and use that exists on their websites. For these reasons, it is important for researchers to communicate the value and importance of research to the public.

12.2 The Nature of Survey Participation

Numerous theories have been posited as to why individuals participate in research. These theories have borrowed heavily from other social science models. Accordingly, as the number of explanations of cooperation has grown, so has the sophistication of techniques and methods aimed at eliciting participation from respondents. Popular survey participation theories include general rational theories of participation, leverage saliency theory, and social exchange theory.

Rational Theories of Participation Rational theories of survey participation describe the respondent as one who calculates the advantages and disadvantages of survey participation. For example, the economic exchange theory of cooperation posits that the respondent weighs the benefits and costs of participating in research and then makes an at least partially rational decision as to whether it is in his or her interest to participate (Lavrakas 2008). Notably, monetary incentives are sometimes justified under the economic exchange perspective based on the idea of fairly and equitably compensating a respondent for the use of their time and effort. However, researchers may also view the use of monetary incentives simply as a device for establishing legitimacy, establishing benefit to participating or demonstrating courtesy, and appreciation toward the respondent.

Rational models of survey participation describe costs of survey participation as including such considerations as the personal amount of time and burden required in participating and the risk associated with providing potentially sensitive or embarrassing information to a stranger. The benefits of survey participation may include examples such as the opportunity to give one's opinion and potentially have influence and personal satisfaction felt from volunteering one's time (Braverman 1996). Though it is generally accepted that individuals do not always take a completely rational approach to decision making, rational models of survey participation may be useful in helping the researcher to consider the benefits and costs their survey plan creates for the respondent.

Leverage Saliency Theory of Participation According to the leverage saliency theory of participation, respondents may be attracted to participate, or decline research requests for a variety of different reasons. Each of these reasons may have different degrees of influence on each respondent. Respondents are most likely to accept an invitation to participate when those factors that are particularly salient to them are emphasized during the request (Groves 2000).

The leverage saliency theory of participation is useful in understanding why various individual approaches to motivating survey respondents (e.g., providing survey incentives alone) to participate have had mixed results. Each individual is unique, and has had unique experiences; therefore, they are motivated in unique ways (Groves 2000). Examples of relevant positive motivators include survey topic interest, survey incentive, familiarity and feelings towards survey sponsor, and feelings of personal reward associated with giving one's time to a worthwhile effort.

Likewise, a respondent may be motivated to decline a survey participation request due to factors such as burden, lack of trust, a sense that the effort is not worthwhile or beneficial, or other factors. A respondent is likely to make the decision to participate when the positive or attractive factors outweigh the negative factors. Notably, the leverage saliency theory of participation offers researchers a model for use when considering the different attributes contained in their survey requests (Groves 2000).

Social Exchange Theory of Participation The social exchange theory of participation argues that individuals make their participation decisions based on *expectations* of benefits and costs. Thus, an individual may decide to participate in a survey based on perceived benefits that may accrue to their self or others. Social exchange theory is important to the popular "tailored design method" of survey construction and implementation advocated by Dillman, where the perceived costs of participating in a survey effort are minimized by the researcher who simultaneously attempts to raise and highlight benefits of participation (Dillman 2009).

These and other theories of participation (e.g., cognitive dissonance, equity theory) provide a general framework for researchers in understanding respondent cooperation. Moreover, they provide perspective on why certain groups of individuals may be difficult to elicit cooperation.

12.3 Factors Influencing Participation

Groves and Couper (1998) proposed a framework for participation that included four principal areas: "social environmental influences (e.g., survey taking climate, urbanicity effects), knowledge and social psychological attributes of the sampled persons (e.g., civic duty, interest in politics), survey protocols (e.g., mode of data collection, incentives, burden of interview), and the selection and training of interviewers" (in interviewer-administered surveys) (Dillman 2002). Each of these aspects can have an influence on whether or not the respondent is successfully contacted, and if so, whether or not they will be persuaded to participate in the survey as requested.

Furthermore, Dillman (2002) give several additional influences on respondent cooperation; these include:

- the opportunity cost of participating in the survey (i.e., what the respondent perceives as the relative value of their time if spent doing another activity),
- the sense of reciprocity that the respondent feels towards the interviewer or entity making the survey request (i.e., social exchange, as described above),
- whether or not the topic of the survey is salient to the respondent and would thus interest them in participating,
- the skill of the interviewer in interviewer-administered survey modes in eliciting respondent cooperation,
- experiences with prior survey waves in panel, or longitudinal surveys, and, lastly,
- gatekeepers and organizational policies respecting survey participation in establishment surveys.

Finally, several design features of surveys are listed as additional factors that affect respondent cooperation. These involve two categories of processes that are listed according to their effect on respondent cooperation. The first include

field protocols that impact the likelihood of making contact with a respondent (the length of time the survey is in the field, the timing and call schedule procedures, and interviewer-workloads). The second grouping of factors include aspects of the survey design that may directly affect willingness to participate, such as knowledge of the survey sponsor, whether prenotifications or advanced notices have been employed, whether incentives are offered, protocols for follow-up with reluctant respondents, and other techniques (Dillman 2002).

Profile of Respondents Different respondent characteristics may relate to respondent cooperation. Age, education, race and ethnicity, gender, and place of residence may all have a relationship with an individual's likelihood to participate in a survey or research effort (Johnson 2008). For example, studies have found evidence that cultural differences, such as those approximated by race, ethnicity or national boundaries, are linked to survey nonresponse under certain conditions.

Johnson (2002) analyzed several studies and found little evidence among cross sectional surveys for differences in respondent cooperation among minority groups, for example, but did not find that minority groups are more subject to attrition in panel studies (Johnson 2002). A number of theories may help to explain why different groups may vary in their tendency to participate in opinion research. Among these are cultural explanations that posit differences in nonresponse tendencies among groups with individualistic versus collectivist orientations, groups that tend to be compliant versus accepting of power, and groups that tend to come from flat versus hierarchical systems of authority (Johnson 2002).

Cultures may also tend to differ across communication styles, where the importance of context, environmental cues, nonverbal behavior, norms for disclosure of information to others and under what circumstances disclosure is appropriate, fear or mistrust of other groups, willingness to help other groups or individuals within their own group, and other social norms that may be relevant to interaction with others. Importantly, in addition to cooperation, culture may play an additional role in survey participation in terms of its influence on preferred communication styles and the result this has on the researcher's opportunity to interview the respondent (Johnson 2002).

Additionally, other sociodemographic characteristics may relate to respondent cooperation. Roose (2002) summarize the conventional beliefs and empirical evidence, explaining:

- Gender: most studies find either no influence from gender on participation, or a small amount of impact leading to slightly higher participation by women.
- Age: older individuals tend to have higher noncooperation rates, but also tend to be easier to contact, due to their reduced mobility.
- Education: lower education levels tend to result in reduced levels of cooperation.

Ultimately, respondent cooperation for both surveys as well as other forms of research depends upon a mix of respondent characteristics and experiences, survey design qualities and implementation procedures, and related circumstances and environmental factors. Researchers have realized both the importance of respondent cooperation and the difficulty in obtaining high participation in their research. As a result, the majority of research involves deliberate planning in order to ensure that cooperation is at the forefront of design considerations.

12.4 Trends in Declining Respondent Cooperation

Researchers in marketing research, public opinion polling, government policy, academia, and other "segments" of (what is commonly referred to as) the "research profession" have been concerned over declining cooperation for many years. Respondent cooperation has been both a popular area and topic for academic inquiry, research and publications as well as for professional conferences, trade magazines, and other media. In fact, the issue has been considered so critical to researchers that an entire organization was formed to address the growing problem of respondent cooperation.

In 1992, an umbrella association named the Council for Marketing and Opinion Research (CMOR) was formed by the American Marketing Association (AMA), Council of American Survey Research Organizations (CASRO), Advertising Research Foundation (ARF), and Marketing Research Association (MRA) specifically to address respondent cooperation issues.[1] CMOR developed professional guidance for researchers in eliciting cooperation from respondents, hosted numerous respondent cooperation themed workshops and conferences attended by researchers from diverse backgrounds, conducted basic research into problems of nonresponse, and even launched a public relations campaign under the auspices of the "Your Opinion Counts" program on behalf of opinion researchers.

Researchers have invested great amounts of effort in maximizing respondent cooperation. Perhaps most significantly, the results of these labors have produced a generous understanding of methods and techniques for designing research surveys that appeal to respondents. Still, concerns over respondent cooperation have persisted, driven both by common knowledge, as well as empirical data (de Leeuw 2002).

Perceptions of Research The public's view of and attitudes toward research may be tied to both their willingness to participate in research studies (i.e., respondent cooperation) as well as their honesty when participating. Moreover, it also serves as an indicator of society's willingness to tolerate research activities. Several public opinion polls have gauged attitudes towards survey research and have found mixed results.

For example, Louis Harris and Associates found that, between 1998 and 2006, the public steadily lost trust in public opinion pollsters as professionals that "tell the truth." Initially, in 1998, over half of the American population (55%) trusted pollsters. By 2006, trust levels were down to 34%, a 21% point decline that resulted in comparable levels of trust with such professions as members of Congress (35%) and trade union leaders (30%) (Kim 2011).

Moreover, in three surveys conducted by the MRA between 2001 and 2006, only 27–29% of the American public felt as though survey organizations would respect their rights to privacy (Kim 2011; MRA 2006). This is a foreboding statistic, since privacy and trust are critical components to the research process. Researchers rely on respondents to give them honest information, sometimes regarding sensitive, or embarrassing topics. Respondents may be wary to provide researchers with information unless they have confidence that their data will remain confidential.

However, the usefulness of polls and surveys fair better than the perception of pollsters and trust of survey organizations. For example, the MRA study found that between 60 and 65% of Americans believe that polls give the public an opportunity to provide feedback into public policy issues. Additionally, between 67 and 68% in the same period indicated a belief that surveys help manufacturers produce better products and services (Kim 2011; MRA 2006).

12.4.1 Response from Researchers

Researchers have taken several approaches to alleviating concerns that relate to respondent cooperation. These have included attempts at raising the willingness of the public to participate as well as developing coping strategies for low respondent cooperation. Efforts to handle issues of respondent cooperation have included: professional standards of practice, public relations campaigns, the development of techniques aimed at improving respondent cooperation and evaluation, and management techniques for nonresponse as it occurs in sample surveys.

Professional Standards. One of the fundamental ways in which researchers have attempted to improve favorability with the public is through ensuring a level of protection for respondents. Professional and trade associations have been the traditional mechanism for developing these protections, since they provide a conduit for independent researchers, research

[1] CMOR was found with two principal missions: respondent cooperation and government affairs/advocacy. The organization would later merge with the MRA.

organizations, and other stakeholders to build consensus around ethical practices.

In the United States, there are several associations of researchers and research organizations that have developed enforceable codes of ethics. To be a member of one of these associations, an individual or company must agree to uphold, and be accountable to the organization's rules. For example, among a lengthy list of additional standards, members of the MRA must agree to uphold the following (Marketing Research Association 2007a, b):

- Will protect the rights and privacy of respondents.
- Will treat respondents in a professional manner.
- Information will not be revealed that could be used to identify respondents without proper authorization.
- Will take all reasonable precautions that respondents are in no way directly harmed or adversely affected as a result of their participation in a marketing research project.
- Will not abuse public confidence in opinion and marketing research.
- Will identify surveys and other methods of data collection as such and not attempt to collect data through casual or conversational means other than for bona fide mystery shopping assignments.
- Will respect the respondent's right to withdraw or to refuse to cooperate at any stage of the study and will not use any procedure or technique to coerce or imply that cooperation is obligatory.
- Will treat the respondent with respect and not influence a respondent's opinion or attitude on any issue through direct or indirect attempts, including the framing of questions.
- Will ensure that information collected during any study will not be used for any sales, solicitations, or Push Polling.

In addition to professional association standards, researchers must comply with relevant Federal and state laws that relate to research methods as well as the rules of their organization, institution, and/or sponsor. Together, these ethical practices provide a baseline that helps to define the practice of opinion research as well as ensure a consistent level of protections for respondents. (Earlier in this handbook Oldendick—Chap. 3—discussed the essentials of survey ethics while providing a clear account of what is allowed and not allowed in survey methodology).

Research Industry Practitioner Public Relations Efforts Researchers have embarked on a number of high profile public relations campaigns in order to encourage respondent cooperation. For example, CMOR developed the "Industry Identifier" program which promotes a recognizable image and "Your Opinion Counts" (YOC) expression aimed at improving society's understanding and acceptance of research through education and outreach.

The Industry Identifier allows organizational members, who may be research provider and other stakeholder entities, to utilize the YOC seal and expression as long as they subscribe to an accepted industry enforceable code of ethics (currently only including MRA), agree to uphold a "respondent bill of rights," and gain acceptance to the program through the submission of an application. The respondent bill of rights is written below (Marketing Research Association 2011).

- *Your privacy and the privacy of your answers will be respected and maintained.*
- *Your name, address, phone number, e-mail, personal information, or individual responses will not be disclosed to anyone outside the research project without your permission.*
- *You will always be informed in advance if an interview is to be audio recorded or video recorded (as in the case of telephone or in-person studies). Additionally, you will be told of the intended use of the recording.*
- *Upon request, you will be informed of the privacy policy that applies to your participation in the research study.*
- *The researcher will be identified to you. You will be told the name of the research organization and the general nature of the survey.*
- *You will not be sold anything, or asked for money, under the guise of research.*
- *You will be contacted at reasonable times, but if the time is inconvenient, you may ask to be re-contacted at a more convenient time.*

- *Your decision to participate in a study, answer specific questions, be re-contacted at another time, or discontinue your participation will be respected.*
- *You are assured that the highest standards of professional conduct will be upheld in the collection and reporting of information you provide.*

MRA maintains and enforces the "Your Opinion Counts" program through adjudication of complaints made toward members of the program by respondents and members of the public-generally submitted through the YOC website. MRA also participates in public outreach in terms of educating the public of the importance of research.

Past public relations efforts have included press and media communications, promotional efforts, grass-roots campaigns launched via association members to the public, and even the development of a cartoon designed to illustrate the usefulness of survey and opinion research. However, the most far-reaching of efforts may arguably be through program members whom communicate the YOC seal and expression to the public on a daily basis while initiating surveys and making survey requests.

While the YOC effort is an example of a significant attempt to improve society-level respondent cooperation from the for-profit and non-profit sectors. Additional attempts have also been made from other sources. One particularly well-known outreach campaign occurred by the Bureau of the Census (US Census) before and during the year 2010 enumeration.

US Census Respondent Cooperation Efforts A census of the US population is mandated by the Constitution of the United States to occur every 10 years (Article 1, Section 2). This activity is carried out by the Bureau of the Census, which also conducts several other important surveys of the population that include the Economic Census, Census of Governments, American Community Survey, and other surveys (US Bureau of the Census 2011). Census data is used for such purposes as determining the distribution of Congressional seats, making decisions about the provision of community services, and determining the distribution of federal funds to local, state and tribal governments (US Bureau of the Census 2011).

With these and many other important uses on the line, it is critical that the US Census make as complete a count of the US population as possible. As such, the organization puts forth tremendous effort in eliciting high respondent cooperation. Recent Census enumerations have cost in the several billions of dollars to achieve this end. However, even with extraordinary techniques and the force and potential threat of mandatory legal compliance (respondents are technically forced by law to complete their census requests), the US Census still achieves imperfect respondent cooperation. For example, the final response rates for the year 2000 and year 1999 Censuses were 67 and 65%, respectively (United States Bureau of the Census 2004).

The decennial census first attempts to collect data through a postal mail survey, followed by in-person enumerator/interviewer follow-up. For the most recent Census (year 2010), it is estimated that every percentage point of mail survey response (versus in-person follow-up) saves about 85 million dollars (Nasser 2010). This is because, for the 2010 Census, it cost about 0.42 cents to process a postal form compared to about $57 dollars for an in-person Census interview (Scott 2010).

To encourage individuals to participate in the US Census effort, the Bureau launched an extensive and broad-reaching public relations campaign geared to both educate, and encourage the public to participate in the Census. Efforts included the use of famous individuals and spokespersons, respected community groups, a dedicated and easy-to-understand website, as well as a host of other initiatives. A few of the innovations employed by the Census included:

(1) An online interactive map of the United States that listed participation rates for neighborhoods, towns, or states, which allowed members of the public to see the relative performance of their communities in participating in the US Census.
(2) Featured personal stories from individuals, which demonstrated to members of the public that their peers participated and found value in the Census.

(3) A Census "road-tour" of the USA which featured a number of unique live and online events, including social media network outreach and blogging, online videos, and picture posting.
(4) A clear and accessible Census website that covered important information about the US Census Bureau, the census effort, and other important topics, such as information privacy, and security.

Additionally, the Census outreach program included television and media commercials, interviews, a blog regularly updated by the Director of the Census, Robert M. Groves, and, a very inclusive language assistance program. (Glaser 2010).

Recent Perspectives on Respondent Cooperation Recent perspectives on respondent cooperation have focused beyond improving the willingness of individuals to participate in research, and have also considered the manner and quality of which they participate. For example, the MRA, a professional trade association found in 1957 representing around 2000 for-profit, government, and academic researchers officially expanded their definition of respondent cooperation in 2009: *Respondent Cooperation is the degree and quality with which the public participates in research* (Glaser 2009).

The expanded focus on quality of participation is driven largely by the rising popularity of Web surveying. Notably, web surveys have had a profound effect on the manner and quality of respondent participation as well as the motivators of participation. Since Internet surveys are self-administered, there is no interviewer to help assist the respondent or motivate them to put the required cognitive effort into answering each question. Moreover, many respondents find Web formatted surveys to be a tedious undertaking, particularly when they are designed in a redundant, and burdensome manner.

Additionally, the common practice of incentivizing web survey respondents for the sake of cooperation has succeeded in raising many surveys response rates. However, as explained below, incentives may have also played a role in the decline of response quality. For this reason, researchers must weigh the response rate advantages that incentives may provide along with the potential for a decrease in response quality. This is particularly true for those researchers that utilize volunteer opt-in panels as sample sources.

It is important to note that, the view that respondent cooperation refers to the manner and quality of which a respondent participates (in addition to willingness) is distinctly one that has gained popularity in and among marketing researchers and other nonacademic oriented practitioners. Traditional views in the social sciences continue to focus on respondent cooperation as a concept that specifically relates to the willingness of an individual to participate in research. However, issues surrounding the honest and thoughtful participation of respondents are still very important and often fall under the heading of "measurement error."

Special Cooperation Issues for Volunteer Opt-in Panels It has become very common to incentivize Web respondents, particularly those that belong to volunteer opt-in panels who participate under a regular basis. Respondents in these panels typically register with a panel provider organization and participate over-time to accumulate prizes or awards. Since many of these respondents are primarily motivated to participate in research for monetary incentives, and not altruistic or other motivations, data quality problems may arise.

One particularly critical issue to researchers that utilize volunteer opt-in panels, rather than sample drawn from other sources, is that some respondents may find it advantageous to provide dishonest answers on surveys in hopes of qualifying for the research (i.e., meeting the screening criteria the researcher employs to target a specific group). Individuals that provide deliberately dishonest or false information in order to qualify for panel membership or survey screening criteria are sometimes called "professional respondents" by survey researchers.

An additional, equally concerning issue that is particularly relevant to panel sample is satisficing, where the respondent may not put forth

sufficient effort in answering survey questions. Again, this is thought to relate to the respondent's primary motivation being the completion of many surveys, quickly; rather than, the deliberate and purposeful participation in research to contribute to product, service, or policy advancements. (Later in this handbook—Chap. 20—Toepoel presents a thorough account of such online panel surveys and their potential limitations and sources of bias).

Weighing Respondent Cooperation and Quality Researchers have begun to understand both the error that nonresponse produces as well as the situations where nonresponse may pose particular threats to a research effort. The traditional nonresponse paradigm has shifted. In the past, researchers believed that the larger the nonresponse to a survey effort, the greater the error. Today, it is widely accepted that a certain amount of nonresponse in surveys is inevitable, but that nonresponse only poses practical threats to the quality of the research where it is systematically linked to both the survey variables of interest, as well as the likelihood of respondents participating.

Researchers now must consider whether the techniques they would employ to raise cooperation for a survey effort may be creating more error than they are reducing. As stated by Groves, "*blind pursuit of high response rates in probability samples is unwise; informed pursuit of high response rates is wise*" (Groves 2006). There are a wide variety of techniques and tools available to researchers in their efforts to achieve high response rates for survey projects. To be employed effectively, with the goal of improving overall quality, the researcher must consider techniques in the light of nonresponse biases-not simply basic respondent cooperation. Readers are referred to the sources available at the end of this chapter for further information on this subject.

12.4.2 Future Implications

The future viability of research is tied to respondent cooperation in important ways. First, traditional concerns over respondent cooperation and nonresponse to surveys persist. Moreover, the ease and cheapness with which surveys can now be deployed, due to cost effective Internet methods, only stands to increase the number of survey requests individuals receive. Ultimately, more requests may prove to burden the public and lessen their willingness to respond to surveys. Moreover, the novelty of participating in a survey may decrease as opportunities rise.

Second, information privacy threats have risen as data is more easily and cheaply stored and transmitted through computers. The results of these technology advances are that researchers must grow to consider information privacy and security much more carefully as they design and conduct research. Information security breaches, like those now commonly reported in the media, are likely to continue for the foreseeable future. Without assurances that researchers can be trusted to protect personal and sensitive information, the public may grow to become highly distrustful of research professionals.

Respondent cooperation is a fundamental concern to researchers. However, it is traditionally considered as most important in the context of sample surveys aimed at projecting findings to a target population of interest. Other methods of information gathering and analysis are currently growing in popularity and being refined, such as data mining, social media research and other passive, and observational techniques. Oftentimes, these leading-edge methods do not require traditional respondent cooperation and have the potential to reduce or eliminate the burden of participation for the respondent.

New methods, such as those described above, may assist the researcher and offer them complementary tools to answer their research questions. Importantly, they may also be used in conjunction with traditional research methods to help evaluate the accuracy and reliability of surveys and other techniques. As societal patterns and technologies have changed, so have both the challenges and opportunities facing survey and opinion researchers.

References

American Association for Public Opinion Research. (2011). *Standard definitions.* Retrieved April 6, 2011, from AAPOR http://www.aapor.org/Standard_Definitions2.htm

Association Collaborative Effort. (2009). *ACE glossary*. Retrieved April 6, 2011, from ACE http://www.aceinfo.org/definitions.cfm

Braverman, M. T. (1996). *Advances in survey research*. San Francisco: Jossey-Bass Publishers.

de Leeuw, E. A. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves (Ed.), *Survey nonresponse* (pp. 41–54). New York: Wiley.

Dillman, D. A. (2002). Survey nonresponse in design, data collection, and analysis. In D. A. Dillman (Ed.), *Survey nonresponse* (pp. 8–12). New York: Wiley.

Dillman, D. A. (2009). *Internet, mail, and mixed-mode surveys: The Tailored design method*. Hoboken: Wiley.

ESOMAR. (2011). *ESOMAR professional standards.* Retrieved April 06, 2011, from ESOMAR http://www.esomar.org/index.php/professional-standards.html

Glaser, P. (2009). *Respondent cooperation: A new definition for a new relationship*. Glastonbury: Marketing Research Association.

Glaser, P. (2010). Census update 2010. Internal Report submitted, Marketing research Association. CT.

Groves, R. M. (2000). Leverage saliency theory of survey participation: Description and illustration. *Public Opinion Quarterly, 64*, 299–308.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646–675.

Groves, R. M., & Couper, M. P. (1998). *Non-response in household interview surveys*. New-York: Wiley.

Johnson, T. P. (2002). Culture and survey nonresponse. In R. M. Groves (Ed.), *Survey nonresponse* (pp. 55–69). New York: Wiley.

Johnson, D. R. (2008). *Using weights in the analysis of survey data.* Retrieved May 24, 2011, from Population research institute, Penn State University, http://help.pop.psu.edu/help-by-statistical-method/weighting/Introduction%20to%20survey%20weights%20pri%20version.ppt

Kim, J. G. (2011). Trends in surveys on surveys. *Public Opinion Quarterly, 33*, 165–191.

Lavrakas, P. (2008). Economic exchange theory. In P. Lavrakas (Ed.), *Enclyclopedia of survey research methods*. Thousand Oaks, CA: Sage.

Marketing Research Association. (2007a). *MRA expanded code of market research.* Retrieved April 6, 2011, from Marketing Research Association http://www.mra-net.org/resources/documents/CodeMRStandards.pdf

Marketing Research Association. (2007b). *The code of marketing research standards*. Glastonbury: MRA.

Marketing Research Association. (2011). *Your opinion counts: Respondent bill of rights*. Retrieved May 22, 2011, from Your Opinion Counts, http://www.youropinioncounts.org/index.cfm?p=rights

MRA. (2006). *MRA research profession image study*. Glastonbury: Author.

Nasser, H. E. (2010). *2010 Census response rate surprisingly close to 2000 rate*. Retrieved May 22, 2011, from USA Today, http://www.usatoday.com/news/nation/census/2010-04-20-census-participation-rate_N.htm

Roose, H. W. (2002). *Response behavior in audience research: A two-stage design for the explanation of non-response.* Retrieved May 23, 2011, from http://mrvar.fdv.uni-lj.si/pub/mz/mz18/roose.pdf

Scott, M. (2010). *States could lose $1,400 per person missed in 2010 census count*. Retrieved May 22, 2011, from Daily Finance, http://www.dailyfinance.com/2010/03/29/states-could-lose-1-400-per-person-missed-in-2010-census-count/

Singer, E. (2006). Introduction: Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*, 640–641.

Smith, T. W. (1999). *Developing nonresponse standards*. Chicago: NORC Report.

United States Bureau of the Census. (2004). *2000 Census response rates*. Retrieved May 22, 2011, from Census.gov, http://www.census.gov/dmd/www/response/2000response.html

US Bureau of the Census. (2011). *About us*. Retrieved May 22, 2011, from Census.gov, http://www.census.gov/aboutus/

Further Reading

Dillman, D. A. (2009). *Internet, mail, and mixed-mode surveys: The Tailored design method*. Hoboken, NJ: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves et al. (2002). *Survey nonresponse*. New York: Wiley.

Effects of Incentives in Surveys

13

Vera Toepoel

"Total Survey Error (TSE) refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. In this context, a survey error is defined as the deviation of a survey response from its underlying true value" (Biemer 2010, p. 817). Survey resources can be allocated to reduce TSE. Nonresponse is an important indicator of TSE, and incentives are widely used to increase response rates. This chapter discusses the theories behind incentive effects, the possible forms of incentives and related effects, estimates the optimal amount of incentives, handles different modes of data collection, the relation between incentives and data quality, and ends with best practices and a view towards the future in order to help survey researchers in identifying if, how and how much incentives should be used in their surveys. The topic of incentives is also discussed in many of the other chapters in this handbook, as it relates to nonresponse error, increase of response rates, and conversion of refusals.

V. Toepoel (✉)
Leisure Studies, Tilburg University,
Warandelaan 2, P.O.Box 90153,
5000LE Tilburg, The Netherlands
e-mail: V.Toepoel@uvt.nl

13.1 Theories of Incentive Effects

The power of survey data is the cumulative result of many individual decisions to participate in a survey. There is a large literature on methods to increase cooperation in surveys. Groves et al. (1992) outline a model that specifies factors influencing survey participation. They integrate the influence of sociodemographic and survey design factors on the one hand, with the impact of psychological components on the other. *Societal-level factors* are characteristics of a culture in which the survey takes place. For example, the number of surveys conducted in a country and the perceived legitimacy of surveys influence the survey taking climate. An "over surveying" effect takes place when individuals are confronted with too many requests for filling out surveys, which can lower the overall response rates in a particular culture. *Design choices* may have an important impact on survey participation as well. The mode of (initial) contact, persuasion strategies used, length, and difficulty of the survey, layout, interviewer qualities and attributes (in face-to-face and telephone surveys), and topic of the survey (Dillman 2007) etc., influence whether a respondent is willing to fill out a survey. Third, response rates have been shown to vary across *characteristics of the sample persons*, such as age, gender, income, and education (Groves 1989). In addition to these sociodemographic and survey design factors, Groves et al. (1992) identify psychological concepts relevant

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_13, © Springer Science+Business Media New York 2012

to survey responding: *compliance with requests, helping tendencies, and opinion change*. Singer and Bossarte (2006) use different terms for similar concepts: altruism (fulfilling a social obligation) and egoistic reasons (respondents like to give their opinion or are in it for the money).

Groves et al. (1992) argue that people follow simple heuristics in deciding to participate in a survey:

- *reciprocation* (copy behavior that is received or seen from others),
- *consistency* (a desire to behave consistent with one's attitudes and beliefs),
- *social validation* (use beliefs of similar others as standards of comparison),
- *authority* (more compliance if a request comes from a legitimate authority),
- *scarcity* (opportunities that appear to be scarce are perceived as more valuable), and
- *liking* (people are more eager to comply towards those they like).

The use of incentives for increasing response rates is primarily based on the principle of reciprocation. The reciprocity heuristic for compliance implies that respondents are more willing to comply with a survey request to the extent that the compliance constitutes the repayment of a perceived gift, favor, or concession (Groves et al. 1992, p. 480). Incentives can be used to abide to people's norm of reciprocity because they create a sense of obligation to potential respondents. This heuristic also implies that prepaid incentives should work better than postpaid incentives in increasing response rates; since giving a prepaid incentive builds a psychological obligation to reciprocate that can be fulfilled by filling out the survey. Jobber and Saunders (1988) state that it is the paying of an incentive rather than the amount paid which is the major trigger for the dissonance and reciprocity which underlie incentive effectiveness.

Dillman (1978) refers to a similar principle with *social exchange theory*, which presumes that actions of individuals are motivated by the return these actions are expected to bring from others. This theory supposes prepaid incentives to be a symbol of trust that the potential respondent is willing to answer the survey. Dillman argues that there are three things that determine the decision to participate in a survey: rewards (what the respondents expects to gain from the experience), costs (what is used to obtain the reward), and trust (the expectation that in the long run the rewards will outweigh the costs). According to Dillman, researchers should minimize the costs of responding, maximize the rewards for doing so, and establish trust that rewards will be delivered.

Economic exchange theory (Biner and Kidd 1994) proposes that respondents choose to participate in a survey after making a rational cost-benefit calculation (Ryu et al. 2005). An invited individual considers an incentive as a compensation for the time and effort required for filling out the questionnaire. This might suggest that incentives particularly attract low-income individuals, and that higher amounts of incentives might elicit higher response rates (in contrast to Jobbers and Saunders' (1988) statement that it is about the paying of an incentive rather than the amount paid). In addition, economic exchange theory implies that money is more powerful than nonmonetary incentives; since cash has a universally understood value.

Groves et al. (2000) outline "*leverage-saliency theory*" to define motives for survey participation. This theory suggests that the effect of a survey attribute on a person's response decision depends on the importance each person places on the attribute and how salient it is made during the survey request (Medway et al. 2011). For example, people might differ in the importance they place on survey specific factors such as topic and sponsorship, privacy issues, interest, etc. This would suggest that incentives can have various effects for different subgroups and across designs. For example, interview-administered surveys may show smaller effects of incentives since interviewers can play a role in persuading people to participate by making special features of the survey more salient (Ryu et al. 2005). Singer et al. (1999a) argue that incentives are particularly effective when other motives of participation (e.g. intrinsic motivation) are absent.

Following Göritz (2005) and Sanchez-Fernandez et al. (2010), *Reactant theory* (Brehm and Brehm 1981) can be applied to postpaid or

Table 13.1 Theories of incentive effects

Theory	Expected incentive effect
Heuristic rule for compliance (Groves et al. 1992) Respondents should be more willing to comply with a survey request to the extent that the compliance constitutes the repayment of a perceived gift, favor, or concession.	Giving a prepaid incentive builds a psychological obligation to reciprocate that can be fulfilled by filling out the survey.
Social exchange theory (Dillman, 1978) actions of individuals are motivated by the return these actions are expected to bring from others.	Prepaid incentives are a symbol of trust that the potential respondent is willing to answer the survey and should therefore be successful in increasing response rates.
Economic exchange theory (Biner and Kidd 1994) Respondents choose to participate in a survey after making a rational cost-benefit calculation.	Higher amounts of incentives should elicit higher response rates. Money is more powerful than non-monetary incentives; since cash has a universally understood value.
Leverage-saliency theory (Groves et al. 2000) The effect of a survey attribute on a person's response decision depends on the importance each person places on the attribute and how salient it is made during the survey request.	Incentives can have various effects for different subgroups and across designs.
Reactant theory (Brehm and Brehm 1981) If behavioral freedom is threatened, individuals will exert an adverse state of excitement (reactance) that will cause them to want to reduce this excitement by reacting to the threatened freedom.	If individuals are invited to take part in a survey where incentives are offered in exchange for a response (postpaid), individuals would react and fill out the survey, and react to the experienced threat of freedom.

conditional incentives. Reactant theory states that if behavioral freedom is threatened, individuals will exert an adverse state of excitement (reactance) that will cause them to want to reduce this excitement by reacting to the threatened freedom. Therefore, if individuals are invited to take part in a survey where incentives are offered in exchange for a response, individuals would react and fill out the survey, and react to the experienced threat of freedom (Table 13.1).

13.2 Types of Incentives

Church (1993) divides incentives into four types. Monetary and nonmonetary incentives provided with the survey and monetary and nonmonetary incentives given on the return of the questionnaire. Nonmonetary incentives are diverse in nature and can range from lotteries, donations to charities, and gifts such as iPods, pens, stamps, and so on. Prepaid or unconditional incentives are given to all individuals invited to take part in a survey regardless of whether they finally decide to complete the survey. Postpaid or conditional are only given to those individuals who take part in the survey (and finish it).

Many experiments have been conducted on the role of incentives in stimulating survey participation. The overall results suggest that prepaid incentives are more powerful than postpaid (conditional) incentives and that money is more powerful than nonmonetary incentives (Eyerman et al. 2005; Göritz 2004; James and Bolstein 1992; Jobber and Saunders 1988; Larson and Chow 2003; Messer and Dillman 2010; Petrolia and Bhattacharjee 2009; Singer et al. 2000; Warriner et al. 1996). This conclusion is confirmed by three meta-analyses. Church (1993), in his meta-analysis of 74 mail surveys, finds an overall increase in response rates of 12 % points between incentive and control conditions. Prepaid monetary incentives work best (an average increase in response rate of 19 % points), followed by prepaid nonmonetary incentives (8 % points), postpaid monetary incentives (5 % points) and nonmonetary (1 % point) incentives, as can be seen in Fig. 13.1. While Church (1993) estimates the results of incentives in mail surveys, Singer et al. (1999a)

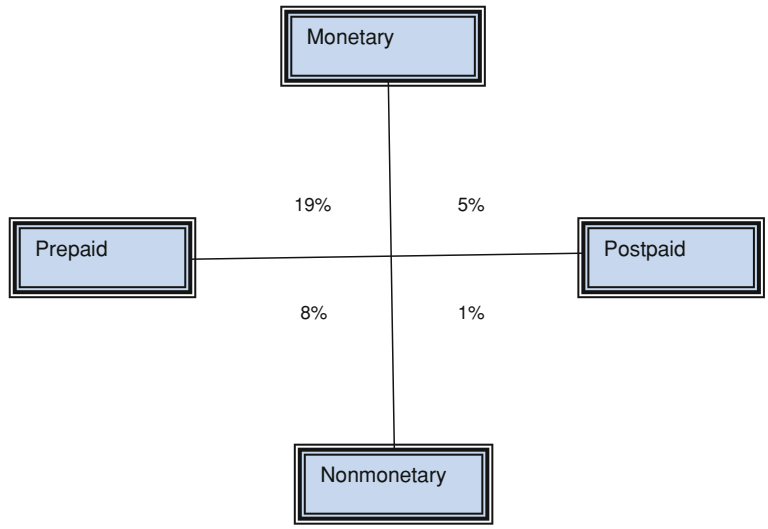


Fig. 13.1 Types of incentives and their effectiveness on response rates (Church 1993)

finds similar effects in a meta-analysis in interview-administered surveys and Göritz (2006) in Web surveys. Empirical results confirm the abovementioned theories.

Göritz (2005) conducted an experiment on the timing of the incentive. She fielded five online experiments in which two conditions were contrasted: one group received an e-mail invitation that all participants are eligible for an incentive (= unconditional incentive), whereas the other group received an invitation in which participants were told that participants who responded would receive an incentive (= conditional incentive). When combining the five experiments in a meta-analysis she finds that the groups with conditional incentives show lower response rates than the unconditional incentives groups. This confirms the results in the literature showing that prepaid incentives work better than postpaid incentives.

Donations to charity and lottery tickets were new measures of increasing response rates in the 1980s. Hubbard and Little (1998) showed that people who were offered a donation to charity as an incentive had a lower response than the no incentive control group. Regardless of this effect, Warriner et al. (1996) used charitable donations since they thought the topic of their survey, environmental issues, offered some promise for success. They argued that donations to charity offer respondents the opportunity to confirm to a norm of social responsibility by having caused the charitable donation to be made. They did not find a positive effect of charitable donations on response rates, however. Göritz and Neumann (2011) find a significant negative effect of donations on response rates in a longitudinal online experiment. The negative effect was larger for a predetermined charity than for a selectable charity. In all, donations to charity do not seem to work in increasing response rates. Donating an incentive to charity could overstrain respondent's willingness to help others (altruism). Respondents might not appreciate the incentives not coming in their own hands; they might feel they should get the incentive because they themselves had the burden of completing the survey.

Research on lottery incentives indicates that they have little to no effect on response rates either (Church 1993; Harris et al. 2008; James and Bolstein 1992; Singer et al. 2000; Warriner et al. 1996). Only Budowski and Scherpenzeel (2004) found the inclusion of a lottery ticket to improve response rate by 4 % ($p < .10$) in a longitudinal panel study. This could be due to the fact that in a panel, trust between the research agency and the respondent is already established and panel members are more used to be paid after completing the survey. Although the evidence of

Table 13.2 Use of lottery incentives amongst institutional researchers

Number of surveys with prizes	*N*	%	Types of prizes	%
0	222	63	Cash	22
1	75	21	Gift certificate-national	12
2	31	9	Gift certificate-local	23
3	11	3	Gift certificate-school	57
4	5	1	Travel prize	6
5+	3	1	Electronics	6
Total	347	100	Clothing	15

Source Adapted from Porter and Whitcomb (2003)

lottery incentive effects is scarce, many people still apply this type of incentives in their surveys. Porter and Whitcomb (2003) asked institutional researchers about their use of lottery incentives. Table 13.2 shows that about one-third of the researchers used lottery incentives in their surveys. The most popular type of lottery incentive was a school related gift certificate.

Although there is little to no empirical evidence that lottery incentives work, use of them appears to be common practice. Researchers may feel the need to provide respondents with an incentive and are attracted to the lower overall incentive costs associated with lottery tickets. They seem to discard the lack of effect of this type of incentives.

The effect of sending nonmonetary gifts such as coffee, books, pens, key rings, golf balls, tie clips, stamps, or regional park passes in advance is relatively unclear. For most studies they do not boost response rates. (Singer et al. 1999a). Sending such types of incentives does not seem to be worth the cost and effort involved. In all, prepaid monetary incentives are the only type of incentive to show a structural effect on response rates, thereby reducing the TSE.

13.3 Amount of Incentives

Since prepaid monetary incentives work best in increasing response rates, the question rises what the amount of the prepaid incentive should be. The most commonly used prepaid monetary incentive is $1 or $2 (Porter and Whitcomb 2003), as also suggested by Dillman (1978) in his Tailored Design Method. Some surveys use larger incentives of $5 (Singer et al. 2000) or $10 (Warriner et al. 1996).

Warriner et al. (1996) found a prepaid incentive in the amount of $2 and $5 significantly increasing response rates, but the increment for using $10 in place of $5 was negligible. This result is partly replicated by Trussel and Lavrakas (2004). They compared levels of incentives varying from $0 to $10 in a mail survey, with the exception that there was no $9 condition. They focused on three different types of household addresses that originally were sampled via a random digit dial (RDD) frame: (1) households that agreed to participate; (2) households that were never contacted in the RDD stage despite numerous contact attempts; and (3) households that had refused to participate when contacted in the RDD phase. Results are presented in Table 13.3.

Among accepting households, all incentive conditions had a significantly higher likelihood of cooperation compared to the no-incentive condition. In addition, the conditions in which at least $5 was sent showed significantly higher cooperation rates than the conditions in which less than $5 was sent, but the differences between the $6 and $10 conditions were not significant. Among noncontacted households there were no significant differences between the $1 and $4 conditions and the no incentive condition. The $5–$10 conditions showed a significant increase compared to the no incentive condition. Among refusing households, the $4–$10 conditions showed a higher cooperation rate than the no incentive condition. In all, those respondents

Table 13.3 Cooperation rates on mail surveys for household types

Condition	Incentive amount (in $)	Accepting households cooperation rate in %	Noncontacted households cooperation rate in %	Refusing households cooperation rate in %
A	0	37.9 b,c,d,e,f,g,h,i,j ($n = 541$)	10.1 f,g,i,j ($n = 99$)	5.5 c,d,e,f,g,h,i,j ($n = 201$)
B	1	50.6 a,f,g,h,i,j ($n = 896$)	12.8 j ($n = 188$)	7.7 e,f,g,h,i,j ($n = 364$)
C	2	50.3 a,f,g,h,i,j ($n = 1634$)	9.2 d,e,f,g,h,i,j ($n = 305$)	10.2 e,f,g,h,i,j ($n = 742$)
D	3	51.3 a,f,g,h,i,j ($n = 2132$)	14.5 c,j ($n = 408$)	11.0 e,f,g,h,i ($n = 990$)
E	4	52.8 a,f,g,h,i,j ($n = 6759$)	13.3 c,f,g,i,j ($n = 1311$)	14.8 a,b,c,d,e,f,g,h,j ($n = 3004$)
F	5	55.2 a,b,c,d,e,f ($n = 7085$)	16.6 a,c,e,j ($n = 1308$)	16.8 a,b,c,d,e,h,j ($n = 3237$)
G	6	56.6 a,b,c,d,e ($n = 6997$)	16.7 a,c,e,j ($n = 1392$)	17.1 a,,c,d,e,h,j (3294)
H	7	57.0 a,b,c,d,e ($n = 2408$)	14.8 c,j ($n = 473$)	21.2 a,b,c,d,e,f,g ($n = 1133$)
I	8	58.5 a,b,c,d,e,f ($n = 1096$)	19.6 a,c,e ($n = 204$)	18.6 a,b,c,d ($n = 414$)
J	10	58.2 a,b,c,d,e ($n = 692$)	26.4 a,b,c,d,e,f,g,h ($n = 121$)	23.5 a,b,c,d,e,f,g ($n = 247$)

Source Trussel and Lavrakas (2004)
a $p < .05$ compared to $0
b $p < .05$ compared to $1
c $p < .05$ compared to $2
d $p < .05$ compared to $3
e $p < .05$ compared to $4
f $p < .05$ compared to $5
g $p < .05$ compared to $6
h $p < .05$ compared to $7
i $p < .05$ compared to $8
j $p < .05$ compared to $10

who agreed to cooperate did in fact show higher cooperation rates than no contacted or refusing households. For this type, $1 can be used to increase response rates. For noncontacted and refusing households, who have a neutral or negative experience at the first contact, each incremental dollar has a slightly larger impact beyond the previous amount that was $1 less in value. This would suggest that for these groups, researchers should send as much as one's budget can afford (defined by Trussel and Lavrakas (2004) up to $10). The study by Trussel and Lavrakas (*ibid*) is important since it demonstrates the importance of prior contact in mediating the effects of incentives on response rates.

Jobber, Saunders, and Mitchell (2004) conducted a meta-analysis on 30 experiments with prepaid monetary incentives. They found that the inclusion of any incentive, regardless of amount, raises the response rate by an average of 15 % across all studies. They conclude that the inclusion of an incentive will lower the cost per completed return if the proportionate rise in cost is smaller than the proportionate gain in response. "So, for example, if printing and mailing cost $1.50 per questionnaire… and there is a baseline response rate of 20%, then a 25 cent or even a $1 incentive would be justified by the expected rise of 15–17% in the response rate… However, if we estimate printing and mailing costs of $30 or more…and a baseline response rate of 20%, then incentives of $20 or more might be justified." (Jobber et al. 2004, p. 24).

Saunders et al. (2006) optimize the amount of monetary incentives in mail surveys. They have developed a cost-benefit model based on 36

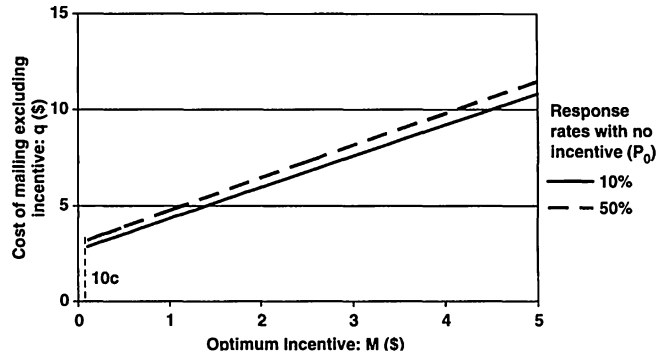


Fig. 13.2 The relationship between mailing costs and the optimum incentive for response rates (without incentive) of 10 and 50 % found by Saunders et al. (2006). *Source*: Saunders et al. 2006, p. 1228

experiments with prepaid monetary incentives to estimate the optimum incentive when attempting to minimize overall survey costs for a given sample size. Data were collected from 1975 to 2002.

"The amount to be spent on a monetary incentive depends upon the cost of mailing[1] without an incentive. The cost (v) of a survey is

$$v = \frac{n}{Pm}(q + M) + q_0 \qquad (13.1)$$

where n is the sample size required, P_M is the expected response rate with incentive M, q is the variable cost of mailing excluding the incentive, and q_0 is the fixed cost of mailing excluding the incentive." (Saunders et al. 2006, p. 1228). The optimum monetary incentive is given by: "

$$q = \frac{1}{m}\left[\frac{(M+1)^{m+1}}{I^i(1-P_0)/P_0} + M(1-m) + 1\right] \qquad (13.2)$$

…This means the optimum monetary incentive (M) depends on: the variable cost of mailing a questionnaire excluding the incentive (q); the survey response rate with no monetary incentive (P_0); and the coefficient for the mail incentive (m). I^i is a constant reflecting the impact of providing any mail incentive."(Saunders et al. 2006, p. 1228).

Although the survey response rate without incentive has some influence on the optimum prepaid monetary incentive, survey costs are bigger drivers for the optimum amounts of money spend. This is demonstrated by Saunders et al. (2006) in Fig. 13.2. The figure shows the difference in the optimum prepaid monetary incentive for response rates (without incentive) of 10 and 50 %. The figure clearly shows that survey costs are more important in estimating the optimum amount of a prepaid incentive than the response rate without incentive. This is a good thing when a researcher can make a better prediction of expected mailing costs than the expected response rate. The latter is often difficult to predict. When survey costs are low, the optimum prepaid monetary incentive is low, since a large value incentive would represent a high proportion of overall costs. When survey costs are high, a large incentive becomes feasible as it accounts for a smaller percentage of total costs (Saunders et al. 2006). In other words, as the variable survey costs increase, so does the size of the optimum prepaid monetary incentive. Saunders et al. (2006) conclude that for mailing costs below $2.50, a minimum incentive level (e.g. 10 cents) should be used. One could argue, however, if a 10 cent incentive could do the trick.

[1] Saunders et al. (2006) base their analysis on mail surveys, but results can easily be extended to other modes of administration, with Web-based surveys having lower variable costs and telephone and face-to-face surveys having higher variable costs than mail surveys.

13.4 Incentive Effects and Data Quality

Petrolia and Bhattacharjee (2009) found incentives to have little effect on item nonresponse. Sanchez-Fernandez et al. (2010) confirm this result and conclude that the use of preincentives or postincentives, either singly or together, has no significant effect on response quality measured in terms of the number of missing data. This is confirmed by Medway et al. (2011). Göritz (2005) compares conditional and unconditional incentives and finds no significant differences with regard to data quality (measured by retention, item nonresponse, length of answers, and stereotypical answering of grid questions) either. In addition, Göritz (2004) finds no significant differences in the number of omitted items and number of response sets between monetary and nonmonetary incentives, nor did the incentive types affect survey outcome. Also, the survey outcome was independent of the amount of incentive. There seems to be no relation between incentives and data quality.

13.5 Incentive Effects and Survey Modes

Overall, there seems to be little evidence that incentive effects do not carry over between modes of data collection. Meta-analyses of Church (1993) in mail surveys, Singer et al. (1999a) in interviewer-administered surveys (face-to-face and telephone), and Göritz (2006) in Web surveys show similar effects. In all modes prepaid monetary incentives work best. Researchers should be aware; however, that interviewers have the ability to overrule incentive effects by persuading people to participate in interviewer-administered surveys, and that online surveys impose challenges since money cannot be given in the hand.

Ryu et al. (2005), for example, found a monetary (cash) and a nonmonetary (regional park pass) incentive to have similar effects in a face-to-face survey, while the difference in response rate between the two incentive types was significant in a mail survey (67 % for the cash incentive group and 45 % for the pass incentive group). These authors argue that the persuasive abilities of the interviewers were likely to swamp any effect of the incentives in the face-to-face interviews. Thus, interviewers could make some attributes of a survey more salient (leverage-saliency theory), thereby influencing the effects of incentives.

In addition to differences in incentive effects between interviewer and self-administered surveys, there is some evidence that online surveys have some specific issues a researcher should be aware of. Göritz (2006) concludes that incentives seem to be less effective in Web surveys than in other modes of administration. Comparing her results with the results of meta-analyses of Church (1993) and Yu and Cooper (1983) in offline surveys, she concludes that promised nonmonetary incentives seem to work better in online than in offline studies. This could be due to the fact that online surveys are often conducted in panels, where the same people respond to surveys in a regular pace. Panel members have already agreed to take part in surveys in principle and trust between the researcher and respondent is already established. In addition, online panel members often expect to receive an incentive after their participation. Promised incentives may therefore work better in online panel surveys than in other modes of data collection.

Birnholtz et al. (2004) argue that online surveys restrict the range of possible incentives in that they must be in electronic form. Göritz and Neumann (2011) discuss different types of monetary incentives that can be rewarded via the Web: a bank transfer, an intermediate such as PayPal, or loyalty points that can be redeemed against money after a minimum number of points have been collected. The authors argue that money that is paid electronically may not have the same value to respondents as cash because it is more abstract and is not in the hand. Therefore, (prepaid) electronic money is not (yet) comparable to (prepaid) rewards offline (Göritz 2006). In addition, noncash incentives may impose redemption costs on participants:

"direct cost of incentive redemption (e.g., Paypal fees for issuing a check), (b) the cost in terms of time required to set up an account or to type in gift certificate information, (c) the inconvenience of having to redeem an incentive such as a gift certificate with a particular online merchant, and (d) the potential cost difference, for example, a $5 gift certificate is redeemed for something that costs more." (Birnholtz et al. 2004, p. 356). Many vendors of online incentives specify a minimum value, for example, gift certificates often do not sell for less than $5. When postal addresses are known, one could always use the possibility to send a letter and deliver a prepaid incentive with it. Birnholtz et al. (2004) argue that researchers should weigh the additional costs of paper invitations and cash incentives against the use of (higher) electronic incentives. If a $1 bill could yield similar or better response rates than a $5 gift certificate, the additional cost of the incentive would surely outweigh the printing, postage, and handling costs of paper invitations and a smaller value cash incentive.

13.6 Cross Sectional and Longitudinal Designs

Longitudinal surveys are designed to return to survey the same people on several occasions (e.g. waves). Incentives are particularly likely to be used in longitudinal surveys to motivate respondents for repeated cooperation since the usefulness of longitudinal surveys is reduced when respondents refuse to respond, resulting in higher levels of TSE. Researchers have to be extremely careful in designing incentive schemes for longitudinal surveys. Ones the decision has been taken to use a particular incentive structure it might be difficult to alter that structure without having an (adverse) effect on response rates. Thus, the decision to use a particular type of incentive has long-term financial consequences. Some major longitudinal surveys such as the Canadian Survey of Labor and Income Dynamics, the Swiss Household Panel, and the British Birth Cohort Studies have never used incentives (Laurie and Lynn 2008). These panels rely on the altruistic behavior of respondents, their authority, or use other motivational measures (such as media communication or multiple contacts).

The effect of incentives in longitudinal surveys is more complex than in independent cross-sections. Effects on response rates can fade-out, be constant, or cumulative. There are little experiments with incentives in longitudinal surveys. Göritz et al. (2008) experimented with incentives in a three-wave longitudinal experiment. The promise of a payment had a negative effect in the first wave, but a positive effect in the second wave. In the third wave there was no significant difference between the incentive group and the control group. They used an electronic payment (PayPal) as incentive and argue that respondents might have been unfamiliar with the procedure and have been reluctant to answer a survey on which the incentive is difficult to collect. Being invited to the second wave applied only to people who had responded to the first wave, in other words those whom the payment did not prevent them from responding in the first wave. Since there was no effect of incentives in the third wave, one could question the effect of electronic postpaid incentives in longitudinal surveys.

Göritz and Wolff (2007) experimented with lotteries as incentives in longitudinal Web surveys. They found a direct positive effect of the lottery on response only in the first wave but no further direct effects in later waves. The lottery had indirect effects in further waves; however, because responding at a given wave makes it more likely to respond at the next wave, in other words the positive effect of the lottery at the first wave carried over to the subsequent waves. This process can be characterized by a Markov chain. A Markov chain implies that people who take part in a first wave are more likely to cooperate in subsequent waves than are nonrespondents. The authors argue that over time respondents learn that the chance of actually winning a prize is low and readjust their cost-benefit calculation

in accordance with their experience. Therefore, lotteries cannot be recommended in a longitudinal survey.

Zagorsky and Rhoton (2008) experimented with paying incentives to reluctant respondents. An extra incentive of $20 or $40 significantly boosted response rates. The no incentive condition had a response rate of 36 %, versus 43 % for the $20 condition, and 49 % for the $40 condition. Although it is of crucial importance to retain respondents in a longitudinal survey, one could argue whether it is ethical to pay initial nonrespondents. Singer et al. (1999a) offer two views on the issue of offering refusal conversion payments to respondents. "On one side is an economic perspective, according to which a refusal is an ipso facto indication that the survey has less utility to the refusers, or is perceived as more burdensome by them. This being the case, it is seen as entirely appropriate to offer compensation to refusers but not to cooperate respondents, whose utility functions (as evidenced by their behavior) favor participation even without compensation. In contrast to the economic perspective is a social psychological perspective, according to which refusal conversion payments represent a violation of equity expectations. …In addition, such payments are believed likely to alienate cooperative respondents if they become aware of the practice" (Singer et al. 1999a, p. 252).

Question

Zagorsky and Rhoton (2008) experimented with providing incentives to nonrespondents in a longitudinal survey. They use three experimental conditions: (1) no incentive, (2) $20 prepaid incentive, (3) $40 prepaid incentive. Their use of incentives significantly improved response rates. The no incentive condition had a response rate of 36 %, versus 43 % for the $20 condition, and 49 % for the $40 condition. Providing a little over $8,700 of incentives resulted in 75 additional cases, or $116 per case. The initial response was 1,072 cases. Do you think the extra incentive for nonrespondents was worth the costs?

13.7 Incentive Effects for Different Populations

It is conceivable that the effect of incentives differs per country or culture. Incentive effects depend on factors such as mentality, legal situations, fiscal regulations, casual earnings, number of surveys conducted, Internet-penetration rate and literacy (for online surveys), norms, attitudes towards institutions, need for cognition, and need to evaluate etc. However, no empirical evidence exist that incentive effects do not carry over between different cultures.

The empirical results with regard to incentive effects on different subpopulations are somewhat inconclusive. In his meta-analysis on 74 studies, Church (1993) found no significant difference in effect of incentives between surveys of different length, study type, sample composition, or year of publication. According to Church, results can therefore be generalized across different samples and time periods. In addition, other studies that focus on the relation between incentive effects and different subpopulations found no significant effects. For example, Eyerman et al. (2005) found a monetary incentive neither introduced additional differences in cooperation propensities, nor did it eliminate differences among demographic subgroups. Sanchez-Fernandez et al. (2010) found no differences with regard to the effects of pre and postincentives based on draws on response distribution with regard to sex and income. Zagorsky and Rhoton (2008) found no differences in response rates between a no incentive group, $20, and $40 group with regard to age, cohort, race, and education. In addition, Messer and Dillman (2010) found no effect of a $5 cash incentive for gender, age, income, and race.

Messer and Dillman (2010) found a significant incentive effect for education: respondents with higher levels of education had higher completion rates without incentives where respondents with lower levels of education had higher levels of cooperation in the incentive condition. Petrolia and Bhattacharjee (2009) also found that incentives tended to bias the sample

in favor of less educated respondents. In addition, Ryu et al. (2005) found that cash incentives resulted in higher response rates of low educated groups. In addition, they found single persons and nonworking respondents to react stronger on a monetary incentive compared to a nonmonetary incentive. Göritz (2004) found that older respondents were more sensitive to gifts and lotteries than younger respondents. She argues that older respondents might be better of financially and can therefore afford to be more playful. Respondents with higher education were more attracted to loyalty points. Higher educated respondents might be more sensitive to the expected value of the outcome. Men were more sensitive to winning gifts than women. This difference might be explained by the rather technical character of the gifts offered.

The results above are confirmed by Singer et al. (1999b). In their meta-analysis they found some studies in which incentives boosted response percentages for specific groups while other studies showed no significant differences between demographic groups as a result of incentives used. Since people with lower education and singles are often characterized as hard-to-reach groups, one could argue that if any effect could be apparent, the inclusion of incentives reduces nonresponse bias and as a consequence TSE.

There is little evidence that incentive effects vary for different sample types. For example, Porter and Whitcomb (2003) use lottery incentives on a student population and argue that students may be more price sensitive than an average person. Their results are in line with experiments on the general population; however, and show that students do not show higher response rates when a lottery incentive is offered. Another difference that can be made between different sample types is that of consumer and organization samples. Saunders et al. (2006) found no differences in incentive effects between consumer and organizational populations. With the rise of Internet surveys and the professional panels associated with it, some researchers claim that commercial online panels use professional respondents who are in it for the money. Online panels are often not drawn from a probability sample but rely on self-selection samples. Göritz (2004) found no evidence; however, that self- and non-self-selected panel members differ in their susceptibility to different incentive conditions (bonus points, money, and gift lotteries). There is no reason to assume that incentive effects differ between commercial and noncommercial panels.

Case: Leisure Panel

In the Netherlands, three universities started to set up a research panel, which is called the Leisure Panel (www.leisurepanel.nl). Students (4,500 in total with a national spread, note that the Internet penetration rate in the Netherlands is more than 90 %) were used to recruit panel members. Each student was send an e-mail in which they were asked to forward the mail to their family members, (online) friends, neighbors etc., and ask them to become a member of the panel. In an incentive experiment, 1,000 students were randomly divided into three groups: one group received an announcement letter by mail with a €5 cash incentive, one group received an announcement letter with a lottery ticket for the Dutch National lottery (costs were €3 per ticket), and one group received an announcement letter without incentive.

The incentive significantly boosted response rates: 7 % response versus 3 % in the no incentive condition (Chi Square = 7.59, $p < 01$, $N = 1{,}000$). There was no significant difference in effect between the €5 cash incentive and the lottery incentive group (both had a response rate of 7 %). Of course, the lottery incentive was less expensive than the cash incentive (€1,000 vs. €1,665, respectively).

What do you suggest the researchers should do in the next wave of recruitment?

A. Send no incentives, since response rates are very low (even with incentives) ,and administration costs without incentives are negligible).

B. Send the lottery ticket as an incentive since it boosted response rates from 3 % to 7 % and worked evenly well as the €5 cash incentive which is more expensive.
C. Divide the students into two groups and send the lottery ticket to one group and the €5 cash incentive to the other, since both perform equally well in the first wave of recruitment and more evidence is needed before making a final decision on incentives.
D. Other, namely….

13.8 Discussion

Survey researchers must decide whether the cost of providing incentives to persuade potential respondents to participate in a survey and the relative increase in response rates is worth the investment. Literature shows that prepaid monetary incentives work in increasing response rates, potentially attracting hard-to-reach-groups, without a loss of data quality. Results carry over to different modes of data collection (paper, Web, face-to-face, telephone) and different populations (general population, student samples, organizational surveys) and types of panels (commercial versus noncommercial). Nonmonetary and postpaid incentives do not seem to work very well. Researchers should be particularly careful when wanting to use one of these types of incentives. They are probably not worth the energy involved.

This chapter provides a formula that researchers can use to calculate the optimum amount of prepaid incentive. When considering the effectiveness of incentives, response rates and costs per completion have to be taken into account. For example, if a representative sample is drawn and contacted with great effort, maximizing response rates with incentives will probably override considerations for saving some money for incentives. The optimum amount of incentives is often more dependent on (variable) survey costs than it is on expected response rates. This is a good thing, since researchers are often better able to predict survey costs than future response rates.

Although many studies prove that incentives can significantly improve response rates, thereby reducing TSE, researchers have other survey design features at their disposal that might cost less, and be evenly effective in increasing response rates and reducing TSE. Understanding the decision to participate in a survey is the key to successful surveying. According to Groves et al. (1992), researchers should consider the following (in addition to cost-benefit analyses):

1. respondents' moods
2. respondents' feeling of obligation
3. respondents' liking towards the survey/interviewer/sponsor
4. characteristics of the sample persons
5. helping tendencies
6. perceptions if survey participation is normative
7. scarcity of opportunities to fill out similar surveys
8. whether participation in a survey is consistent with previous behavior
9. whether participation in surveys is consistent with behavior of others
10. opinion change
11. maintaining interaction between survey office and respondent

Larson and Chow (2003) conclude that follow-up mailings are preferred over monetary incentives given a limited budget for survey administration. This is also found by James and Bolstein (1992) who demonstrate that a personalized survey with four mailings results in a response rate as high as a one-wave mailing with an enclosed incentive of any amount (up to $50). On the other hand, if there is limited time for survey administration, monetary incentives may be preferred over follow-up mailings. Dillman (1978) already proposed in the late 1970s in his "Tailored Design Method" that follow-up mailings or multiple contacts are of crucial importance when it comes to responses to surveys. His method is still widely applied today.

Multiple contacts may therefore outweigh the effects and costs of incentives. This could be a positive factor for survey budgets.

Researchers should acknowledge that most potential respondents do not systematically review all pro's and con's of complying to a survey request, but follow one or two highly prominent heuristics on which they base their decision to cooperate. Abiding to the previous defined heuristics might be more cost-effective than implying large incentive schemes.

13.9 Glance at the Future of Incentives

Evidence suggests that surveys face decreasing response rates (Curtin et al. 2005). This can be explained by higher refusal rates and an increase in number of contact attempts, which in turn may be due to oversurveying and call or spam blocking technologies. Kroph and Bair (2005) notice a parallel between declining cooperation rates for surveys and a decline in civic participation (see also the work of Robert Putnam 1995, 2000) that has been seen in the United States and similar western cultures over the past 30 years. The authors find evidence that norms of cooperation matter in the decision to participate in a survey. With the overall decline in response rates, incentives as external motivator might become a more powerful tool in increasing response rates. When respondents become less intrinsically motivated, extrinsic motivation becomes more salient.

As a result of the rise of the Internet, Web surveys have become available to the masses. Everyone can conduct a survey nowadays; it takes little more than an e-mail file and a survey software program to conduct an online survey. With the rise of Web surveys, two things are likely to appear when it comes to incentives. First, Web surveys are often conducted in panels (repeatedly asking the same respondents to answer survey questions). This has associated consequences for the use of incentives, as we have discussed in the previous paragraphs. Second, Web surveys make it pretty difficult to provide respondents with prepaid cash incentives. Other ways of providing respondents with prepaid incentives, for example, gift vouchers or PayPal money transfers, may become a more dominant tool for sending incentives although they have less power than money in the hand.

Questions

1. Compare and contrast existing theories on incentive effects. What does each theory predict about different types of incentives?
2. Define different types of incentives and discuss which type is generally most effective? Can you come up with an example for which another type of incentive might be more effective to use?
3. What determines the optimal amount of incentives?
4. Do incentives affect data quality?
5. What is the difference in incentive effects between cross-sectional and longitudinal designs?
6. Are there differences in incentive effects between different survey modes, e.g. face-to-face, telephone, mail or web?
7. Can you expect different incentive effects for subpopulations, e.g. differences when it comes to age, gender, education, and race?
8. Are incentive effects likely to differ between different sample types, for example student population versus general population or consumer population versus organization population?
9. Do trained respondents react differently to incentives than untrained respondents?
10. What do you think the future will hold for incentives and related effects?

References

Biemer, P. (2010). Total survey error design, implementation, and evaluation. *Public Opinion Quarterly, 74*, 817–848.

Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychological and Marketing, 11*, 483–492.

Birnholtz, J. P., Horn, D. B., Finholt, T. A., & Bae, S. J. (2004). The effects of cash, electronic, and paper gift

certificates as respondent incentives for a web-based survey of technologically sophisticated respondents. *Social Science Computer Review, 22*, 355–362.

Brehm, S. S., & Brehm, J. W. (1981). *A theory of psychological reactance*. New York: Academic Press.

Budowski, M., & Scherpenzeel, A. (2004). *Encouraging and maintaining participation in household surveys: the case of the swiss household panel, working paper, 1_04*. Neuchatel: Swiss Household Panel.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly, 57*, 62–79.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69*, 87–98.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Dillman, D. A. (2007). *Mail and internet surveys. The Tailored design method*. New York: Wiley.

Eyerman, J., Bowman, K., Butler, D., & Wright, D. (2005). The differential impact of incentives on refusals: Results from the 2001 national household survey on drug abuse incentive experiment. *Journal of Economic and Social Measurement, 30*, 157–169.

Göritz, A. S. (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research, 46*, 327–345.

Göritz, A. S. (2005). Contingent versus unconditional incentives in www-studies. *Metodolosky Zvezki, 2*, 1–14.

Göritz, A. S. (2006). Incentives in web surveys: Methodological issues and a review. *International Journal of Internet Science, 1*, 58–70.

Göritz, A. S. & Neumann, B. P. (2011). *The longitudinal effects of incentives on response quantity in online panels*. Retrieved from www.goeritz.net

Göritz, A. S., & Wolff, H.-G. (2007). Lotteries as incentives in longitudinal web surveys. *Social Science Computer Review, 25*, 99–110.

Göritz, A. S., Wolff, H.-G., & Goldstein, D. G. (2008). Individual payments as a longer-term incentive in online panels. *Behavior Research Methods, 40*, 1144–1149.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475–495.

Groves, R. M., Singer, E., & Corning A. D. (2000). A leverage-saliency theory of survey participation: description and illustration., *Public Opinion Quarterly, 64*, 299–308.

Harris, I. A., Khoo, O. K., Young, J. M., Solomon, M. J., & Rae, H. (2008). Lottery incentives did not improve response rate to a mailed survey: a randomized controlled trial. *Journal of Clinical Epidemiology, 61*, 609–610.

Hubbard, R., & Little, E. L. (1998). Promised contributions to charity and mail survey responses: replication with extension. *Public Opinion Quarterly, 52*, 223–230.

James, J., & Bolstein, R. (1992). Large monetary incentives and their effect on mail survey response rates. *Public Opinion Quarterly, 56*, 442–453.

Jobber, D., & Saunders, J. (1988). Modeling the effects of prepaid monetary incentives on mail survey response. *Journal of Operation Research Society, 39*, 365–372.

Jobber, D., Saunders, J., & Mitchel, J. W. (2004). Prepaid monetary incentive effects on mail survey response. *Journal of Business Research, 57*, 21–25.

Kroph, M. E., & Blair, J. (2005). Eliciting survey cooperation: incentives self-interest, and norms of cooperation. *Evaluation Review, 29*, 559–575.

Larson, P. D., & Chow, G. (2003). Total cost/response rate trade-offs in mail survey research: impact of follow-up mailings and monetary incentives. *Industrial Marketing Management, 32*, 533–537.

Laurie, H., & Lynn, P. (2008). The use of respondent incentives on longitudinal surveys. *Institute for Social and Economic Research*, No: 2008-42.

Messer, B., & Dillman, D. A. (2010). Using address based samping to survey the general public by mail vs. 'Web Plus Mail', Technical Report 10–13, Pullman, WA:Social & Economic Sciences Research Center.

Petrolia, D. R., & Bhattacharjee, S. (2009). Revisiting incentive effects. evidence from a random-sample mail survey on cnsumer preferences for fuel ethanol. *Public Opinion Quarterly, 73*, 537–550.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*, 389–407.

Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in america. *Political Science & Politics, 28*, 664–683.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of american community*. New York: Simon & Schuster.

Medway, R. Tourangeau, R., Viera, L. & Marsh, S. (2011). *Exploring the impact of prepaid cash incentives on multiple indicators of data quality*. Paper presented at Phoenix, AZ:AAPOR Conference.

Ryu, E., Couper, M. P., & Marans, R. W. (2005). Survey incentives: cash vs. in-kind; face-to-face vs. mail; response rate vs. nonresponse error. *International Journal of Public Opinion Research, 18*, 89–106.

Sanchez-Fernandez, J., Munoz-eiva, F., Montoro-Rios, F. J., & Ibanez-Zapata, J. A. (2010). An analysis of the effect of pre-incentives and post-incentives based on draws on response to web surveys. *Quality & Quantity, 44*, 357–373.

Saunders, J., Jobber, D., & Mitchell, V. (2006). The optimum prepaid monetary incentives for mail surveys. *The Journal of the Operational Research Society, 57*, 1224–1230.

Singer, E., & Bossarte, R. M. (2006). Incentives for survey participation. when are they "Coercive"? *American Journal of Preventive Medicine, 31*, 411–418.

Singer, E., Groves, R. M., & Corning, A. D. (1999a). Differential incentives: beliefs about practices,

perceptions of equity, and effects on survey participation. *Public Opinion Quarterly, 63*, 251–271.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999b). The effect of incentives in interviewer-mediated surveys. *Journal of Official Statistics, 15*, 217–230.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64*, 171–188.

Trussel, N., & Lavrakas, P. J. (2004). The influence of incremental increases in token cash incentives on mail survey response: is there an optimal amount? *Public Opinion Quarterly, 68*, 349–367.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, no; lotteries, no; cash, yes. main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, *60*, 542–562.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20*, 36–44.

Zakorsky, J. L., & Rhoton, P. (2008). The effects of promised monetary incentives on attrition in a long-term panel survey. *Public Opinion Quarterly, 72*, 502–513.

Further Readings

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly, 57*, 62–79.

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

Göritz, A. S. (2006). Incentives in Web Surveys: Methodological Issues and a Review. *International Journal of Internet Science, 1*, 58–70.

Göritz, A. S. & Neumann, B.P. (2011). "The Longitudinal Effects of Incentives on Response Quantity in Online Panels", retrieved from www.goeritz.net.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly, 56*, 475–495.

Groves, R. M., Singer, E., & Corning A. D. (2000). "A Leverage-Saliency Theory of Survey Participation: Description and Illustration", *Public Opinion Quarterly*, *64*, 299–308.

Ryu, E., Couper, M. P., & Marans, R. W. (2005). Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. *International Journal of Public Opinion Research, 18*, 89–106.

Saunders, J., Jobber, D., & Mitchell, V. (2006). The Optimum Prepaid Monetary Incentives for Mail Surveys. *The Journal of the Operational Research Society, 57*, 1224–1230.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The Effect of Incentives in Interviewer-Mediated Surveys. *Journal of Official Statistics, 15*, 217–230.

Singer, E., Groves, R. M., & Corning, A. D. (1999). Differential Incentives: Beliefs about Practices, Perceptions of Equity, and Effects on Survey Participation. *Public Opinion Quarterly, 63*, 251–271.

Part IV

Survey Designs, Modes and Applications

Designing the Face-to-Face Survey 14

W. Lawrence Neuman

14.1 Survey Types and Strengths/ Limitation of Face-to-Face Interview Survey

A. Survey types and the place of the face-to-face survey

Of the several ways to conduct survey research, the face-to-face interview is the oldest, followed by the self-administered paper questionnaire, with the mail questionnaire close behind. Survey researchers increasingly rely on modes of administration that employ new technologies, such as telephones, computer-assisted interviewing, and Internet-web methods. With the move to survey modes with new technology, Why should we continue to use the face-to-face interview rather than other modes of administration? The face-to-face interview has its limitations but under certain conditions, it has significant advantages over non-interview surveys and over interviews not conducted in person.

B. Face-to-face strengths

Direct human contact is the primary strength of the face-to-face interview. The researcher, or a trained interviewer, personally meets, sees, talks with, and listens to a respondent. Direct human interaction increases respondent cooperation and motivation. Face-to-face surveys have the highest response rates. Respondents find it more difficult to turn down a person standing at the doorway than to toss a mail questionnaire into the trash, hang up the phone, or click to a different computer screen. Krysan et al. (1994) found different types of people may respond or fail to respond to different modes of survey administration (mail versus face-to-face). Lower income, less educated, and minority populations are less likely to participate in non face-to-face modes.

Once an interviewer has gained entry and initial cooperation, mainstream norms of politeness and hospitality encourage continued interaction. The face-to-face format permits the longest survey interviews. A skilled interviewer can stimulate and involve a respondent, build a sense of trust, and create rapport factors that encourage respondents to provide honest answers and disclose information. Respondents who might otherwise be difficult to engage fully in a survey interview can become absorbed, cooperative participants with face-to-face interaction. Not surprisingly, respondents report higher satisfaction with the survey process when interviewed face-to-face rather than by telephone or Web surveys (Holbrook et al. 2003).

Human contact with an interviewer, usually in a respondent's home environment, also provides opportunities for interviewer observation. In other survey modes, a respondent could lie or be deceptive about obvious, visible facts without

W. L. Neuman (✉)
Department of Sociology,
Anthropology and Criminal Justice ,
Asian Studies, University of Wisconsin-Whitewater,
Whitewater, WI, USA
e-mail: neumanl@uww.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_14, © Springer Science+Business Media New York 2012

detection. In the face-to-face interview, an interviewer can observe the respondent's physical characteristics (e.g., skin tone, apparent age, outward gender expression), be attentive to nonverbal clues, and note surroundings. Such details can be significant contextual data unavailable via a telephone interview or Web survey. Was the respondent located in a dingy apartment located on street filled with abandoned cars and boarded up buildings as gun shots are heard in the background or in a spacious, plush mansion with servants located on a huge manicured estate surrounded by a wall?

Computer-assisted Web surveys can show respondents photos, charts, and so forth, but the face-to-face interview offers the greatest flexibility to show charts, photos, and such. If a survey asks a respondent to react to a complex array of visual images or physical objects (e.g., which of these two hand-held remote controls is better for you?), the face-to-face interview is the most effective mode of administration. Dillman and Christian (2005) warn that survey mode differences (e.g., among mail, face-to-face interview, phone interview) can produce substantial differences in respondent answers, especially when using visual presentations. They found large differences by presentation mode when showing different five-point satisfaction scales to respondents.

The face-to-face survey also facilitates interviewers using *probes* to inquire about ambiguous or unclear respondent answers, offering clarifying explanations of complex survey questions, or providing gentle assistance (e.g., helping an elderly person sit down). Sometimes eye contact, a shift in tone of voice, or a smile can ease tension, facilitate understanding, and move the interview process forward. Such human contact extends beyond greater respondent cooperation to improve measurement validity. The respondent who answers with a quizzical facial expression or appears hesitant or unclear about how to interpret a survey question may be able to correctly interpret the survey question as a researcher intended and provided a recognizable, accurate answer if a warm and compassionate interviewer offers assistance.

C. Face-to-face limitations

High costs are a major limitation of the face-to-face interview, because it can cost five or ten times more than alternative modes and take much longer to complete. We face a tradeoff between low-cost, fast modes of survey administration that yield acceptable response rates and data quality, and the high-cost, slow face-to-face survey interview that provides the highest response rates and the richest, most complete data.

Another limitation involves the interviewer. As Tourangeau (2004, p. 790) remarked, "… the big drawback to face-to-face interviews is the interviewer… Interviewers can be both a blessing and a curse." More than other modes, the face-to-face interview has the greatest potential for "interviewer effects" that become "interview bias." Compared to other modes, the face-to-face survey also increases a respondent's tendency to give "socially desirable" i.e., expected, normative answers, and to hide embarrassing personal details (see Holbrook et al. 2003 for contrary evidence). Potential interviewer effects require longer and more careful interviewer training than for other survey interviewing. Again, we face a tradeoff. Direct interaction with another person may best engage, motivate, and assist respondents, and it also carries the greatest potential to influence respondent answers and behaviors.

Survey administration is more complex with a face-to-face survey than with the alternatives. Administrative complexity, in addition to longer interviewer training and involvement, adds to high costs. The face-to-face survey often requires more attempts to contact a respondent prior to the interview because it requires gaining physical access to a respondent. The costs in time, money, and interviewer effort for multiple contacts add up. After locating a respondent, additional costs include transporting, housing, feeding, and preparing the interviewer. A face-to-face interview also introduces potential safety or security concerns avoided by other modes of administration. There are also administrative "overhead" costs to supervise and monitor sending interviewers to diverse geographic

locations and oversee transportation and budget details.

14.2 Models of Face-to-Face Survey Process

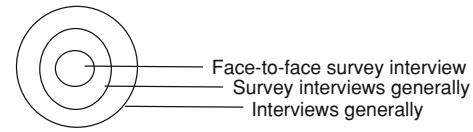


Fig. 14.1 The context of face-to-face survey interviews

A. The process of a face-to-face survey

The face-to-face survey is both a mode of survey research and a type of interview. From both perspectives, it is a multi-phase process that centers on a short-term social interaction between two strangers, an interviewer and respondent. However, the total process includes other people and phases prior to or after the interview encounter. Most prior phases parallel to other modes of survey administration, such as research question clarification, survey question writing, questionnaire layout and design, questionnaire pilot testing, sampling, and respondent contact. For example, question design will vary depending on whether a survey is Web-based or uses face-to-face interviews. Likewise, interviewer training will differ depending on whether face-to-face versus telephone interviewing is used. Such differences are minor, and the major difference involves what takes place during an interview. Next, we consider the interview phase.

1. The Interview as a type of social interaction

An interview is a special type of social interaction. As Schaeffer and Presser (2003, p. 65) remarked, "… the standardized survey interview is a distinct genre of interaction with unique rules, but it shares many features with ordinary interaction because social and conversational norms as well as processes of comprehension, memory, and the like are imported into the interview from the situations in which they are learned and practiced." Implicit in best practices for conducting a face-to-face survey is some model of the interview as a special social encounter with a specific purpose and structure.

All interviews (e.g., clinical or medical, therapeutic, police interrogation or investigative, entertainment, job application, journalistic) are distinct social interactions and speech events. They obey different rules from ordinary social encounters and conversations. In an ordinary conversation, each participant equally asks or answers questions and either participant can introduce new topics, change direction, or digress. In interviews, one participant, the interviewer, asks questions, controls new topic introduction, monitors the pace and direction of the interchange, and restricts digressions. The other participant only provides answers.

A survey research interview differs from other interviews. Kahn and Cannell (1957, p. 16) defined the survey interview as, "a specialized pattern of verbal interaction—initiated for a specific purpose, and focused on some specific content area, with consequent elimination of extraneous material… the role relationship of the interviewer and respondent is highly specialized…" Interviews in general, the survey interview as a type of interview, and the face-to-face survey interview follow increasingly distinct rules, operating within one another like a set of concentric circles (see Fig. 14.1).

To begin, we examine how the interview differs from non-interview social interactions or conversations. Next, we consider how the survey-research interview differs from other interviews. Last, we focus on concerns unique to the face-to-face compared to other survey interviews.

The interview has the following six distinctive characteristics:

- *Recognized institutionalized roles* Each party to the interview is assigned an institutionalized identity and role. One is the interviewer, with an active authoritative questioner role. The other is the respondent, with a passive answerer role. Although other people may be involved (e.g., supporter, observer, assistant, ally), these are the two principal roles of the interview encounter.
- *A task-governed, controlled agenda* The encounter follows a formal or informal script

or sequence with a "task" to accomplish. The agenda may be prefixed or flexible and interactive, but the occupant of the interviewer role constrains or controls the encounter, directs flow of events (e.g., when it starts or ends, when it changes direction), and moves the encounter from a beginning toward an ending or conclusion phase.

- *A restricted range of topics* The encounter has a bounded range of topics. The type of interview and its purpose will constrain the topics expected or permitted, and define "off limits" or forbidden topics. Each party to the encounter has some influence over the appropriate topic range, but the interviewer has far greater control.
- *An appropriate set of behaviors* An interview situation has implicit rules of appropriate behavior. The rules include the acceptability or non-acceptability of physical touching, making eye contact, physical movements, physical distance between persons, loudness of speech, use of obscenities, ability to take notes or make recordings, and so forth.
- *A question–answer interactional sequence* Central to the interview interaction is a turn-taking sequence in which one role occupant asks questions and the other provides answers. The exact form varies, but there is a fundamental imbalance in verbal behavior between the occupants of the two major roles.
- *A bounded encounter* The interview encounter can vary in length. It may be a one-time event or a set of repeated events. However, it is distinct, separate interaction and not part of an on-going, open-ended social relationship. It usually involves strangers or non-intimates and is bounded by a specific purpose and limited duration.

Each participant brings to the interview a set of expectations and assumptions about the appropriate behaviors, topics, and ways of conversing. Capable adults have learned about the interview interaction informally by observing others, viewing media presentations, or hearing casual explanations from others about what to expect. Most occupants of the dominant (interviewer) role received some formal instruction, while occupants of the subordinate (respondent) role only learn about the interview and its expectations of them informally. They rarely receive a formal script or written instructions about how they are to behave.

Most capable adults recognize the variety of interviews that take place in a modern, bureaucratic society and are aware that the purpose, degree of trust, formality, and so forth vary by type of interview. For example, asking about intimate details of one's body or personal hygiene behaviors may be appropriate for a medical interview but not in a job interview. A person may be cautious about revealing potentially incriminating information in an interrogation interview by a police detective more than in a confidential therapeutic interview with a psychiatrist or priest/minister/rabbi. Although their interview knowledge may be incomplete or partly accurate and acquired informally without uniform experiences, most adults know basic interview rules and expectations as a specific type of social interaction and distinguish among interview types.

2. Survey research interview

A survey research interview differs from other types in purpose, respondent and interviewer roles, and standardization.

a. Purpose Other interviews screen or rate potential applicants, identify personal concerns to provide tailored assistance, or investigate suspicion of criminal behavior. The purpose of a survey interview is to gather information on a respondent (or a respondent's situation). The requested information can be broad–life history, family background, characteristics such as schooling or income, personal beliefs and attitudes, dreams and aspirations, social relationships, intimate behaviors, and daily habits. Survey interviews solicit information to acquire scientifically relevant knowledge, and in the process, a respondent becomes only a data source for researcher-defined purposes. Respondents are rarely aware of the researcher's interest, needs, and purposes even when deception is not part of a research design. Survey researchers treat respondents as contributors to a larger pool of data. They detach and abstract the data from the individual respondent, stripping

away a web of intimate contextual details that comprise a person's biographical life. They then merge the abstracted data with data taken from other respondents. Next, they apply statistical or other data analysis without further respondent input or participation. The respondent loses ownership and is barred from altering or viewing the raw data he or she contributed.

For example, a survey researcher asks about schooling in an interview to acquire data for a theoretically defined variable, educational attainment. The researcher ignores biographic details of a particular respondent's schooling (e.g., particular courses taken, grades received, classroom experiences, co-curricular activities, and so forth), and removes isolated information about a person (e.g., years of schooling completed). Stripped of the embedded biographic context and removed from the interactional encounter of the interview, survey researchers merge abstracted individual data with that taken from other respondents. They then manipulate the merged data to test hypotheses, construct models of human behavior, advance theoretical knowledge, and so forth. All this occurs outside the direct awareness, control or involvement of the respondent who supplied the original information.

b. Roles The structure and norms of the survey interview encounter proscribe respondent and interviewer roles and assign behavioral expectations to each. Interviewer role expectations include being warm, polite, and considerate, and to display modest affective expressions that are simultaneously "professional" i.e., neutral and detached. The role's professional, detached data-gatherer dimension requires role occupants to be careful, deliberate, focused, and task-oriented. An interviewer maintains continuous control over the interaction, including its content, pace, and direction. An interviewer guides, encourages, and cajoles the respondent into compliance with the information extraction task, simultaneously disciplining, instructing, and motivating the respondent.

In addition to the respondent and interviewer the interview involves a third party, the survey researcher who designs the study, provides questions, and orchestrates the process (see Zouwen et al. 2010, p. 72). An interviewer may control what occurs in an interview interaction, but the unseen, non-present researcher is directing the interviewer's actions. The interviewer only represents or is a surrogate for the absent survey researcher. The researcher provides an interviewer with employment, selects which respondents to interview, stipulates interview behavior, supplies the questions to ask, and instructs how answers are to be recorded. Like a hidden puppeteer, the researcher pulls an interviewer-accomplice's strings to extract desired information from respondents.

The proscribed survey interviewer role is to be a detached professional, diligent and meticulous employee, and warm friendly motivator. The respondent role is to be a cooperative, candid, and honest information source. The ideal respondent is compliant, answers every question honestly, and discloses personal information accurately and fully. He or she is interested, attentive and informed, and correctly interprets every survey question exactly as intended by the researcher through the intermediary of an interviewer. The respondent expects nothing in return for supplying survey researchers with data.

3. The standardized format

To understand the practice of survey interviewing, we first should step back and recognize that survey research is part of a larger historic enterprise or "project" of the modern era. The project is to gather objectified data about people, including their characteristics, behaviors, feelings and attitudes for use by experts and bureaucratic officials who acquire knowledge on, advance and test theories about, and make decisions regarding social-political issues (see Alonso and Starr 1987; Gubrium and Holstein 2002). Persons with official or expert authority use the extracted abstract information for their own purposes. Gubrium and Holstein (2002, p. 13) called this the "vessel-of-answers model" of the respondent. In it, the respondent is a passive subject who supplies detailed intimate information that is subsequently removed from her/his control. Few active social researchers debate such meta-methodological

issues (for exceptions see Barnes 1979; Burawoy 2005; Clawson et al. 2007).

Over time, survey researchers shifted survey practice to be more rule-based, explicit, and standardized. As they sought to create a more positivist "scientific" and objective method for collecting data, they increasingly standardized procedures. This produced more objectified data that better fit the paradigmatic assumptions of positivist social science. We can view the standardized format survey as part of a larger historical movement to produce objective, quantifiable knowledge about the social-political world, or what Porter (1995) astutely described as "mechanical objectivity."

Mechanical objectivity makes us follow explicit data collection rules that are neutral and fair. Its purpose is to remove what may be considered subjective personal moral judgments about individuals and yield objective, quantifiable information consistent with positivist scientific or bureaucratic purposes (see Schaeffer and Maynard 2002).

a. Why standardize?

Early survey interviews were ad hoc and idiosyncratic, varying greatly by interviewer. Gradually, survey researchers learned that different interviewers could substantially influence the survey data. Suchman and Jordan (1990, p. 233) observed, "… the constraints on the interview are all imposed in the interest of standardization. Standardization is what identifies the interview process as a scientific procedure." The reason for standardizing is to control interviewer effects that can arise in two ways.

A first way arises from behavioral variation across specific interviewers. This can introduce unwanted variation into the data. Positivist-oriented survey researchers are only interested in data variation caused by different respondent answers, and seek data variation that reflects actual variation among respondent attitudes, behaviors, or characteristics. They treat other sources of data variation, such as that from specific interviewers, as sources of measurement error.

A second way interviewers can influence data arises from an interviewer's personal opinions, attitudes, and expectations about a respondent. "Standardization addresses both components of interview error. The rules of standardization attempt to hold the behavior of the interviewer constant, thereby reducing variable error… Standardization also reduces bias by reducing the number of opportunities for interviewer expectations or opinions to intrude on the process by which the respondent's answer is generated" (Maynard and Schaeffer 2002, pp. 5–6).

Thus, by standardizing the survey researchers try to remove interviewer variation and make the specific interviewer irrelevant to data collection. They assume that such variation represents unwanted "noise" or "error." Standardizing interviewer behavior is just part of standardizing the entire survey process, including questions and respondent actions. In a standardized interview, an interviewer does the following (Schaeffer and Maynard 2002, p. 580): read each question neutrally and exactly as written by the researcher, follow question sequences or branches in question order as directed without deviation, use only acceptable non-directive probes for inadequate respondent answers, record respondent answers without interviewer discretion, and be non-judgmental toward all respondent answers. This supposedly eliminates or reduces interviewer influence and pushes respondent replies into a predetermined range for easy data analysis.

In the standardized interview format, interviewers repeat each question word-by-word without variation in a specified sequence while respondents select from preset choices, unless an open-ended option is offered. It rests on the assumption that variations in question wording, question sequencing, interviewer behavior, or non-conforming respondent answers are potential sources of measurement error, errors that could invalidate the survey data. The underlying goal is to extract data in an uncontaminated antiseptic manner from a conforming, complacent respondent, much as one might extract a DNA or blood sample from an inert patient for subsequent analysis.

A model of how to conduct face-to-face survey interviews rests on assumptions from one of two alternative survey research paradigms. The two paradigms come from different epistemological

assumptions (positivist and interpretative) (see Della Porta and Keating 2008; Neuman 2011, pp. 90–119; Steinmerz 2005). We next examine the two survey paradigms.

b. The stimulus-response, information retrieval paradigm

The dominant, conventional paradigm relies on paired stimulus (question from interviewer) and response (answer from respondent) model. As Mishler (1986, pp. 14–15) observed, "… by and large, research on problems of the interview has been framed with the stimulus–response paradigm… The intent… is to find ways to standardize the stimulus, or perhaps a better term, to neutralize it. That is, the aim is to ascertain respondent's 'true' opinions and minimize possible distortions and biases in responses that may result from question or interviewer variables that interfere with the respondents' ability or wishes to express their 'real' or 'true' views." The stimulus–response, information retrieval (SRIR) paradigm uses a naïve assumption model of the interview (Foddy 1993, p. 13). Its assumptions are naïve due to a simplistic and non-problematic view of the survey interview and include the following:

(1) Researchers have clearly conceptualized all variables into specific survey questions;
(2) Effects due to wording bias or question order have been removed from survey questions;
(3) The process of being interviewed and the specific interviewer and interview interaction have no effect on the respondent or his/her answers, and
(4) All respondents (a) are highly motivated and willing to answer any question asked, (b) possess complete information and can accurately answer all questions asked, (c) understand the questions exactly as they are asked, (d) give truthful answers when they do not know the researcher's hypotheses and have no hints or suggestions from interviewers, and (e) give answers in an interview that perfectly match their true attitudes, behaviors, and beliefs.

Although some survey researchers have questioned the degree to which specific assumptions hold, mainstream survey research proceeds as if they hold true most, if not all the time. To examine seriously such assumptions might disrupt the enterprise of survey research and raise questions about the validity of the data routinely used by social scientists and bureaucratic organizations. The primary advance in the SRIR paradigm has been to move beyond a simple behaviorist "black box" view of respondent mental processes and into cognitive information processing models (Sudman et al. 1996). In recent decades, survey researchers have explored the mental processes a respondent applies when retrieving information relevant to the survey question (stimulus) and providing a response (see Tourangeau et al. 2000).

c. An constructionist, collaborative paradigm

An alternative paradigm views the survey interview less as an individualized stimulus–response, information-retrieval process than as a special kind of social encounter, one that obeys some but not all the rules of a conversation. Schegloff (1990, p. 248) termed this "talk-in-interaction." In contrast to the SRIR paradigm, the alternative constructionist, collaborative (CC) paradigm holds that, "an interview no matter how standardized, remains fundamentally a linguistic and interactional event" (Suchman and Jordan 1990, p. 240). The CC paradigm treats the interview primarily as an interactional encounter between two persons who collaborate to construct meaning through a fluid, dynamic exchange. In this exchange, "interviewers, along with respondents, deploy resources of an interactional substrate-basics skills for engaging meaningfully-in pursuit of answers. These skills are presumed by and yet largely ignored in the rules of standardized interviewing" (Maynard and Schaeffer 2002, p. 9). To accomplish a meaningful interchange, respondent and interviewer work together, deploying tacit practices to handle unforeseen, unpredictable events (see Gathman et al. 2008; Lavin and Maynard 2001).

In the SRIR paradigm, the survey interview is understood as an individual cognitive retrieval process. By contrast, the CC paradigm views it as social interaction in which two people mutually generate meaning in a constrained social situation through on-going conversational dynamics.

In the CC paradigm, interview participants draw on an array of social skills and resources from ordinary life and daily conversations to generate situation-specific meanings. Like other human interchanges, the interview is a fluid encounter filled with opportunities for interactional slippage and misunderstanding. Luckily, people posses a variety of skills and interactional resources that they can draw upon to repair, correct, and realign communication flows and thereby construct mutual understanding. Schober and Conrad (2002, p. 75) contrasted the SRIR paradigm, or message model, with the CC paradigm. In the CC paradigm, "conversational partners provide each other with evidence of whether the conversation is on track—whether they are understanding one another well enough for current purposes… the theory of standardization assumes a view of communication, the message model, that does not hold in spontaneous conversation… "

In the CC paradigm, an interview extends processes already present in ordinary conversation rather than blocking them as occurs in the standardized survey format. The CC paradigm supports use of a "conversational" or "flexible" format interview (see Tourangeau et al. 2000, pp. 56–59). Taking the CC paradigm seriously would force a major re-examination of the standardized format along with its goal of eliminating from the interviewer-respondent interaction what it considers to be "measurement error."

Paradoxically, the interactional skills and interpersonal resources that the CC paradigm treats as essential for achieving a meaningful survey interview are what the SRIR paradigm seeks to control, restrict, or eliminate through standardization. "Compared with ordinary conversation, the survey interview suppresses those interactional resources that routinely mediate uncertainty of relevance and interpretation" (Suchman and Jordan 1992, p. 242). The two paradigms suggest different interviewer behaviors. The chart below compares the paradigms on three features, proscribed interviewer behavior, assumptions about the questioning process, and assumptions about respondent answering.

Feature	Paradigm of survey interview	
	Stimulus–response, information retrieval	Constructivist, collaborative
Proscribed interviewer behavior	Behavior is uniform and standardized with exact wording. Interviewers are to follow a script closely and use minimal, preset probes.	Interaction is conversational and flexible with active rewording. Interviewers seek to convey authentic intention and meanings from the researcher and use extended, tailored probes.
Assumption of questioning process in the interview	A proper stimulus will elicit an appropriate mental task and verbal response by the respondent. The researcher creates the stimulus and the interviewer is only a neutral conveyer of it.	The interviewer and respondent engage in a mutual collaboration relying on verbal/non-verbal interaction skills and resources. Together they try to create a shared understanding of the researcher's intended meaning and appropriate responses to it.
Assumption of answering process in the interview	Respondent interprets a standard question as worded with restricted	Respondent constructs, jointly with the interviewer, a meaningful answer that both

(continued)

	interviewer behavior then internally formulates an appropriate response (more than once if necessary) then delivers it for acceptance.	addresses the requirements of the survey question and constraints of the interview situation.

14.3 Survey Questions Specific to Face-to-Face Surveys

A. General Survey Question Writing Principles

Dozens books describe how to write survey questionnaires (see Babbie 1990; Bradburn and Sudman 1988; Dillman 2000; Czaja and Blair 2005; Fowler 2009; Oishi 2002; Rossi et al. 1983; Sudman and Bradburn 1983 as examples). The following section will review a few basics and focus on issues relevant to the face-to-face survey interview. Writing good survey questions takes practice and is a mixture of art and science, as discussed by Gideon earlier in Chap. 7. A "good" question is easily comprehensible to respondents; it unambiguously and clearly communicates the survey researcher's intent. Three principles guide writing good survey questions for the face-to-face interview, whichever paradigm is used:

(1) Keep the respondent's perspective in mind. Avoid questions that might confuse or frustrate respondents. This means exercising care if the respondents are heterogeneous, have life situations very different from the researcher, or have priorities other than completing a survey research interview.
(2) Keep the interviewer in mind. Interviewers are allies who act as intermediates between the researcher and respondent. Provide specific directions to guide the interviewer and to make it easy to follow a researcher's intent in a dynamic face-to-face situation.
(3) Treat the entire interview schedule/questionnaire as an integrated whole, with a start and finish, and the questions flowing smoothly from one to another.

The first principle is critical. Survey questions should be clear, relevant, and meaningful to all respondents. However, we face a dilemma. If respondents have diverse social backgrounds and deploy varying frames of reference, using identical question wording may not carry an identical meaning for all respondents. Wording that some respondents find clear may create confusion for others due to differences in language usage/vocabulary, frames of reference, or life situations. As Schober and Conrad (1997) noted, misunderstandings can arise because a researcher's concepts may not map well onto a respondent's circumstances. The issue is especially important for non-mainstream respondents (i.e., not white, middle-class heterosexuals) and people whose subjective, lived reality does not mesh well with most mainstream survey researchers. "Conceiving of a 'standardized' subject behind the respondent casts the subject as a mere vessel of answers that can only be expressed in conventional terms, in relation to the standards that are assumed to be in place. This cheats the experiences of those whose lives are not lived in accord with, or may even be lived against the standard" (Dunbar et al. 2002, p. 295). Whether it is a group of elites (see Odenahl and Shaw 2002) or marginalized members of society, the great diversity of respondent lived experience and social position may not mesh well with using a single standardized survey for everyone.

Survey question writing takes skill, practice, patience, and creativity. The following list of ten things to avoid when writing survey questions includes only frequent potential problems–avoid jargon, slang, and abbreviations; avoid ambiguity, confusion, and vagueness; avoid emotional language and prestige bias; avoid double-barreled questions; avoid leading questions; avoid questions beyond respondents' capabilities; avoid false premises; avoid asking about distant future intentions; avoid double negatives; and avoid overlapping or unbalanced response categories. Such lists are framed as things to avoid, because studies to improve survey research focus on question problems and warn

against what not to do, while their proscriptive advice is general and vague.

B. Interview Schedule Design Principles

Interview schedule design includes the following four issues: layout, interview flow, interviewer instructions, and data recording.

1. Layout An interviewer has an interview schedule in front of him/her during the interview. It is simultaneously a script of what to say to a respondent, a map to guide the interview process, and an instrument for data gathering. Graphic layout concerns, such as font size, use of lines and diagrams, and placement of white space should make everything easy and clear for the interviewer. An interviewer might be in a calm, quiet setting with a serious, attentive respondent or in a noisy, confusing, fast-changing social situation with an annoyed, distracted respondent. A layout should support and assist the interviewer in the worst of situations. A layout that seems clear and straightforward under ideal conditions can easily become unintentionally confusing with unclear contingency/skip questions, difficult to convey questions, and awkward response categories in the worst of situations. Besides being a script and data collection instrument, an interview schedule acts as a prop to reinforce the researcher's authoritative voice, and it can be a "security blanket" for the uncertain interviewer.

2. Flow All interview schedules have a beginning with an introduction, a sequence of topics and questions, and a closing. Between the beginning and closing, the schedule guides the interview's pace and flow. After an introduction explaining the survey, it is best to make opening questions pleasant, interesting, and easy to answer (see Chap. 6 by Stalans on framing and the importance of introduction to the survey process). This helps respondents feel comfortable with the questionnaire, the interviewer, and the interview process. It is best to avoid asking many boring background questions or sensitive issue questions at the beginning. It is important to organize questions into common topics. Mixing questions on different topics creates unnecessary confusion. A good interview schedule orients respondents by grouping questions on the same topic together. Ideally, the interviewer's script introduces each section with a short introductory statement (e.g., "Now I would like to ask you questions about housing"). Question topics should flow smoothly and logically to assist respondents' memory and comfort levels. As the interview moves toward the closing, avoid frightening, embarrassing, or upsetting respondents with many sensitive issue questions. It is always best to end with a "thank you" and an offer to address any respondent concerns or questions.

3. Instructions Interviewer instructions are essential; however, instruction that are highly cryptic or overly elaborate can be worse than no instructions. Instructions should be in a different font/color for easy identification. Four types of interviewer instructions include the following: (1) instructions about introducing the survey or its sections (e.g., Hand the respondent the informed consent form and wait several minutes for her/him to read it, then ask whether there are any questions about it); (2) instructions about interview flow and direction (e.g., now show respondent Fig. 14.1, if yes is answered skip to next page question 4, tell the respondent "thank you, we have everything we need"); (3) instructions about respondent behavior or reactions (e.g., if respondent hesitates, appears confused, or appears uncertain, pause 10 seconds then offer this alternative wording); (4) instructions about using probes when respondents give incomplete or highly ambiguous answers (e.g., if respondent fails to give a complete answer, pause then repeat the question a second time, if still unclear ask "could you tell me more about that").

4. Data Researchers think ahead to data recording and analysis as they prepare an interview schedule/questionnaire. Interviewer errors often involve mistakes in marking respondent answers, so clear and easy to mark check boxes for closed-ended responses are best. Interviewer training usually drills in the importance of accuracy and completeness in recording respondent answers. Open-ended answers require extra attention. It is important to offer sufficient space and allot time if interviewers are

to write in open-ended responses. The interviewer must listen carefully, have legible writing, and record what a respondent says verbatim without editing, correcting grammar, or removing slang. The interviewer should never summarize or paraphrase since this causes a loss of information or distorts answers. For example, the respondent says, "I'm really concerned about my daughter's heart problem. She's only 10 years old and already she has trouble climbing stairs. I don't know what she'll do when she gets older. Heart surgery is too risky for her and it costs so much. She'll have to learn to live with it." If the interviewer writes, "concerned about daughter's health," much is lost. Audio recordings have advantages (e.g., removes worry about time to write responses, missing full responses, illegible writing) and disadvantages (e.g., a recording may be obtrusive and threatening to respondents, may have mechanical problems, include background noise or garbled responses).

C. Social Desireability and Questions on Sensitive Topics

Direct human interaction intensifies two negative survey effects, social desirability, and threatening questions, i.e., questions on sensitive issues. Social desirability, or respondents giving what they believe to be socially acceptable, ideal, or normatively appropriate answers instead of the truth, is greater in face-to-face than other survey modes (Heerwegh 2009). It may be that the most honest responses come from anonymous (e.g., self-administration or computer-based) rather than from personal contact (e.g., phone or face-to-face interviews) modes of survey administration (see Tourangeau et al. 2010; Tourangeau and Smith 1996; Tourangeau and Yang 2007, but see Holbrook et al. 2003). Respondents may be reluctant to answer completely and truthfully questions about sensitive issues (i.e., sexual behavior, drug or alcohol use, mental health problems, law violations, or socially unpopular behavior) or ones that threaten their presentation of self. As Willson et al. (2009) suggested, some respondents answer questions based on a self-image as much as from their actual behaviors. Some qualitative researchers suggest self-disclosure by the interviewer to elicit honest answers to sensitive issues (Dunbar et al. 2002; Reinharz and Chase 2002).

Survey researchers created an array of techniques to adjust question wording when respondents may hide or underreport behaviors or attitudes. In the face-to-face survey, interviewers can increase honest answering on sensitive issues in the following four ways: create comfort and trust, use enhanced phrasing, establish a desensitizing context, and use anonymous questioning methods. We briefly examine each next.

1. Create Comfort and Trust An interviewer can establish trust and a comfortable setting before asking threatening questions by displaying a calm, professional demeanor, reminding respondents of guarantees of anonymity and confidentiality, and emphasizing the need for honest answers. Researchers should place sensitive questions after a "warm-up period" of non-threatening questions and allow time for interviewers to establish rapport.

2. Use Enhanced Phrasing Indirectly asking about an issue may help to obtain honest answers. For example, instead of asking, "Have you shoplifted?"—which can carry an accusatory tone and uses the emotional word *shoplift,* that names an illegal act, interviewers might ask, "Have you even taken anything from a store without paying for it?" This describes the behavior, avoids emotional words, and opens the possibility that the behavior occurred under acceptable conditions (e.g., accidently forgetting to pay).

3. Establish a Desensitizing Context Contextual information can reduce threat. One way is to first ask about more serious issues than ones of real interest. For example, a respondent may hesitate to answer a question about engaging in shoplifting, but if the same question follows a list of more serious crimes (e.g., armed robbery, burglary), shoplifting appears to be less serious. A respondent may answer honestly because in context shoplifting appears to be a minor, almost trivial offense.

4. Anonymous Answering Some researchers have the interviewer to carry a laptop computer or other answer-marking device. For sensitive

questions, the interviewer reads questions, but the respondent indicates answers to them, him or herself on a keyboard or other device out of the interviewer's line of vision and without giving a verbal response (Couper et al. 2003).

14.4 The Face-to-Face Interview Process

A. Interviewing in Survey Research

As noted above, the survey interview is a social interaction in which each participant has assigned roles and tasks. Interviewer role expectations are made explicit in formal training sessions but respondent expectations are less clear, and an interviewer may have to motivate, instruct, or guide respondents. The face-to-face survey mode best allows an interviewer to motivate, guide, reassure, and make "on the fly" adjustments based on respondent answers or inquiries.

B. Phases of the Survey Interview Process

1. Introduction An interview proceeds through stages, beginning with an introduction and entry. The interviewer gets in the door, shows authorization, and reassures and secures cooperation from the respondent. He or she is prepared for reactions such as "How did you pick me?" "What good will this do?" "I don't know about this." "What's this about, anyway?" The interviewer explains why he or she must interview a specific respondent and not a substitute. Couper (1997) found that introductory statements by interviewers had a large impact on respondent cooperation and data quality, but are little studied. A decade later, Blohm et al. (2007, p. 96) observed, "reasons for different response rates among interviewer are still not totally clear." Financial incentives to respondents appear to increase respondent participation without serious side effects (Davern et al. 2003), although interviewers may not endorse such incentive payments (Lynn 2001; also see Toepoel Chap. 13 for a more thorough discussion on the effects of incentives). Experience and more cooperative, extroverted interviewers may reduce refusals (Sinibaldi et al. 2009) as well as "aversion training" for interviewers (Mayer and O'Brien 2001).

Two theories explain the initial cooperation-motivation process, Social Exchange and Leverage Saliency. *Social Exchange Theory* within the total design method (Dillman 2000) makes rational-choice assumptions and treats the interview as an exchange relationship in which respondents behave based on anticipated benefits for their efforts and cooperation. To increase cooperation and response rates, researchers should minimize the burdens by making participation easy and maximize rewards by providing benefits (i.e., feelings of esteem, material incentives, and emotional rewards) for cooperation.

Leverage Saliency Theory holds that a survey interview's salience varies by respondent (Groves et al. 2000). Respondents value, positively or negatively, specific aspects of the survey differently (e.g., length of time, topic of survey, sponsor, etc.). To maximize cooperation, researchers should identify and present positively valued aspects early in the interview process. Two practical implications are sponsorship and tailoring. Sponsorship refers to identifying the organization carrying out or paying for the survey. Tailoring refers to adjusting what interviewers say in an introduction to specific respondents. In tailoring, interviewers are sensitive to a range of household types and concerns, and learn to "read" the setting and various verbal and nonverbal cues, they then shift quickly to alternative scripts and tailor the presentation to a specific respondent (Groves et al. 2004).

The CC paradigm suggests a related way to increase motivation and cooperation with conversational format interviews–the interviewer first conducts ethnographic field research into the lived experiences of respondents before and during the interview (Dunbar et al. 2002). As interviewers grasp the respondent's subjective experiences and lifeworld, they relate to and motivate respondents in highly empathetic ways that are outside the constraints of a standardized interview.

2. Body of Interview Once an interviewer has made initial contact and secured cooperation, the next task is to motivate, direct, and guide respondents through the process, trying to prevent break-offs and reduce satisficing. Survey break-offs, i.e. when a respondent begins the survey but stops prior to completion, is an extreme form of item nonresponse, or when a respondent fails to answer specific items. As Peytchev (2009) suggested, break-off is related to demands and frustration with the survey process. Satisficing (Krosnick 1999) focuses on the respondent's cognitive tasks of answering survey questions. Respondents who initially agree to an interview may become fatigued or lose interest as task difficulty increases. They reduce effort and seek less burdensome, "easy" answers (e.g., middle response, nonresponse, or a pattern of acquiesce). Holbrook et al. (2003) suggest that satisficing is less a problem in face-to-face than other interviewing (e.g., telephone) because of personal motivation. Instead of satisficing, Moore (2008) found an "extreme response style" in which respondents tended to pick the extreme answer choices on a Likert scale.

In the standardized format interview, the interviewer uses exact questionnaire wording (without rephrasing) and asks questions in order (without returning to or skipping questions). He or she moves at a comfortable pace and provides limited nondirective, neutral feedback to maintain interest. In addition to asking questions, the interviewer accurately records answers. This is easy for closed-ended questions, for which interviewers just mark the correct box. For open-ended questions, the interviewer listens carefully, and then records in legible writing what the respondent said verbatim without correcting grammar or slang.

Probes, (an interviewer's request for a respondent to clarify an ambiguous answer, an incomplete answer, or an irrelevant response) can be controversial in standardized format interviews. Probes are a deviation from a standard script and vary by respondent. They are only used because some respondents interpret questions differently than intended by the survey designer, fail to answer completely, or do not adhere to role expectations. A probe requires that the interviewer understand the intent of a survey question, recognize inappropriate or incomplete responses, and apply specific follow-up actions in an attempt to elicit an appropriate or complete response (see further discussion of probes in section c below). Fixed-answer questionnaires that rely on the na assumption model imply simple,-mechanical responses; however, this makes it difficult for interviewers to detect respondent misunderstanding because answers may fit into preset categories.

It is possible to reduce unanticipated or unusable responses and to detect misunderstanding by supplementing closed-ended questions with some open-ended ones and the use of periodic probes. Yet, the open-ended questions and probes require additional time and well-trained interviewers, and they may produce nonstandard, difficult-to-quantify responses.

3. Exiting In the last stage, the interviewer thanks the respondent and leaves, and then goes to a quiet, private place to edit the questionnaire, and record other details. Interviewers may write a thumbnail sketch of the respondent and interview situation. This includes the respondent's attitude (e.g., serious, angry, or laughing) and unusual circumstances (e.g., "Telephone rang at question 27 and respondent talked for four minutes before the interview started again" "Teenage son entered room, sat at opposite end, turned on television with the volume loud, and watched a baseball game"). An interviewer also records personal feelings and suspicions (e.g., "I felt angry about the answer given to question 5," "Respondent became nervous and fidgeted when questioned about his marriage").

C. Guidance and Flow Control Issues.

In the standardized format interview, probes are short neutral requests to clarify an ambiguous answer and obtain a complete, relevant response. There are many types of probes. A three- to five-second pause is often effective. Nonverbal communication (e.g., tilt of head, raised eyebrows, or eye contact) may also work well. The interviewer can repeat the question or repeat the reply and then pause, or ask a neutral

question, such as, "Any other reasons?" "Can you tell me more about that?" "What do you mean?" "Could you explain more for me?" Different survey topics or question complexity may require varying amounts of interviewer intervention or aid. Complex, cognitively changeling questions may require greater assistance. As Zouwen et al. (2010) found, greater complexity requires interviewer intervention, which produces greater variation across interviewers.

Conversational or flexible format interviews encourage extensive tailored probing and active assistance to minimize respondent misunderstanding. After asking survey questions as written, the interviewer inquires into respondent understanding and offers explanations, rewordings, or long probes. A conversation ensues with the goal of clarification and reaching a shared understanding.

There is a critical tension between the uses of a standardized versus a conversational-flexible format interview. In the standardized format, researchers seek consistency across interviews in order to maximize measurement reliability. They assume that by having interviewers strictly follow a script and read survey questions identically they also maximize measurement reliability. However, formal behavioral consistency by an interviewer is not the same as substantive, interpretative consistency across respondents. As Conrad and Schober (2000) observed, respondent "comprehension can be made more consistent—and responses more comparable—when certain interviewer behaviors (discussions about the meaning of questions) are *less* consistent."

Paradoxically, non-standardized flexible-conversational interviewing may increase reliability over the standardized interview format by improving substantive, interpretative consistency (i.e., how respondents interpret the meaning of questions). This is because interpretative meaning across respondents may be more consistent when each respondent does not hear the exact same survey question. Schober and Conrad (1997) argued that a standardized format interview may work for simple questions in which a researcher's intentions map close to respondents' understandings. For more complex questions or cognitive tasks where such mapping is unlikely, the conversational format interview is probably most effective (also see Zouwen et al. 2010).

An assumption of the standardized format is that a private interaction occurs between two individuals, the respondent and the interviewer, but this is not always the case. The literature on the presence of others, "third" parties (usually spouses or children), during interviews has been mixed, but generally shows little effect. Pollner and Adams (1997) found little effect of spousal presence, while Zipp and Toth (2002) found that spousal presence had an influence on some responses, producing greater interspouse agreement. Pollner and Adams (1997, p. 624) said, "the findings offer comfort to researchers who ask questions about marital relations as a spouse looks on, they are unsettling from the view point of much social psychology, which underscores the sensitivity of actors to their interpersonal contexts."

14.5 Issues in the Face-to-Face Survey Interview

A. Interviewer Characteristics

Many studies examined the influence of specific interviewer characteristics. Data suggest that an interviewer's socially salient visible characteristics, including race and gender, might influence respondent answers for questions about issues related to race or gender. For example, African American and Hispanic American respondents express different policy positions on race-or ethnic-related issues depending on the apparent race or ethnicity of the interviewer (see Neuman 2011, pp. 347–349). It appears that interviewers of the same racial-ethnic group as a respondent get more honest answers than do interviewers from a different background (but see Dunbar et al. 2002). Yet, as Weisberg (2005, p. 61) noted, "Interviewer matching is rarely used in the United States, except when it is necessary to use interviewers who can speak another language.

" Gender may also affect interviews in terms of relevant issues, such as revealing sexual behavior, and support for gender-related collective action or gender equality. In cultures in which masculinity implies control, authority, and rationality along with minimal verbal responses and concealed emotions, flexible interviewing can be difficult. "An interview situation is both an opportunity for signifying masculinity and a peculiar type of encounter in which masculinity is threatened" (Schwalbe and Wolkomir 2002, p. 205).

Interviewer characteristics can influence answers in ways other than affecting candor and honesty. For example, an interviewer with visible disabilities caused respondents to lower their self-reported level of "happiness," compared to when they answered a self-administered questionnaire. Apparently, they did not want to sound too well off compared to the interviewer (Sudman et al. 1996, pp. 74–76). Flores-Macias and Lawson (2008) found asymmetric interviewer gender effects—affecting men more than women on gender sensitive items and dependent on geographic location (urban cosmopolitan versus other). They also suggest that interviewer characteristics vary by national culture.

Interviewer features such as age, physical appearance (attractive or not, neatness), and clothing style may influence respondents. The male with a shaved head and beard might garner differ reactions than the same male clean shaven with a modest, neat haircut. A 22-year-old interviewing an 80-year-old might get different responses than a 60-year-old who has more shared life experiences that could create closer bonds and shared understanding. Interviewer training recommends avoiding both highly formal attire and very informal, sloppy attire, instead aiming for the middle, and being nondescript. This rests on an assumption that nondescript dress will best neutralize interviewer effects.

Interviewer characteristics may not influence respondents in a simple way, but may interact with specific respondent characteristics or question topics. In addition, influence may take the form of nonresponse instead of altering a respondent's answer. As West and Olson (2010, p. 1022) remarked "… interviewer related variance on some key survey items may be due to nonresponse error variance, that is, differences in respondent characteristics across interviewers, rather than measurement differences." They found that interviewers of certain ages were more successful with respondents of certain ages.

Researchers who rely on the standardized format interview try to minimize variation across interviewers under the assumption that it will also reduce "interviewer effects." However, "interviewer effects" may remain even when the same interviewer is interviewing all respondents due to the dynamic social interaction that occurs between an interviewer and a specific respondent (Zouwen et al. 2010). In panel surveys, in which the same respondent has been interviewed over time, an interviewer's knowledge of past information about a respondent lessens their burden and reduces errors. Yet, such improvements in interviewer understanding and data accuracy that result when an interviewer starts with detailed information about a specific respondent contract the principles of the standardized format, i.e., treating all respondents identically (see Sala et al. 2011).

Interviewers vary in motivation, cognitive abilities, and experience. Researchers have sought highly motivated interviewers with strong cognitive abilities and experience. However, the issue of interviewer experience is complex. Olson and Peytchev (2007) found that interviewers' behaviors and perceptions evolved as they gained experience with a survey, and this affected data quality. Over time, interviewer familiarity with and repetition of the same survey questions can create boredom. With more time in the field, the interviewer becomes careless, less attentive, and speeds up the interview pace. A second possibility is that as interviewers learn more about the survey and responses, they modify their attitudes or behaviors in subsequent interviews. Olson and Peytchev also found that interviewer experience was negatively associated with interview length and interviewer interest. With more experience, interviewers tended to speed through interviews and perceive respondents as disinterested.

Researchers want interviewers who obey instructions diligently, engage and guide respondents, and produce quality data. Rosen et al. (2011) examined whether providing interviewers with monetary incentives for completed interviews changed interviewer behavior by reducing nonresponse, but they found no effect of incentives. However, Mayer and O'Brien (2001) found that greater interviewer training could reduce nonresponse rates.

14.6 Face-to-Face Survey Improvement

Strategies to reduce survey errors and misunderstandings began with pretesting or pilot testing the survey questionnaire in the 1950s. Early researchers administered a survey to a small group (e.g., 20–25) of surrogate respondents prior to full-scale implementation. They then looked for high refusal rates, many "don't know" answers, or hesitation and bewilderment by pretest respondents to specific survey questions. These problems indicated a need to revise specific questionnaire items. As Presser et al. (2004) noted, two shifts in pretesting occurred since the 1980s (also see Tourangeau 2003).

The first shift has been from trying to fix specific questionnaire items to improving overall data quality and reducing all forms of survey error. Error reduction and quality improvement mean capturing the respondent's "true" answer to the question's meaning as intended by a survey researcher, thus eliminating any miscommunication or miscomprehension. In short, the goal is a perfect match between a researcher's intended inquiry and a respondent's honest, candid, full response to that inquiry.

In the standardized format, the shift has meant eliminating deviations from a match believed to originate in the mediation of the interviewer, survey interview situation, or question wording. Researchers tried to write clearer survey questions, enhance respondent motivation to cooperate, improve the accuracy of respondent comprehension, reduce "interference noise" from the interviewer or interview situation, and raise the precision of data recording. In the conversational format, the shift has meant seeking interviewers who truly understand a researcher's intent and improving the interviewer-respondent communication-interaction.

The second shift has been to scrutinize the survey interview itself with cognitive interviewing and conversational analysis. Cognitive interviewing assumes the SRIR paradigm and a standardized format interview. It views the interview primarily as a cognitive information retrieval process. Conversational analysis (CA) assumes the CC paradigm and a conversational format interview. It views the interview as a specialized type of conversation/social interaction.

1. Cognitive Interviewing

Cognitive interviewing relies on psychological models of the cognitive tasks performed by a respondent when answering survey questions. Beatty and Willis (2007, p. 288) defined it as, "administering draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that the author intends." Cognitive interviewing occurs prior to full survey administration. A researcher attempts to improve the efficiency of extracting accurate information from a respondent who is completing a sequence of cognitive tasks. Tourangeau et al. (2000, p. 8) identified four steps in the process: **comprehension**–respondent attends to question and processes it by creating links to relevant concepts; **retrieval**–respondent generates a retrieval strategy, uses memory and cues to gather specific information; **judgment**–respondent assesses completeness and relevance of information, integrates information, and makes an estimate of adequacy based on partial information; and **response**–respondent makes a judgment about appropriate answer category in the survey question, then edits and delivers a verbal response

Two major forms of cognitive interviewing are think-aloud and interviewer probes (see Beatty 2004; Beatty and Willis 2007; Willis 2005). In the "think aloud" method, a surrogate respondent participates in a survey interview pretest but simultaneously narrates his/her thinking while processing the question and delivering a response. For example, a respondent is asked the open-ended question, "When did you last see a doctor?" A respondent says, "Let me see, I had my annual checkup in August. Then in October, I went to the clinic for a flu shot, but I only saw a nurse not a doctor. I had a tooth extracted by an oral surgeon in January. He was really more of a dentist than a doctor. After a question from a pharmacist, I also phoned my doctor to change my medication dosage in January, but did not go in for a visit. I got new glasses in February. My eye doctor is an ophthalmologist with an MD degree. In early March, I went to my wife's doctor with her for a consultation on her upcoming operation. So I last saw a doctor myself in February, unless you do not count eye doctors, then the last time I saw a doctor myself was in October." This reveals three issues: (a) the potential confusion between "doctor" and other medical professionals; (b) the meaning of "seeing" a doctor as a personal physical visit by the respondent, and (c) the respondent's chronological reconstruction of medical-related events over several months. A researcher can use this data to redesign and reword the survey question. It also may compel the researcher to clarify his/her exact intent behind the survey question.

Interview probing has a highly skilled interviewer ask multiple probe questions about a pretest interview after a respondent has completed it. For example, an interviewer might ask, "in the question, *When did you last see a doctor*? you answered three weeks ago. How did you arrive at that answer?" Through a series of follow-up probe questions, the interviewer reconstructs the twists and turns of the respondent's decision-making process. Each method has its advantages (see Willis 2004).

Related methods of pretesting include expert panels, focus groups, and behavior codings. Expert panels are teams of highly experienced survey researchers that review a questionnaire for potential errors. In the focus group six to twelve surrogate respondents have an open-ended group discussion on a topic mediated by an observer, and it is recorded. The discussion can reveal respondent thinking on a topic. In behavioral coding, a researcher codes a transcript, audio or videotape of an interview for specific actions or statements–such as an interviewer misreading questions, a respondent's requests for clarification, long pauses by a respondent before answering, requests to repeat answers or questions, and so forth. Cognitive interviewing and related methods vary in results. There is some question about whether they have improved questionnaires (see Presser et al. 2004).

2. Conversational Analysis

Conversational Analysis (CA) views the survey interview as a contextualized social-conversational interaction and holistic process in which interviewer and respondent engage in joint meaning production. It questions assumptions of the standardized format (see Moore and Maynard 2002; Schaffer 2002) and is rooted the different theoretical assumptions. As Houtkoop-Steenstra (2000, p. 17–18) remarked, "… the so-called inadequate interviewer and respondent behavior is primarily an artifact of incorrect theoretical assumptions about the nature of conversational interaction." CA draws on the ethnomethodological and sociolinguistic traditions. Ethnomethodology assumes that all human interaction is a fluid, on-going accomplishment produced jointly by participants. The interviewer, respondent, and indirectly the physically absent survey researcher mutually construct meanings and understandings within the constraints of an interview conversation.

CA rejects the separation assumed in a cognitive, information-processing model. As Schober and Conrad (2002, p. 68) noted, "… by completely separating interviewer behavior, respondent behavior, and question wording, this

prevailing approach relies on a view of communication that has been discredited." In CA, an interview is a dynamic, integrative process in which all participants draw upon common sense and tacit knowledge, as well as explicit survey questions, to produce mutual understanding. In short, features that cognitive interviewing treats as being problematic or sources of errors to eliminate, CA sees as essential for successful communication, including survey interviews.

Rather than try to eliminate "errors" as respondent and interviewer jointly attempt to process an absent researcher's request, CA seeks to uncover the ordered processes by which people construct meanings in a particular type of social encounter. CA treats the interview as a "site" of human interaction in which people combine elements of human communication, (i.e., complements, laughter, rationality, and emotion) for a purpose (Gathman et al. 2008).

CA describes/analyzes the interaction itself, at the same time, its analysis can inform and improve both conversational and standardized format interviews (Maynard and Schaeffer 2000). "As powerful as information-processing models have been in helping us understand how survey questions are answered, they can usefully be supplemented by paying attentions to the social aspects of surveys" (Schaeffer and Presser 2003, p. 68).

In CA, a researcher conducts a detailed examination of interview transcripts or recordings. This is similar to the behavioral coding method described above. A researcher notes communication slips, adjustments, corrections, lapses, and so forth. However, rather than treat these as "errors" for elimination, in CA they are vital features of communication processes that participants have drawn on from ordinary conversations as they work to accomplish coherence and mutual understanding.

In the conversational interview format, an interviewer has an active role to assist the respondent in comprehending survey questions' meaning as well as what the survey researcher intended. Interviewer and respondent together seek an accurate understanding of the researcher's intentions. Once a respondent understands and answers a question, the interviewer clarifies the answer, and then conveys it back to the researcher. Thus, interview "improvement" in CA does not push for a closer alignment between interviewer behavior and a standardized script; rather, it seeks a closer alignment between the researcher's intended question meaning and an accurate, meaningful answer from a specific respondent.

CA researchers examine interview interactions to uncover possible impediments to the respondent's involvement or misunderstandings of the researcher's inquiry. For example, Schaeffer and Thomson (1992) examined "don't know" response categories in surveys and revealed a range of respondent uncertainty-neutrality, ambivalence, and lack of clarity. Houtkoop-Steenstra (2000) explored how survey questioning deviates from the normal flow of conversations, such that a standardized format introduces communication disruptions and greater confusion by blocking normal conversational repair actions. Maynard et al. (2010) examined how respondents understood requests to participate in survey interviews as either encouraging or discouraging.

The holistic perspective of CA incorporates all meaningful behavior in the interview encounter. As Mathiowetz (1998, p. 48) observed, "… a respondent's behavior during the course of a survey interview provides a wealth of data that extends beyond the substantive responses to the questionnaire." Wortham et al. (2011) found that the "denotation text" of survey questions–responses and an "interactional text" of all verbal exchange and actions during an interview occasion each reveals distinct and important information about a respondent. They warned against ignoring the interactional text.

In summary, cognitive interviewing and CA both can improve face-to-face survey interviews but they rest on different models of the interview. Cognitive interviewing tries to improve how the respondent processes a sequence of cognitive-behavioral tasks and seeks smooth, accurate information retrieval. CA examines the

dynamics of the interview interaction. It can help to improve the flow of communication and enhance interviewer-respondent mutual understanding. We can apply both techniques to both standardized and conversational format interviews.

14.7 Planning and Design in Survey Management

A. Before the first interview

Planning begins with the decision to conduct a face-to-face interview instead of other modes. The decision triggers issues of respondent sampling, questionnaire design, interviewer selection and training, cost estimates, and logistics.

1. Respondent selection The primary selection difference between face-to-face and other modes is that face-to-face interviewing requires physically locating the respondent and transporting the interviewer to that location. In addition to identifying and screening respondents (e.g., have properly sampled person, sufficient mental competence, language facility), an agreed-upon physical location and interview time must be established. Although often in the respondent's home, another "neutral" location (e.g., an office, library, school classroom) that offers security, privacy, and comfort for both interviewer and respondent is possible. The interviewer, who carries photo identification, must verify respondent identity and secure cooperation for survey participation. This may entail describing the survey purpose and sponsor, outlining the interview process, providing assurances of confidentiality, and obtaining informed consent.

B. During the survey interview

The interview follows either a standardized or conversational format. In the standardized format, an interviewer reads each question as written and in order, remains neutral and unobtrusive, rarely uses probes, and strictly follows a preestablished protocol; respondent requests for clarification are addressed minimally and in a standardized manner. In the conversational format, an interviewer devotes significant time and effort to establishing rapport and clarifying the purpose of the interview. Interviewer and respondent go through questions together with the interviewer ensuring that the respondent understands each question as the researcher intended; the interviewer continuously adjusts the interview based on respondent needs and rewords questions to maximize respondent understanding.

C. After the interview

Once the interview ends and the interviewer exits, the next phase is recording immediate reactions, reviewing the interview schedule, and creating a record of how the interview went. The survey planner builds in time for the interviewer to complete these tasks and interviewer training includes a protocol for immediate post interview processing and documentation. An ending protocol might include the following: record the interview ending time and call the supervisor to report interview completion; document the interview event, such as any problems (e.g., disruptions, other people entering) or unusual situation; write a short evaluation of the interview session; review the interview schedule, noting blanks, adding clarifying notes, or qualifications.

D. Quality control and management issues

Quality control begins with survey design. In addition to writing good survey questions, providing quality interview schedules, and training the interviewer, quality control includes interviewer supervision. Prior to the interview, the supervisor meets with interviewers, addresses questions and reviews procedure, and oversees the interviewer's transportation and personal safety issues. After the interview, the supervisor verifies the respondent contact, reviews the interview schedule and interviewer comments, and addresses interviewer concerns.

14.8 Conclusion

The face-to-face survey has distinct benefits not present in other modes of survey administration, but it comes with costs or risks. Its dominant feature is direct contact with another person in an interview. Two paradigms of the interview

encounter (Stimulus–response/information retrieval and Constructionist-collaborative) imply using alternative face-to-face interview formats (standardized or conversational) to achieve high measurement reliability. The two paradigms also imply different ways to examine and improve survey interviews, i.e., cognitive interviewing or conversational analysis. We can use insights from the paradigms in complementary ways. An exclusive reliance on the dominant stimulus–response/information-retrieval paradigm and strict adherence to a standardized format may achieve high levels of behavioral consistency. However, this may limit our ability to acquire rich survey data with greater substantive consistency. A conversational interview format, as supported by the constructionist-collaborative paradigm, is only possible in the face-to-face survey and may offer the greatest substantive consistency. It requires more time and higher costs than a standardized format interview, and it may yield data that mixes variation originating in a respondent's responses with variation across interview situations.

References

Alonso, W., & Starr, P. (Eds.). (1987). *The politics of numbers*. New York:Russell Sage Foundation.

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmonth CA:Wadsworth.

Barnes, J. A. (1979). *Who should know what? social science, practice and ethics*. New York:Cambridge University Press.

Beatty, P. (2004). The dynamics of cognitive interviewing. In S. Presser, J. Rothgeb, M. Couper, J. Lesser, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 45–66). New York:Wiley.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: the practice of cognitive interviewing. *Public Opinion Quarterly, 71*, 287–311.

Blohm, M., Hox, J., & Koch, A. (2007). The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face household surveys. *International Journal of Public Opinion Research, 19*, 97–111.

Bradburn, N. M., & Sudman, S. (1988). *Polls and surveys*. San Francisco:Jossey-Bass.

Burawoy, M. (2005). For public sociology. *American Sociological Review, 70*, 4–28.

Clawson, D., Zussman, R., Misra, J., Gerstel, N., & Stokes, R. (Eds.) (2007). *Public sociology. Fifteen eminent sociologists debate politics and the profession in the twenty-first century*. Berkeley:University of California Press.

Conrad, F. G. & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, *64*, 1–28.

Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion Quarterly, 61*, 317–338.

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly, 67*, 385–395.

Czaja, R., & Blair, J. (2005). *Designing surveys: A guide to decisions and procedures*. Thousand Oaks, CA:Pine Forge.

Davern, M., Rockwood, T. H., Sherrod, R., & Campbell, S. (2003). Prepaid monetary incentives and data quality in face-to-face interviews. *Public Opinion Quarterly, 67*, 139–147.

Della Porta, D., & Keating, M. (Eds.). (2008). *Approaches and methods in the social sciences: A pluralist perspective*. New York: Cambridge University Press.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17*, 30–52.

Dunbar, C., Rodriquez, D., & Parker, L. (2002). Race, subjectivity and the interview process. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 279–297). Thousand Oaks, CA: Sage.

Flores-Macias, F., & Lawson, C. (2008). Effects of interviewer gender on survey responses: findings from a household survey in Mexico. *International Journal of Public Opinion Research*, *20*, 100–110.

Foddy, W. (1993). *Constructing questions for interview and questionnaires: Theory and practice in social research*. New York: Cambridge University Press.

Fowler, F. J, Jr. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.

Gathman, E. C. H., Maynard, D. W., & Schaeffer, N. C. (2008). The respondents are all above average: Compliment sequences in a survey interview. *Research on Language & Social Interaction, 41*, 271–301.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*, 2–31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. *Public Opinion Quarterly, 64*, 299–308.

Gubrium, J. F., & Holstein, J. A. (2002). From the individual interview to the interview society. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 3–31). Thousand Oaks, CA: Sage.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research, 1*, 111–121.

Holbrook, A., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly, 67*, 79–125.

Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire*. New York: Cambridge University Press.

Kahn, R. L., & Cannell, C. F. (1957). *The dynamics of interviewing: Theory, technique and cases*. New York:Wiley.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567.

Krysan, M., Schuman, H., Scott, L., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58*, 381–399.

Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review, 66*, 453–479.

Lynn, P. (2001). The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research, 13*, 326–336.

Mathiowetz, N.A. (1998). Respondent expressions of uncertainty. *Public Opinion Quarterly, 62*, 47–56.

Mayer, T. S., & O'Brien, E., (2001). *Interviewer refusal aversion training to increase survey participation*. Proceedings of the annual meeting of the American Statistical Association

Maynard, D. W., Freese, J., & Schaeffer, N. C. (2010). Calling for participation requests, blocking moves and radical (Inter) action in survey introductions. *American Sociological Review, 75*, 791–814.

Maynard, D. W., & Schaeffer, N. K. (2000). Toward a sociology of social scientific knowledge: Survey research and ethnomethodology's asymmetric alternates. *Social Studies of Science, 30*, 323–370.

Maynard, D. W., & Schaeffer, N. K. (2002). Standardization and Its discontents. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge* (pp. 3–45). New York: Wiley.

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Moore, R. J., & Maynard, D. W. (2002). Achieving understanding in the standardized interview repair sequences. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge* (pp. 281–311). New York: Wiley.

Moors, G. (2008). Exploring the effect of middle response category on response style in attitude measurement. *Quality & Quantity, 42*, 779–794.

Neuman, W. L. (2011). *Social research methods: Quantitative and qualitative approaches* (7th ed.). Boston: Allyn & Bacon.

Odenahl, T., & Shaw, A. M. (2002). Interviewing elites. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 299–316). Thousand Oaks, CA: Sage.

Oishi, S. M. (2002). *How to conduct in-person interviews for surveys* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*, 273–286.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly, 73*, 74–97.

Pollner, M., & Adams, R. E. (1997). The effects of spouse presence on appraisals of emotional support and household strain. *Public Opinion Quarterly, 61*, 615–626.

Porter, T. M. (1995). *Trust in numbers*. Princeton, NJ: Princeton University Press.

Presser, S., Rothgeb, J., Couper, M., Lesser, J., Martin, E., Martin, J., & E. Singer. (2004). The dynamics of cognitive interviewing. In S. Presser, J. Rothgeb, M. Couper, J. Lesser, E. Martin, J. Martin & E. Singer (Eds.) *Methods for testing and evaluating survey questionnaires* (pp. 45–66). New York: Wiley.

Reinharz, S., & Chase, S. E. (2002). Interviewing women. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 221–235). Thousand Oaks, CA: Sage.

Rosen, J., Murphy, J., Peytchev, A., Riley, S., & Lindblad, M. (2011). The effects of differential interviewer incentives on a field data collection effort. *Field Methods, 23*, 24–36.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.). (1983). *Handbook of survey research*. New York: Academic Press.

Sala, E., Uhrig, S. C. N., & Lynn, P. (2011). It is time computers do clever things!: The impact of dependent interviewing on interviewer burden. *Field Methods, 23*, 3–23.

Schaeffer, N. C., & Maynard, D. W. (2002). Standardization and interaction in the survey interview. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 577–597). Thousand Oaks, CA: Sage.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65–88.

Schaeffer, N. C., & Thomson, E. (1992). The discovery of grounded uncertainty: Developing standardized questions about strength of fertility motivation. *Sociological Methodology, 22*, 37–82.

Schaffer, N. C. (2002). Conversation with a purpose—or conversation? Interaction in the standardized interview. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge* (pp. 95–123). New York: Wiley.

Schegloff, E. A. (1990). Interactional troubles in face-to-face survey interviews: Comment. *Journal of American Statistical Association, 85*, 248–250.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*, 576–602.

Schober, M. F., & Conrad, F. G. (2002). A collaborative view of standardized survey interviews. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge* (pp. 67–94). New York: Wiley.

Schwalbe, M. L., & Wolkomir, M. (2002). Interviewing men. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 203–219). Thousand Oaks, CA: Sage.

Sinibaldi, J., Jackle, A., Tipping, S. & Lynn, P. (2009). *Interviewer characteristics, their doorstep behavior, and survey co-operation*. American Association for Public Opinion Research annual meeting, (May 2009).

Steinmerz, G. (Ed.). (2005). *The politics of method in the human sciences: Positivism and its epistemological others*. Durham, NC: Duke University Press.

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*, 232–241.

Suchman, L., & Jordan, B. (1992). Validity and the collaborative construction of meaning in face-to-face surveys. In J. Tanur (Ed.), *Questions about questions: inquiries into the cognitive bases of surveys* (pp. 241–267). New York: Russell Sage Foundation.

Sudman, S., & Bradburn, N. M. (1983). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey research*. San Francisco: Jossey-Bass.

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research, 15*, 3–7.

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775–801.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents. *Public Opinion Quarterly, 74*, 413–432.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

Tourangeau, R., & Yang, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.

Weisberg, H. F. (2005). *The total survey error approach: A guide to the new science of survey research*. Chicago, IL: University of Chicago Press.

West, B. T., & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly, 74*, 1004–1026.

Willis, G. B. (2004). Cognitive interviewing revisited: A useful technique in theory? In S. Presser, J. Rothgeb, M. Couper, J. Lesser, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 23–44). New York: Wiley.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Willson, S., Stussman, B., Maitland, A., & Nahin, R. L. (2009). Role of self-concept in answering survey questions on complementary and alternative medicine: Challenges to and strategies for improving data quality. *Journal of Alternative and Complementary Medicine, 15*, 1319–1325.

Wortham, S., Mortimer, K., Lee, K., Allard, E., & White, K. D. (2011). Interviews as interactional data. *Language in Society, 40*, 39–50.

Zipp, J. F., & Toth, J. (2002). She said, he said, they said: The impact of spousal presence in survey research. *Public Opinion Quarterly, 66*, 177–208.

Zouwen, J., Smit, J., & Draisma, S. (2010). The effect of the question topic on interviewer behavior: An interaction analysis of control activities of interviewers. *Quality & Quantity, 44*, 71–85.

Repeated Cross-Sectional Surveys Using FTF

15

Ineke Stoop and Eric Harrison

15.1 Introduction

15.1.1 About the Chapter

This chapter is about designing a cross-national face-to-face survey. Different aspects of survey design have been covered in a wide range of books and journal articles (Groves 1989; Biemer and Lyberg 2003; De Leeuw et al. 2008; Marsden and Wright 2010). In fact, almost everything that has been written on survey methodology and survey quality is directly or indirectly relevant for survey design.

Given that there is already a substantial body of literature on this subject, we focus here on one particular category of surveys, namely those that are administered face-to-face. This narrows the scope substantially, as there is no need to pay attention to difficult and challenging issues such as the design of online questionnaires. Still, considering the sheer volume of the literature on interviewing and interviewers, even this is a very broad topic. And it is even broader here, because we focus especially on *cross-national* surveys, by which we mean surveys that are designed to compare countries and that are fielded in different countries, cultures and languages.

For this reason we cannot and will not attempt to give a complete and detailed overview of all topics pertaining to the design of cross-national surveys. Instead we will give a general overview of the issues under consideration and refer to the existing literature. We will go into more detail when the cross-national perspective becomes important. Throughout the chapter we will pay particular attention to the European Social Survey (ESS), from which we will draw many examples. This survey is particularly relevant because it specifically aims to achieve high quality, its design aspects are extensively documented and both authors have been involved in this survey over many years.

In this chapter we look first at the different steps in survey design, summarised in the concept of the survey life cycle. Although the focus is on face-to-face surveys, we then briefly compare face-to-face administration with other data collection modes. Section 15.2 presents a short overview of the ESS. Section 15.3 covers the first important aspect of surveys, namely how to make sure that the final respondents represent the target population (or how to minimise non-observation errors); Sect. 15.4 the second key aspect, namely how to make sure that the answers to survey

I. Stoop (✉)
The Netherlands Institute for Social Research/SCP,
PO Box 16164, 2500 BD, The Hague,
The Netherlands
e-mail: i.stoop@scp.nl

E. Harrison
Centre for Comparative Social Surveys,
City University, Northampton Square,
London, EC1V 0HB, UK
e-mail: eric.harrison.1@city.ac.uk

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_15, © Springer Science+Business Media New York 2012



Fig. 15.1 Survey Life Cycle (Survey Research Center 2010)

questions represent the concepts we wish to investigate (or how to minimise observation errors). The final section presents some thoughts on the future of face-to-face surveys.

15.1.2 The Survey Life Cycle

Many steps must be taken in order to design a survey, and even more to design a survey that will be fielded in different countries. Each step will have an impact on the feasibility, timing, costs and quality of the survey data. Figure 15.1 gives a concise overview of the different steps in survey design and implementation (Survey Research Center 2010). This overview is taken from a very useful website comprising Cross-Cultural Survey Guidelines (www.ccsg.isr.umich.edu).

As is clear from Fig. 15.1, these guidelines assume that data will be collected by face-to-face interviewers. However, as Fig. 15.2 shows, in many countries the face-to-face mode now represents a small proportion of the overall expenditure on quantitative data collection in the market research industry. As face-to-face surveys are an expensive mode of data collection (see Sect. 15.1.3), this means that the proportion of actual *data* collected face-to-face in the industry is even smaller. Despite this, face-to-face data collection is the one survey mode that is almost universally applicable, in every country and in every cultural group, and for this reason it remains the dominant paradigm in cross-national social surveys, such as the ESS, the Adult Literacy and Life Skills Survey, the Eurobarometer, the European Values Studies and World Values Surveys, the Survey on Health, Ageing and Retirement in Europe and the World Mental Health Survey (Koch et al. 2009). As face-to-face administration is now becoming a rarity in several countries, it is questionable how long this can continue.

One of the first steps in designing a survey is the choice of the data collection mode. This largely determines the funds required and the

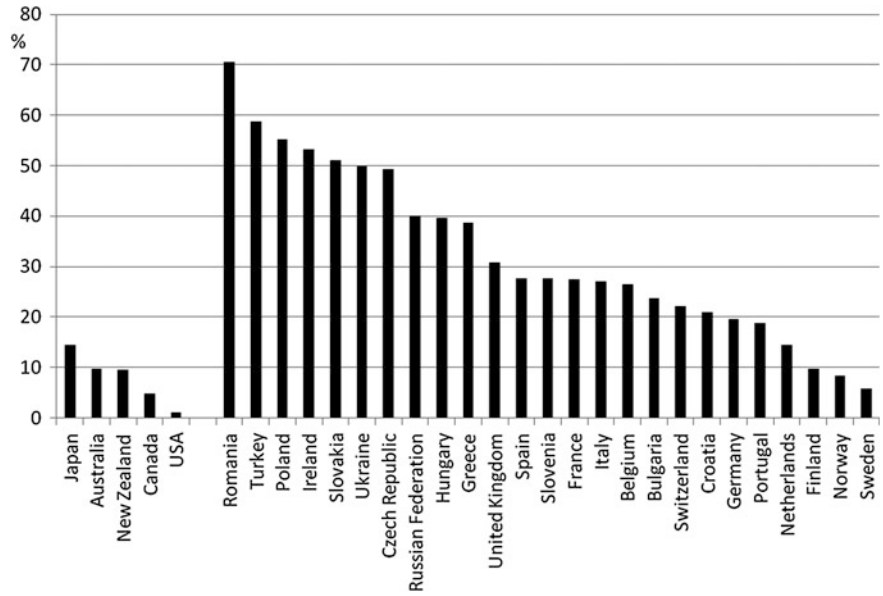


Fig. 15.2 Expenditure by face-to-face as share of spend total quantitative (%). Countries participating in the European Social Survey, and other large western countries

duration of the fieldwork. This decision will have a major impact on many of the steps in the survey life cycle and also on the quality of the survey outcomes. We will present a short overview of the advantages and disadvantages of different data collection modes in Sect. 15.1.3, but will generally assume that this choice had already been made, and will then focus on face-to-face surveys, and in particular cross-national face-to-face surveys.

This chapter does not cover every step in the survey life cycle. Firstly, it focuses on the design phase, and not on the implementation. This means that issues such as non-response adjustment, data dissemination and archiving receive very little attention. It should be noted, however, that even these issues will have to be partially addressed during the design phase. The collection of auxiliary data to adjust for non-response will need to be part of the design phase, as will obtaining informed consent from respondents to permit the dissemination of the data. A second limitation is that this chapter focuses only on those steps that are specific to face-to-face surveys.

In describing aspects of the design of face-to-face surveys, this chapter draws its examples almost exclusively from the European Social Survey (henceforth 'ESS'). This is a study that strives to maintain the highest standards (and indeed to improve survey methodology more widely), is exclusively based on face-to-face interviewing and has been implemented in more than 30 European countries. Since this chapter draws so heavily on the ESS, Sect. 15.2 will give a short overview of the most salient features of this study.

15.1.3 Face-To-Face Versus Other Data Collection Modes

Face-to-face (or personal visit) interview surveys (for short face-to-face surveys) are the oldest mode of data collection (Groves et al. 2009). They were soon accompanied by mail surveys, and in the late 1960s telephone surveys became more common. With the arrival of computers in the survey process, computer-assisted forms of face-to-face and telephone surveys were introduced: Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Telephone Interviewing (CATI). The old-fashioned computer-less face-to-face survey become known as PAPI (Paper and Pencil Interviewing).

It should be noted that the involvement of the interviewer, and thus interviewer effects, in the PAPI approach is much greater than in the CAPI approach (see Neuman's previous, Chap. 14).

Face-to-face and telephone surveys require the presence of an interviewer, whereas mail and drop-off surveys (where the interviewer leaves behind a questionnaire that is completed by the respondent) are self-completion surveys. New forms of self-completion surveys came into being at the end of the last century: Computer Assisted Self Interviewing (CASI), web surveys, interviewer-less forms of telephone interviewing based on audio-CASI (taped text) and Interactive Voice Recording (see Groves et al. 2009, pp. 150–153). Increasingly, mixed mode models of data collection are also being implemented.

De Leeuw (2008) analyses what really makes a face-to-face interview. She distinguishes *interviewer effects*, *media-related factors* and factors that influence *information transmission*. *Interviewers* can guide respondents through a standardised interview, probe and clarify when required and check whether the respondent has understood the questions. Interestingly, a number of recent studies have shed additional light on this probing and providing clarification. De Leeuw et al. (2011) showed that probing after a DK ('don't know') response is perfectly possible in telephone and web surveys. Conrad et al. (2006), Peytchev et al. (2010) studied clarification in web surveys. They found that clarification can easily be given in a web survey, but respondents seem inclined to feel that even a little effort is too much to obtain a definition of a concept or a clarification (mouse rollovers are more effective than clicks). Strikingly, 'respondents who understand that the information contained in definitions may complicate the process of answering the question (e.g. the definition may require respondents to revise their understanding of a concept) try to avoid this extra effort' (Peytchev et al. 2010, p. 644). This could be one of the main differences between face-to-face an online surveys: without the presence of an interviewer respondents are keen to go through an online survey as quickly as possible.

With regard *to media-related factors*, De Leeuw (2008) mentions familiarity with the medium or data collection mode, the locus of control, silences and sincerity of purpose. Familiarity with the medium will definitely change over time and differ across countries. The locus of control in face-to-face surveys is the interviewer. This could mean that answers in self-completion surveys are more correct because the respondents record these themselves. On the other hand, the presence of an interviewer precludes the possibility that respondents will look up the answers to knowledge items, can increase the likelihood that respondents will look up their income data, has some inkling as to whether an answer is correct (weight, age) and can make sure that answers are given by the respondents themselves and not by a family member or as 'fun for the family'. In face-to-face situations, silences are much less of a problem than in telephone surveys, because the interviewer can see what is going on, and can try to bridge the silences. And finally, an interviewer can convey the sincerity of purpose and communicate trustworthiness much more than is possible in other modes.

Finally, with regard to *information transmission* De Leeuw (2008) distinguishes the presentation of information, channels of communication and the communication flow. Face-to-face surveys provide information aurally (sounds), orally (speaking), and visually (showcards, non-verbal communication). This has many advantages, but reading a long list of characteristics may take a lot of time (and be somewhat boring). With regard to the channels of communication a distinction can be made between verbal, communication, nonverbal communication and paralinguistic communication (tone, timing, emphasis, humming). These different channels can help the conversation along and motivate the respondent, but they can also cause error (different emphasis by different interviewers, nonverbal disapproval of answers given). Finally, in face-to-face surveys nonverbal cues can indicate that an answer is final, or more probing might be useful.

Despite the proliferation of data collection modes, face-to-face interviewing has for a long time been viewed as the gold standard (De Leeuw 2005; Schober and Conrad 2008). The ESS, for instance, which aims '… to produce rigorous data about trends over time in people's underlying values within and between European nations, and to rectify longstanding deficits in the rigour and equivalence of comparative quantitative research, especially in attitude studies' (Jowell et al. 2007, p. 9) has explicitly opted for face-to-face interviewing. As Biemer and Lyberg (2003, p. 189) stated '… face-to-face interviewing is often associated with good data quality and is viewed by many survey researchers as the preferred mode of data collection for most survey topics'. From their work and the work of others; however, it is clear that face-to-face surveys have advantages but also important disadvantages.

De Leeuw (2008, pp. 133–134) gives a useful overview of the pros and cons of different modes of data collection. With respect to face-to-face surveys it is important to take into account the triple role of face-to-face interviewers. In the *recruitment phase* interviewers have to visit respondents in their homes. Travel is usually a high-cost component of this approach, especially when establishing contact requires a large number of visits, or when sample units live in sparsely populated areas. Once the cooperation of the respondents has been obtained, interviewers have to *interview* respondents at a convenient time, if possible without too many disturbances and interference from others. And finally, they have to perform a lot of *administrative tasks*, such as selecting target respondents, keeping call records, collecting observational data and completing information on the course of the interview. Face-to-face surveys require a large budget because visiting sample units and conducting interviews requires a lot of interviewer time and involves substantial travel costs. This brings us to the pros and cons of conducting an interview face-to-face.

In the *recruitment phase* interviewers have to establish contact with the sample units and obtain their cooperation. The value of face-to-face surveys as a means of obtaining high response rates is almost universally acknowledged. Holbrook et al. (2003, pp. 94, 113) ascribe the higher response rates in face-to-face surveys partly to the reassuring effect emanating from having an interviewer on the doorstep, while Groves et al. (2009, p. 166) attribute it to the greater credibility that the interviewers are who they say they are. There is a good deal of evidence that tailoring persuasion strategies to particular respondents is very important in gaining response and that this strategy can be optimally deployed in face-to-face contact. On the other hand, there is some evidence that interviewers can increase non-response bias because individual interviewers are more successful with some groups than with others (West and Olson 2010, see also Chap. 9 by Stoop on unit non-response due to refusal). The main disadvantages of using interviewers are costs and the length of time required to run a face-to-face survey.

During the *interview* the interviewer can help the respondent, probe for more complete and accurate answers, use showcards, provide explanations when required, hold the respondent's attention and build confidence and rapport during the interview. Broken-off face-to-face interviews are rare, and satisficing—choosing the answer that requires least thought to minimise the time spent answering questions (Krosnick et al. 1996)—occurs less often than in other modes (Heerwegh and Loosveldt 2008; Roberts 2007), especially in long or cognitively burdensome surveys.

In this phase, the greatest asset of the survey interview—the presence of an interviewer—can also be its greatest weakness (De Leeuw 2008, p. 134). Face-to-face surveys may generate social desirability bias, especially for sensitive questions (De Leeuw 2008). Interviewers have been shown to have an impact on the answers given by respondents. This 'interviewer effect' is shown by intra-interviewer cluster effects (see Biemer and Lyberg 2003; Schaeffer et al. 2010). One reason why good interviewers manage to obtain response and guide respondents through a long questionnaire is that they establish rapport. It is not certain; however, whether this rapport leads to more honest answers (Anderson 2008).

Interviewers also have to perform a number of *administrative tasks*. They have to record the calls they make on contact forms, provide information on reasons for refusal, assess whether non-cooperative target persons would be willing to participate in the survey at a more convenient time, answer questions on the interview process and collect observational data on the dwelling and neighbourhood. As interviewers are on the spot, observe what is going on and have contact with many refusals, they are an ideal source of information. The main challenge here is to find a balance between maximum information on the survey process and an acceptable burden for the interviewer.

Despite the fact that face-to-face surveys are expensive—and often very expensive—and that fieldwork periods may have to be extended to provide good response rates, and despite the fact that interviewers may influence survey outcomes, the fact remains that face-to-face surveys are in some cases the best, or even the only option. Telephone surveys always exclude the members of a population without a fixed-line telephone and usually the increasing numbers who own only mobile phones. Web surveys exclude those without access to the Internet, and self-completion surveys generally rule out those sections of the population who are functionally illiterate' (in the Netherlands, for instance, more than 10 % of the population). Face-to-face surveys exclude no one, and allow for long and complex interviews that can take an hour or more. In addition, face-to-face surveys have relatively few coverage problems, and can be based on different types of sampling designs: lists of individuals, households and addresses and also area probability sampling. And finally, they usually have the highest response rates. For a thorough discussion on face-to-face please refer to the previous chapter by Neuman (Chap. 14).

15.2 The European Social Survey

15.2.1 History and Background

The ESS was first conceived in the 1990s in response to a perceived lack of good-quality comparative data about social and political attitudes (ESF 1999). The survey first went into the field in 2002 and five rounds of data collection have now been completed. From the start, the mission of the ESS was clear: to bring to cross-national survey research the highest standards of rigour associated with the best single-nation studies, and in doing so to provide data that would allow the monitoring of variations over time and between countries. In order to achieve this, it was decided that 'the hallmark of the ESS would be consistency across nations and exacting standards. Thus, familiar but inappropriate national variations in methodology were to be firmly resisted' (Jowell et al. 2007). Although fieldwork is devolved to experts in each participating country, there is strong adherence to a uniform set of principles and procedures set out in the Specification for Participating Countries (see e.g. European Social Survey (2011b). The ESS now runs in more than 30 European countries (see http://www.europeansocialsurvey.org).

Despite being a time series, the ESS is not longitudinal in the strict sense of following a panel of selected respondents. Rather, it is a collection of nationally representative cross-sections repeated every 2 years. This has advantages: it avoids the risk of 'putting all its eggs in one basket' by staking everything on achieving a representative sample at the start; it also sidesteps the problems of attrition faced by long-running panels and the distortions introduced by the arrival of new panel members when people re-partner. It also has disadvantages: a fresh sample has to be drawn in each round, and the frame from which this is taken may change between rounds due to availability, legal restrictions, a change of field agency or cost constraints. While the aim is to ensure the best sample design in each round, there may be a loss of comparability between rounds.

15.2.2 Questionnaire Development

The ESS questionnaire comprises both core and rotating modules, taking about 60 min to complete (in its source language of English—in translation it may diverge from this length). The core questionnaire was developed at the beginning of this century, based on proposals

from experts in social research and survey methodology (European Social Survey 2002). In addition to a long section measuring socio-demographic variables, there are three modules measuring media usage, interest in and engagement with politics, attitudes to immigration, satisfaction with life, evaluation of national performance and a range of enduring topics. Around 100 items are devoted to 'rotating modules'—usually two—which cover particular areas in more depth. Academics compete for the chance to help design these rotating modules in collaboration with a team of survey methodologists from the ESS's Core Scientific Team (CST). In addition to the core and rotating modules, there is a short supplementary questionnaire which fields the Schwartz scale of human values (21 items) plus a number of alternative formulations of earlier questions. These either have different wording or offer different numbers or types of response categories. The test items are used to compute the reliability and validity of survey questions (Saris and Gallhofer 2007a, b), and detailed guidelines are given for the translation and documentation of the questionnaire (see Harkness 2007).

15.2.3 Administration Mode

The ESS is administered face-to-face, unlike an increasing number of major surveys but similar to many cross-national surveys (see Sect. 15.1). There are two very strong justifications for this. First, the questionnaire is quite long and the issues it deals with are quite complex, making it less than ideal for administration over the telephone. Secondly, sampling frames for face-to-face interviewing—registers of individuals and households, or files of addresses—tend to have more complete coverage than any list of telephone numbers or e-mails. Opting for face-to-face mode meant that participating countries that had switched to telephone interviewing had to re-invent the practice of field interviewing in order to comply with the specification.

The choice of face-to-face surveys makes data collection in the ESS quite costly. One way to reduce these costs is to recruit sample units by telephone. This is allowed in the ESS in a small number of countries where telephone surveys are the predominant mode of data collection, individual sampling frames are available and the majority of the population is listed in the telephone book (Stoop et al. 2010).

15.2.4 Coverage, Sampling, Non-Response

Basic requirements with regard to coverage, sampling and non-response are set out in the Specification for Participating Countries (European Social Survey 2011b). This states that the survey must be representative of all persons aged 15 years and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language. Translations are required for each language used as a first language by 5 % or more of the population. This means that small language groups may not be able to take part in the survey.

The sample must be selected using strict random probability methods at every stage and respondents must be interviewed face-to-face. Different sampling frames, designs and procedures can be used across countries (see Sect. 15.3.2). An expert sampling panel helps each country to find the optimum solution in their case (see Häder and Lynn 2007), and signs off the final sample design. The minimum 'effective achieved sample size' should be 1,500, or 800 in countries with populations of less than 2 million. The effective sample size is the sample size based on simple random sampling (SRS). This means that the number of actual interviews will be larger in most countries.

According to the Specification, the proportion of noncontacts should not exceed 3 % of all sampled units, and the minimum target response rate should be 70 %. As shown earlier in Chap. 9, this figure is exceeded in some countries, but is not reached in many. Detailed guidelines for response enhancement (Koch et al. 2010), and feedback from the CST based on general experience and previous rounds can help countries to maximise their response rates.

Outcomes of all approaches to addresses, households and individuals in the sample must be defined and recorded according to a

pre-specified set of categories that distinguish between ineligibility, noncontacts, refusals, other contact (but no interview) and interview. Model contact forms are produced by the CST for different sampling frames (see Sect. 15.3.2).

15.2.5 Checklists and Monitoring

Each national funder selects its own fieldwork organisation, rather than entrusting the fieldwork to a single, multinational supplier. This makes it possible to select the organisation in every country that can best conduct the survey according to the specifications (Jowell 2007). As a consequence, in some countries the survey is conducted by the national statistical institute or a university, an option that is not feasible in every country. This brings the risk of remote management and a long chain of command that may breakdown.

Several steps have been taken to make sure that fieldwork is conducted according to the Specification for Participating Countries (European Social Survey 2011b). Firstly, prior to the commencement of the fieldwork, and ideally prior to signing a contract with the fieldwork organisation, a fieldwork checklist has to be completed by every participating country, ensuring maximum compliance with the standards. Dialogue around this process takes into account results—and possible deviations—from previous rounds. This checklist, like the sample design, has to be formally signed off. Secondly, countries have to provide fieldwork projections, based on the results from previous rounds. Thirdly, based on these fieldwork projections bi-weekly progress reports have to be sent to the CST to make sure that no unexpected problems occur. The progress of fieldwork can be derived from the detailed contact forms (see also Stoop et al. 2010).

Context and events

Comparing the results of cross-national surveys is fraught with traps, even when the process has been conducted according to rigorous methodological standards. Even the most experienced comparativist is unlikely to be an expert in the history and culture of more than a handful of countries. Even if we can be confident of the accuracy of the data, how can we explain the stubborn tendency of a particular country to buck a wider trend or a sudden spike or dip in a country's time series? The ESS addresses the first of these problems by gathering sources of data about the aggregate nature of a country—its GDP, the prevailing unemployment rate, its population structure—in order to account for otherwise unexplained 'country differences'.

The MacroDataGuide, which can be accessed at www.nsd.uib.no/macrodataguide, is an accommodating starting point for researchers and students looking for data on the environment in which individuals reside and act. The guide will jump-start your data search by providing you with an overview of extant data sources as well as an instant insight into their quality and comparability.



In regards to the second problem, one of the risks of the cross-sectional survey is that the 'snapshot' it produces may be taken at a moment that is atypical in some way. The social and political atmosphere prevailing during the fieldwork period may be unusual for that country, and this can affect the results. This risk is compounded in a survey like the ESS which, because of its stringent methodological standards, may be in the field for a period of some months. If the social climate is disrupted by a patch of unusual political weather, the atmosphere may be markedly different on either side of those events. Respondents interviewed later in the period may change their attitudes accordingly. In order to aid present and future analysts, the ESS records major events reported in the media during the fieldwork period,

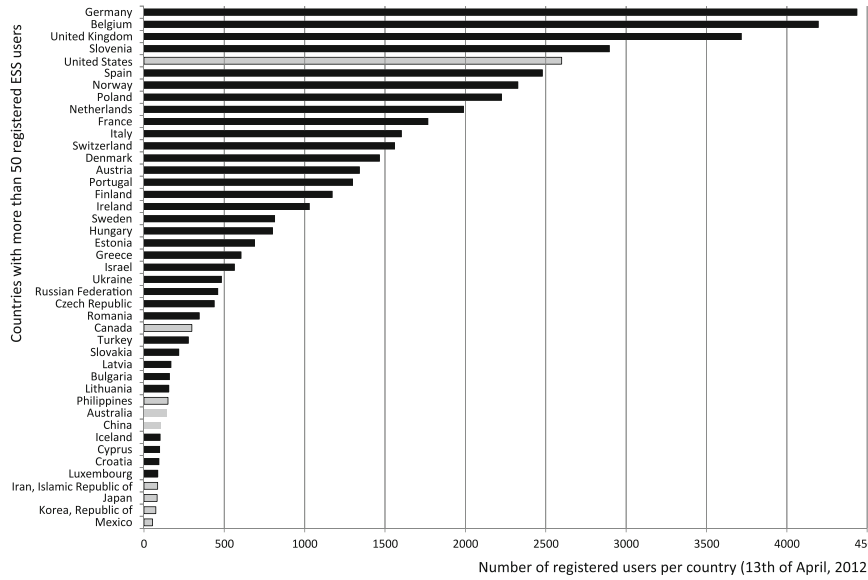


Fig. 15.3 Registered users of ESS data (total: 46.903; non-ESS countries in *grey*)

codes these into categories and indicates which areas of the questionnaire may be impacted and how (Stoop 2007).

15.2.6 Archiving and Data Access

The touchstone of the ESS is free and immediate access to data (Kolsrud et al. 2007). The archiving process is unusually rapid, with the first tranche of data from each round being released around 9 months after the first country has finished the fieldwork. The process is not just about speed; however, there is an ongoing dialogue between the archive and the surveyed countries about the quality of their data. The requirements pertaining to the depositing of the data and paradata are as rigorous as those governing their collection. Thanks to the free access and high quality of its data, ESS has developed into an increasingly popular data resource that now has around 40,000 registered users from all over the world (see Fig. 15.3).

15.2.7 Improvements and Discussions

The ESS operates within a continual quality enhancement loop, repeating steps from the Survey Quality Life Cycle (presented in Sect. 15.1.1) after each round of data collection. National coordinators are extensively briefed before every round and supported during fieldwork. At the end of the fieldwork period all participating countries feedback their experiences and the CST reports on the performance indicators from that round. Ongoing issues arising from fieldwork—such as response rates, the merits of attempts at refusal conversion, the quality of coding of demographic variables, compliance with fieldwork specifications—are reviewed and acted upon. After Round 4 there was considerable debate about the apparently increasing length of the questionnaire. This resulted in a reduction of 20 items in the length of the Round 5 questionnaire and the setting up of a committee to examine strategies for reducing the costs to countries and the burden on respondents.

Another persistent discussion concerns the mode of data collection. Although face-to-face administration is seen as the gold standard for the ESS, it is to be foreseen that at least some countries will have to rely on other data collection modes in the future. To test the consequences of this, a large number of experiments have now been conducted, (see e.g. Jäckle et al. 2010).

15.3 Representing the Target Population

15.3.1 Coverage and Exclusion

Groves (1989) starts his theoretical overview of populations with the *population of inference*, for instance Spanish citizens in 2010. The *target population* is the finite set of persons that will be studied in a survey. Persons who cannot be contacted or who will not be able to participate are generally excluded from the target population, for example, persons living abroad and those living in institutions (residential care homes, prisons). The *frame population* is the set of persons for whom some enumeration can be made prior to the selection of the survey sample. After the sample has been drawn from the frame population, ineligible units have to be re-moved, such as incorrect addresses or persons who are not Spanish citizens. Those who then respond to the survey are the *survey population*: the set of people who, if they had been selected to the survey, would be respondents. Unit non-response is the failure to collect data from units belonging to the frame population and selected to be in a sample. The response rate is the percentage of selected units who participate in the survey.

Excluding far-away regions

In face-to-face surveys travel costs can be a substantial part of the survey costs. For this reason, far-away regions are often excluded from the target population, for instance, those who live on small islands far from the coast. In the ESS, the inhabitants of what are termed the outermost regions which are part of the European Union are excluded: Guadeloupe, French Guiana, Réunion and Martinique (France), the Azores and Madeira (Portugal) and the Canary Islands, Ceuta and Melilla (Spain). The impact on survey outcomes is likely to be small, as these are small subpopulations, although the number of volunteers to conduct interviews in these exotic areas is usually large.

At every stage between *population of inference* and *survey population* groups drop out and operational and definitional problems have to be tackled. Excluding the homeless in advance means that they will not contribute to the non-coverage error when they are not in a sampling frame. Excluding residents of care homes right from the start means that they cannot contribute to the non-response error when they are not able to participate in a survey. In general, excluding difficult groups from the beginning may make life easier for those involved in fieldwork and will result in higher response rates, but will not make the survey any better. In assessing survey quality, the total survey process should be taken into account.

15.3.2 Sampling

One of the advantages of face-to-face-surveys is that many different sampling frames are possible. Population registers, electoral registers, household registers and address lists can all be used. Even where they exclude some groups in the target population, for instance, youngsters who are not yet on the electoral register, these lists can still be used provided additional steps are taken in the field to include people who are not on the list. Even area-based sampling can be used, where enumerators select buildings, households within buildings and persons within households.

The ESS (see also Sect. 15.2.3) uses simple random sampling, systematic random sampling, stratified unclustered random sampling and clustered random sampling (see Häder and Lynn 2007). As long as probability sample designs of comparable national populations are used, samples can be considered as equivalent in a cross-national study (Kish 1997). Nonetheless, the type of sample can have an impact on fieldwork and non-response. When census records are used, for instance, it can happen that a person has moved to another location when the interviewer comes by. In an address-based sample the target persons are those who are living at the particular address at the time the interviewer calls, and changes of address (including migration to another country) will not be a problem.

The different types of sampling frames used in the ESS are extensively documented in the National Technical Summaries (see e.g. European Social Survey 2011a). Table 15.1 gives an overview of sampling frames used in the second round of the ESS, taken from Häder and Lynn (2007).

Except in the case of individual sampling frames, where the person to be recruited is indicated in advance, interviewers can play a major role in the sampling for face-to-face surveys. Firstly, they may have to select addresses, if area sampling is used. Interviewers will, for example, be instructed to select the third address on the left and right from a starting address, or every 10th house in a street. In this case, interviewers may be prone to skip certain unattractive addresses (the 'mean dog' phenomenon) and select a nicer and more promising address. A good procedure in this case would be to let someone else do the enumerating, draft a list of fixed addresses and assign this to the interviewer.

When address registers are used, interviewers may have to randomly select a household at a particular address in the case of multi-household units that share the same address. This could be student flats, comprising many households or people living in a few apartments sharing the same postal address. In this case the interviewer needs someone to provide information on the number of households. Subsequently, one household has to be drawn at random, a procedure that is also prone to errors (and perhaps also to the likelihood of selecting the nice student who opened the door instead of coming back for the possibly not-so-nice student who is not at home). Luckily, multi-unit addresses where several households share the same address are relatively scarce.

More common, indeed quite common, is that the interviewer has to select one person in a household. This selection can be based on a Kish grid (a selection table giving each eligible household member an equal chance of being selected), the 'last birthday method' or any random scheme. In other data collection modes, it can be hard to make sure that the right person is selected and that the right person completes the questionnaire. In face-to-face surveys the interviewer has control over this procedure, which is a good thing. Here too, of course, the temptation might arise to select the person who opens the door instead of an absent household member.

Detailed contact forms are used in the ESS, distinguishing between individual, household and address samples. The selection procedure used for the target respondent also has to be recorded on these forms. This makes it possible to check—to a certain extent—whether random selection has taken place. Figure 15.4 describes the person selection in household samples in ESS Round 4, using the last birthday method (for an overview of all contact forms see www.europeansocialsurvey.org: Fieldwork Documentation). As a complete household roster is part of the survey, it is possible to check whether the right person was selected.

Interviewers can thus play an important role in sampling face-to-face surveys and can help make sure that random sampling takes place. Monitoring of this process will always be required.

15.3.3 Non-Response

In a face-to-face survey many steps can be taken to enhance response rates. Chap. 9, by Stoop on unit non-response due to refusal highlighted the importance of experienced, well-trained interviewers. We will return to the topic of interviewers in Sect. 15.4. Interviewer tasks depend on the nature of the sampling frame. In a sample of individuals they have to establish contact with the target person, establish whether he or she is eligible, find out if he or she is able and willing to participate in the survey and conduct the interview. In a household sampling frame, the interviewer also has to select the target person according to some kind of criterion.

An advance letter and leaflet sent to the sample unit just prior to the first visit are now generally considered good practice (Dillman et al. 2002; Groves and Couper 1998). Design features that may be salient to the respondent can be highlighted in an advance letter, and advance

Table 15.1 Sampling frames used in the second round of the ESS

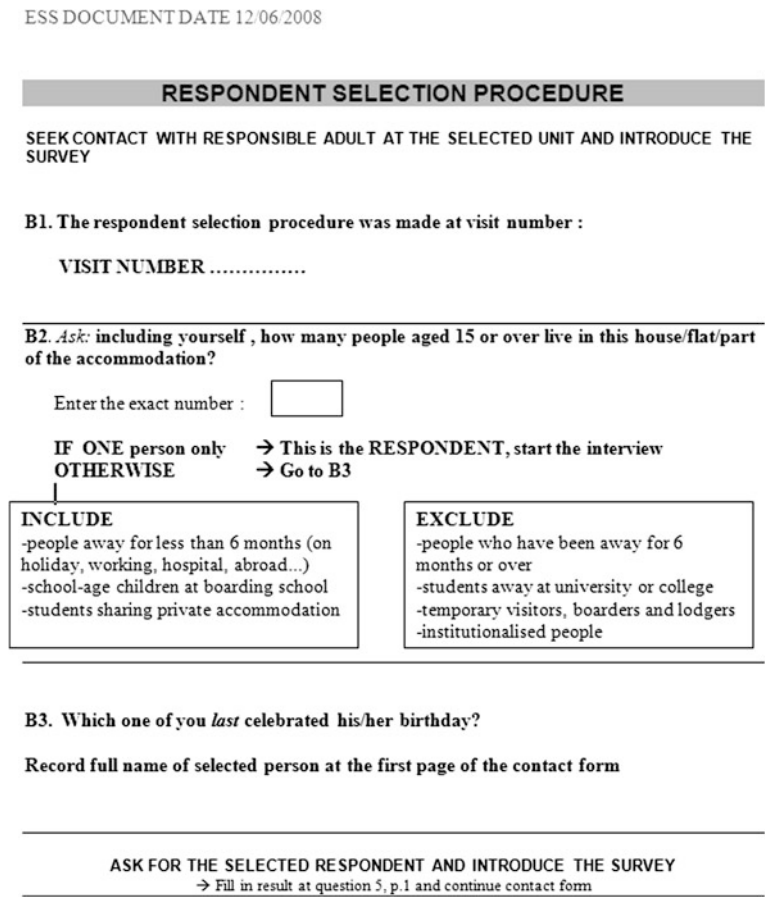
Country	Frame
Austria	Telephone book (additional non-telephone households sampled in the field)
Belgium	National register
Czech Republic	Address register (streets selected, followed by field enumeration)
Denmark	Danish central person register
Estonia	Population register
Finland	Population register
France	None (area-based sampling)
Germany	Registers from local residents' registration office
Greece	None (area-based sampling)
Hungary	Central registry
Iceland	National register
Ireland	National electoral register
Luxembourg	Social security register
Netherlands	Postal address list
Norway	Population register
Poland	National register of Citizens
Portugal	None (area-based sampling)
Slovakia	Central register of citizens
Slovenia	Central register of population
Spain	Continuous census
Sweden	Register of population
Switzerland	Telephone register
Turkey	Cluster of addresses
UK	Postal address list
Ukraine	None (area-based sampling)

Based on Häder and Lynn (2007, p. 39)

letters increase interviewers' confidence when seeking cooperation. Advance letters may also contain an incentive (see also Sect. 15.4.1).

When the interviewer visits the sample unit's address, the latter may be not at home, not able to participate or not willing to participate. People may be unable to participate because they do not speak the fielding language of the survey, or are mentally or physically unable to participate. The design of the survey greatly determines people's ability to participate. Translating questionnaires, adapting the questionnaire for the very elderly or those with limited mental capacities, rephrasing and simplifying questions, or coming back when an ill person has recovered, can substantially reduce the percentage of 'not able'. In many cases, not being able is a minor factor behind non-response (see also Chap. 9, by Stoop).

Non-response due to refusal is a major reason for non-cooperation in surveys. This factor is discussed extensively earlier in this handbook by Stoop in Chap. 9. What remains is non-response due to noncontact. A number of steps can be taken to minimise refusal due to noncontact in face-to-face surveys. One step is to require a sufficient number of calls, and to make sure these calls are spread over different hours of the day and days of the week. This call scheduling is of course much easier in telephone surveys, as in face-to-face surveys each visit can take a considerable amount of time. A second step is to allow sufficient time for the contacting process,



ESS DOCUMENT DATE 12/06/2008

RESPONDENT SELECTION PROCEDURE

SEEK CONTACT WITH RESPONSIBLE ADULT AT THE SELECTED UNIT AND INTRODUCE THE SURVEY

B1. The respondent selection procedure was made at visit number :

VISIT NUMBER .....

B2. *Ask:* **including yourself , how many people aged 15 or over live in this house/flat/part of the accommodation?**

Enter the exact number :

IF ONE person only → This is the RESPONDENT, start the interview
OTHERWISE → Go to B3

INCLUDE
-people away for less than 6 months (on holiday, working, hospital, abroad...)
-school-age children at boarding school
-students sharing private accommodation

EXCLUDE
-people who have been away for 6 months or over
-students away at university or college
-temporary visitors, boarders and lodgers
-institutionalised people

B3. Which one of you *last* celebrated his/her birthday?

Record full name of selected person at the first page of the contact form

ASK FOR THE SELECTED RESPONDENT AND INTRODUCE THE SURVEY
→ Fill in result at question 5, p.1 and continue contact form

Fig. 15.4 Extract ESS contact form round 4, household samples using last birthday method

in order to reach people who have been away for some time (e.g. on holiday).

To increase the contact rate interviewers can leave behind a card with a telephone number, or call the target person if they have a telephone number. In both cases there is a risk that the target person will refuse more easily than in a personal contact. Interviewers can also try to find out from neighbours when the sample units will be at home, or try to assess from the characteristics of the dwelling (yuppie apartment building, old-fashioned plants on the window sill, child's bicycle in the front garden) when the target persons are likely to be at home. Here a dilemma presents itself; on the one hand interviewer observations can help in establishing contact, but on the other hand sticking to prescribed call schedules may better facilitate the modelling of non-response and may prevent interviewers from calling only at times they themselves prefer.

Studies on the best way to schedule calls have mainly been based on telephone surveys, partly because call scheduling is much easier and partly because calling is much less expensive in telephone surveys. One advantage of studying contactability in face-to-face studies, however, is that additional information may be available on the households to be contacted, their dwelling and their neighbourhood. Analyses of noncontact in face-to-face studies and predictions of good times to call can be found in Groves and Couper (1998), Purdon et al. (1999), Stoop (2005), D'arrigo et al. (2009).

Many studies on contactability pertain to a single country. An exception is Stoop et al. (2010), where an overview of contactability in different countries is given. Figure 15.5 shows the contact rate in a number of European countries in the second round of the ESS. To the left of the origin is the contact rate, i.e., the percentage of

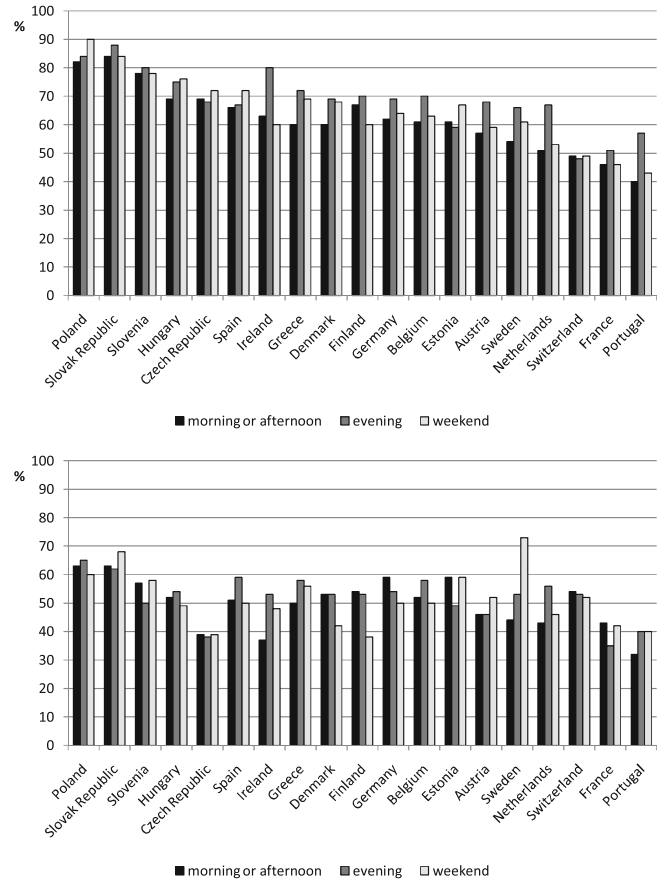


Fig. 15.5 Contact rate by time of the day, ESS Second Round. *Above* call 1, *below* call 2. Contact rate on morning or afternoon weekday, evenings weekday, and weekend days

visits that result in a contact, independent of the outcome) at the first call. To the right is the contact rate at the next call for the sample units who were not contacted at the first call. It is clear that the contact rate differs substantially across countries, that evening calls often, but not always, have the highest contact rate and that the pattern differs for call 1 and call 2. Sweden, for instance, had contact rates at call 1 of 54 % (morning/afternoon), 66 % (evening) and 61 % (weekend); and at call 2 of 44 % (morning/afternoon), 53 % (evening) and 73 % (weekend). In Slovakia, the differences between time slots were much smaller; here the contact rate at the first call was 84 % or more, regardless of the time slot. The figure shows that local circumstances have to be taken into account when prescribing or optimising call schedules in a face-to-face survey.

15.4 Representing the Concepts of Interest

15.4.1 Developing the Questionnaire

Many components of the survey process, such as sampling designs and strategies for minimising non-response, are aimed at ensuring that the data are representative of what could have been collected from the larger population. The questionnaire has a much more direct bearing on the nature of the data itself, however—not only the form in which it is captured, but also its quality. Data quality comes in two varieties: validity (the degree to which it has measured what it purports to measure) and reliability (the degree to which comparable results could be reproduced under similar conditions).

Saris and Gallhofer (2007b) describe questionnaire design in the ESS. A very comprehensive study on questionnaire design in general is given in Saris and Gallhofer (2007a). Harkness et al. (2003, 2010) focus on questionnaire design from a comparative point of view. Drawing on these sources, we can distinguish a number of steps in questionnaire design which, although not necessarily unique to the face-to-face mode, are particularly relevant for cross-national studies.

Literature and conceptual framework

Before questions can be drafted, the literature has to be reviewed and existing instruments compared (Harkness et al. 2003). A conceptual framework then has to be established, and the underlying concepts behind the questions have to be formulated. Theoretical concepts lead to latent constructs, which in turn lead to manifest indicators and finally to the survey questions (Harkness et al. 2010). Information on concepts, constructs, theories and expected relationships is carefully documented for evaluation and re-design in later stages.

Old or new questions

A next step is to decide whether to use existing material or to create new questions (see Harkness et al. 2003; Krosnick et al. 2010). Existing questions have often been tried and tested. When using existing questions, measurement characteristics are known, and survey results can be compared with other surveys. When existing questions are used, they can be *adopted*—taken literally, 'as they are'—or *adapted*. Tested scales, for instance in health research, are often used in the original formulation. It may be necessary, however, to adapt question texts to align them with other languages or other cultures. Adaptations may take four forms: factual (euro or dollar), language-driven (from gender–neutral English to gender-specific French), convention-driven (left–right versus right-left reading) and culture-driven.

> Frequently, adaptations are motivated less by features of the target language than by the need to fit social, cultural or other needs of the new linguistic group to be studied. Examples of adaptation not directly related to linguistic considerations abound. A recent international project proposed fielding the question, 'Can you lift a two liter bottle of water or soda', in multiple countries. The source question itself was not developed cross-culturally (…..). Several locations (countries) noted that (a) the normal size of bottle in their context was 1.5 litre bottles, and not 2 litre bottles, (b) that they were unsure whether the bottle referred to was intended to be glass or plastic (which would affect the lifting task), (c) that 'soda' was not a salient generic concept in their locations, and (d) that the formulation in English which indicates that the bottle is not empty ('bottle of water or soda') needed to become 'a full bottle of water' or 'a bottle full of water' in their translations. However, there was some concern that these much more explicit renderings of 'bottle of water' might alter respondent perceptions of the lifting task.
>
> Survey Research Centre (2010). Adaptation of Survey Instruments, by Janet Harkness.

Development strategy for multi-language questions

When new questions are to be created in a cross-national study, either because questions on a particular topic do not yet exist or are deemed to be of insufficient quality, three strategies can be followed: sequential, parallel and simultaneous. In the sequential approach the questionnaire is developed in one language and then translated into others. This is a simple approach which carries the clear risk that some concepts simply do not apply in other languages, cultures or countries. In a parallel approach, input to the source questionnaire from all target cultures and languages is acknowledged in the development stage. In the simultaneous approach, there is no one single source questionnaire, but different language versions are developed at the same time. The latter is clearly an impossible procedure in a multiple-language study such as the ESS, and here the parallel approach is used. Harkness et al. (2003, p. 25) give a clear overview of the advantages and disadvantages of the different strategies.

Ask the same question or different questions

In a comparative study another distinction can be made, namely between Ask the Same Question (ASQ) and Ask Different Questions (ADQ). It can be argued that different questions have to be asked to measure the same concept in different cultures. Most cross-national studies try to ask the same question, however, in order to pursue equivalence. This leads to long discussions when developing questions in different cultures (see the example above).

Considering whether to ask the same or a different question is also relevant when developing a mixed-mode survey, or a face-to-face survey that may be administered in other modes in the future. According to Dillman (2000), mode effects can be due to the fact that questions tend to be constructed differently for different types of questionnaires. In addition, complex surveys like the ESS make heavy use of showcards. Where these are complex and/or lengthy, they would be difficult to read out over the telephone (although a visual cue like a showcard can be adapted for a self-completion interview on the web). To minimise this problem, Dillman advocates the 'unimode' construction approach to questionnaire design for mixed-mode surveys, i.e., Ask the Same Question in each mode, but draft this question so that it is suitable for administration in all modes.

Item characteristics

When drafting individual survey items, many question characteristics have to be taken into account. There are a few clear 'don'ts' in this regard (no double-barrelled questions, no double negatives, no very long questions, no difficult words or concepts). Still, when '… scrutinised, almost every survey question is subject to criticism' (Schuman and Presser 1981, p.12). There are still many options from which to choose: open or closed questions, inclusion of mid-positions and explicit DK (don't know) options, the form and length of a response scale, the presence of an introduction, the order in which an item appears in a sequence, and so on. These issues and many more are treated in great detail by Saris and Gallhofer (2007b) who provide lots of clear and instructive examples.

Testing

When questions have been drafted, they have to be tested. There are a number of ways of gathering evidence about the performance of a question, producing either quantitative or qualitative data. The first looks at the form and structure of the question—its length, the number of complex words it contains, and so on—and then, based on what is known about the quality of items with similar features that have been fielded before, makes a prediction about how well the question will perform (Saris and Gallhofer 2007a). The Survey Quality Predictor program (SQP) reverses the customary warning about investment products. In this case, past performance *is* a guide to future success.

A second approach that is normally described using the loose term 'cognitive interviewing', uses human subjects to help predict likely problems with questions as they pass through the design process. Prototype items are administered to subjects, who are then encouraged to volunteer additional information by way of clarification. This normally involves one or both of the following techniques: respondents are encouraged to share their cognitive processes in 'think-alouds' as they construct their answer (a technique increasingly used on TV quiz shows to heighten audience engagement); with 'verbal probing' the interviewer waits for the answer, before asking for further detail or enquiring about how the answer was arrived at or what scenario they had in their mind while answering (Willis 2005). Cognitive interviewing can be extremely useful for identifying at an early stage questions that are badly worded, are confusing or are burdensome to respondents. In addition to cognitive interviewing, methods are increasingly being developed to include probing questions in a web survey (e.g. Behr et al. 2011).

A third—and usually the final—method of question testing is the large-scale pilot study, where whole batches of items will be tested on a sample large enough to include meaningful numbers of important subgroups. This is a particularly useful exercise in the context of face-to-face surveys, as questionnaire designers can also benefit by debriefing interviewers who will have recorded not only item non-response and any

queries from respondents, but will also have observed the typical time taken by respondents to answer particular items, whether due to the cognitive complexity, lack of salience of the topic or reluctance to answer.

The data from the pilot study should be analysed in detail to assess both the quality of the questions and the distribution of substantive answers and scale characteristics of question blocks that are expected to measure underlying concepts. 'Problematical questions, whether because they have displayed weak reliability or validity, deviant distributions or weak scales, are sent back to the drawing board. It is on the basis of these pilot analyses that the final source questionnaire is subsequently developed' (Saris and Gallhofer 2007b, p. 55).

Annotation

Once the questionnaire has been finalised in the source language, annotations are added for the translation. These annotations do not form part of the final questionnaire but serve as a guide to translators. Annotations can be used to explain specific concepts (government: 'The people now governing, the present regime') or fairly general ('Should' in the sense of 'ought to'; not in the sense of 'must') (ESS source questionnaire, Round 4).

Translation

The main challenge to cross-national equivalence is the ability of questions to travel into other languages and other cultural contexts. 'Tried and tested' questions from national surveys will frequently fail cross-nationally because they rely on a neat idiom from a single language, or use a term that has too specific a meaning within that national context to be understood in the same way across a large number of countries.

Many cross-national studies rely on a procedure called 'back-translation' to check that their survey translations are adequate. In its simplest form, this means that the translation which has been produced for a target language population is re-(or back-) translated into the source language. One major drawback of this procedure is that an identical or correct back-translation does not necessarily mean that the text in the target language is correct, understandable and covers the same concepts. 'It's raining cats and dogs' as original text and as back-translated text could still mean that the target language text would be 'Il pleut chats et chiens' (in French), which would not make much sense to French-speakers, who would expect 'Il pleut des grenouilles' ('It's raining frogs').

In the ESS the Translation, Review, Adjudication, Pretesting and Documentation (TRAPD) approach has been implemented, developed and promoted by Harkness and her colleagues (Harkness 2003, 2007). TRAPD consists of five interrelated procedures, performed by three key agents: translator, reviewer and adjudicator. The ESS calls for two translators per questionnaire, who translate questions independently and then take part in the subsequent review session(s). Reviewers not only need to have good translation skills, but also need to be familiar with the principles of questionnaire design and the particular study design and the topics covered. They do not produce either of the two draft translations, but attend the review sessions and contribute to and guide the revisions. The adjudicator is the person responsible for the final decisions about which translation options to adopt. Pretesting may again result in modifications before the adjudicator signs off on the version for final fielding. Central in TRAPD is the team approach and the detailed documentation required.

In the ESS, countries which share languages (UK and Ireland, Belgium and France, Austria, Germany and Switzerland) are also required to compare their draft translations in order to minimise unnecessary differences (Harkness 2003).

Translations in Switzerland

Four languages are spoken in Switzerland: German, French, Italian and Romansh. The ESS questionnaire is translated into German, French and Italian. This means that the Swiss translation team has to compare translations from English in three languages to check whether the same concepts are being measured, but also that the German translation has to be discussed with the Germans and Austrians, the French with the

Belgians and French and the Italian with the Italians.

The written German in Switzerland is Swiss Standard German. This Swiss Standard German is rarely spoken (in Parliament and in main news broadcasts). Spoken Swiss German, by contrast, comprises a range of Alemannic dialects (*Mundart*). For the German translation of the questionnaire, specific Swiss wordings and structures therefore have to be found to make it more suitable for a translation on the fly to the Swiss German dialects (including vocabulary and syntax adaptations). The written questionnaire will in practice be administered in different dialects.

Verification

Once a questionnaire has been translated, two types of verification strategies are implemented in the ESS. Firstly, the formal characteristics of each question can be checked by coding each item on the Survey Quality Prediction website (www.sqp.nl). In this way, deviations from the source version can easily be spotted. The second approach entails verification by human native-speakers (Dept 2010), who compare the source version and the translation version. Differences need not necessarily be errors, but they are documented and reported back to the translation team. Differences can occur in a variety of areas: added or missing information, layout and visual issues, grammar and syntax, within-unit and across-unit consistency (same word in source version, or measurement units in different places), register (casual, formal), wording, adaptation to local usage and mistranslation.

15.4.2 Survey Organisations and Interviewers Position

Face-to-face surveys have increasingly been replaced by telephone surveys, and at present online surveys are taking an increasing share of the market, which means that the number of survey organisations in a country with a well-trained interviewer staff can be very small. In the ESS every country selected the best possible fieldwork organisation, given the funds available. This means that the survey is carried out by statistical offices, universities, survey organisations and not-for-profit organisations. This may mean that some organisations have a fixed interviewer staff and some mainly employ free lance interviewers. Despite the emphasis on quality in the ESS, this may also mean that some organisations have to use some inexperienced interviewers to conduct the fieldwork (European Social Survey 2011a).

As a corollary to differences in interviewer position, the pay schemes of interviewers will vary. For instance, 30 countries participated in the fourth round of the ESS, and their practices can be summarised as follows (European Social Survey 2011a):

- Ten countries paid interviewers for completed interviews;
- Nine countries paid for completed interviews and offered some kind of bonus;
- Three countries used staff from a statistical institute who received an hourly wage;
- Eight countries used another payment scheme, which might, for instance, include a general assignment fee.

Bonuses were handed out for several reasons:

- Difficulty in obtaining response (interviews in difficult areas (big cities), hard-to-reach respondents, refusal conversion efforts);
- High individual response rates;
- Additional efforts (if addresses were contacted within 2 weeks after sending the advance letter);
- Extra bonuses for supervisors or for all interviewers, depending on the total response rate.

Why are these organisational and financial issues important? Firstly, because when selecting from among fieldwork organisations it is useful to have an overview of their track records. Secondly, because the characteristics and experience of organisations can have an impact on survey quality. There is evidence, for example, that the type of sponsor (university, government agency, market research organisation) has an effect on the response rate. In addition, organisations with little experience in face-to-face data

collection from a random sample of the population (quota sampling is not so rare in many countries) may underestimate the amount of time and number of interviewers required to conduct the target number of interviews. Thirdly, the payment scheme of interviewers may determine how they allocate their efforts. Interviewers who are paid per completed interview may tend to spend little time on unpromising cases, unless there is an extra bonus when they recruit these unpromising cases. Interviewers who are paid by the hour, on the other hand, may not mind spending a lot of time on unpromising cases but may not do this in the most efficient way.

Given the large difference between countries in the ESS, organisational characteristics and payment schemes do not predict final response rates very well. Statistical institutes that pay interviewers by the hour (Sweden, Norway, Finland) do achieve high response rates (see Chap. 9), but there could be many other factors behind this, and in any case the response rates in these countries are decreasing steadily; some other countries where statistical institutes are not involved now far outscore them.

15.4.3 Preparing Fieldwork

Once the questionnaire has been finalised and the survey organisation is in place, the preparation for the survey can start. First, the questionnaire itself has to be printed or programmed for CAPI, by no means a trivial endeavour (see e.g. Hansen and Lee 2010). A sample has to be drawn (Sect. 15.3.2), call schedules have to be developed (Sect. 15.3.3) and measures have to be taken to maximise the contact rate (Sect. 15.3.3) and minimise the refusal rate (Chap. 9). In addition, other documents have to be prepared, such as advance letters and contact forms (see Fig. 15.4 earlier in this chapter). Furthermore, decisions have to be taken about whether to use respondent incentives, and if so, what kind of incentives and when to offer them to respondents.

Two issues are, however, of particular importance for face-to-face studies, namely interviewer assignment sizes and interviewer briefing and training.

Assignment sizes

Detailed overviews of interviewers' effects on survey errors can be found in Biemer and Lyberg (2003), Groves et al. (2009) and Schaeffer et al. (2010). Interviewers can have an impact on non-response errors and measurement errors. Non-response errors are discussed earlier in Chap. 9. Measurement error in face-to-face surveys has already been touched upon in this chapter, when comparing survey modes. Here we will focus on interviewers' contribution to variance, i.e., the intraclass correlation due to interviewers.

It will be clear that the effect of this intraclass correlation will be larger if interviewers have larger assignment sizes. Biemer and Lyberg (2003, p. 166) give an example which shows that a survey among 6,000 persons that employs 100 interviewers has greater precision than a survey that employs 40 interviewers. In the first case, the average assignment size would be 60 in their example, in the second case 150. They also mention, however, that 'the cost of hiring, training, and supervising more interviewers may be such that the reduction in variance obtained by increasing the number of interviewers is not worth it'. In practice, it may also happen that experienced, successful interviewers have larger assignment sizes and are also deployed to take over difficult cases from less experienced interviewers or to carry out refusal conversion. Limiting their assignment sizes could thus well result in lower response rates.

The maximum assignment size in the ESS is 48. This means that no interviewer should work on more than 48 cases. In practice, this criterion is not always met. Using a weaker criterion of a maximum of 48 achieved interviews per country, it transpired that quite a number of countries exceeded this maximum. The number of countries in which at least one interviewer conducted more than 49 interviews was 12 (out of 22) in ESS Round 1, 15 (out of 26) in ESS Round 2, and 16 (out of 25) in ESS Round 3 (Koch et al. 2009, p. 237). The number of interviewers in the ESS differed greatly across countries as is shown in Fig. 15.6 (although sample sizes were of the same order of magnitude).

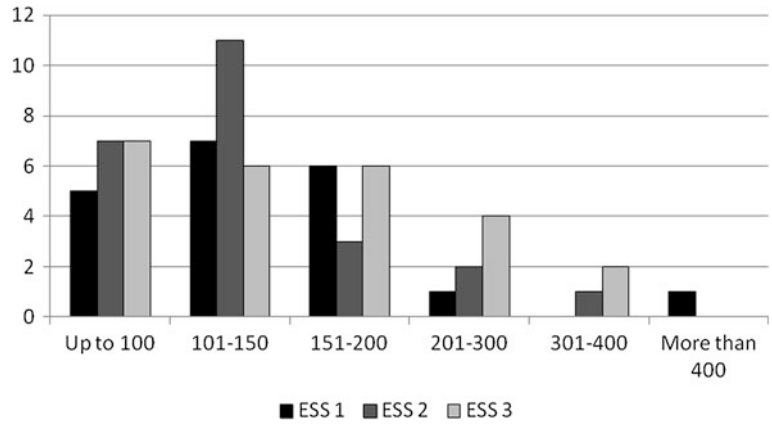


Fig. 15.6 Number of interviewers per country in Round 1, 2 and 3 of the ESS *Source* Stoop et al. (2010, p. 79)

Interviewer briefing and training

An interviewer briefing is used to give specific instructions for a particular survey. This may involve a description of the project, discussion of the questionnaire and practical arrangements. It is usually highly appreciated, and highly motivating, when the substantive researchers, who need the data for their research project, attend these briefings. They can convey information on the background and importance of the study to the interviewers and tell them how the data will be used.

Interviewer training is more general, and more task-oriented. Training is crucially important for interviewers to do a proper job (Groves et al. 2009). This is definitely true for new, inexperienced interviewers. However, additional and adapted training is also advisable for more experienced interviewers.

An illustration of on-line interviewer training can be found on the website of the Behavioral Risk Factors Surveillance Surveys: http://www.cdc.gov/brfss/training/interviewer. Detailed information and results on interviewer training is to be found in Lesser et al. (2008), Loosveldt (2008), Schaeffer et al. (2010).

Content of the interviewer training (from the Round 6 Specifications for Participating Countries (European Social Survey 2011b))

Survey researchers agree in general about the content of an interviewer training. Interviewers must *gain insight in the essential characteristics of a face-to-face interview* (e.g. similarity and difference of an interview with a conversation, awareness of the complexity of the interviewer-respondent interaction) and have a sound grasp of both positive and negative effects of interviewer behaviour on data quality (e.g. positive: interviewers can give direct support to the task performance of the respondent; negative: interviewers can have an impact on the respondent's answers).

To reduce systematic and variable interviewer related error interviewers must perform their task according to some *basic task rules* (e.g. reading questions as worded in the questionnaire, asking every question that applies to the respondent, using prompt cards, recording answers accurately, …). During the training, there is also special attention for *interviewing techniques* related to the manner in which the interviewer must deal with the respondent in order for the respondent to carry out his task well (e.g. giving general and specific instructions, using positive and negative feedback, clarifying questions, probing for more and adequate information …). A thorough discussion of *effective interviewer reaction behaviour to*

inadequate respondent behaviour is another essential part of the training (Loosveldt 2008).

An interviewer training is more than a theoretical presentation. *Practice* is at least as important. By means of discussions of tape and/or video recordings of various interview situations, role-playing with other trainees and conducting 'mock' interviews that are monitored and reviewed one can master by practice the content of the training. We advise to record a few interviews on tape and to thoroughly evaluate these interviews.

15.4.4 Conducting the Interview

There is an extensive body of literature on conducting interviews (see e.g., Loosveldt 2008). We will focus here on three issues: the general issue of standardised interviewing, the more specific topic of 'don't knows' (DKs) and the much less frequently covered topic of the situation during the interview.

Standardised interviewing

Standardised interviewing is usually advocated in face-to-face surveys, implying that results are replicable and that interviewers are interchangeable. Groves et al. (2009) mention three disadvantages of standardised interviewing. Firstly, different people exposed to the same words do not necessarily derive the same meaning from them. Secondly, a standardised interview is an unnatural situation, which may result in a cumbersome and redundant interaction; and thirdly, it can be problematic when the interviewer clearly sees that the respondent does not understand or misunderstands the question. Having to ask a respondent who has just run down the stairs to open the door for the interviewer, and then runs back up them ahead of the interviewer, whether he or she is able to climb stairs, seems rather silly. Answering 'Whatever it means to you' when the respondent asks 'What do you mean by income' may not be very helpful. An alternative would be a more conversational form of interviewing (Conrad and Schober 2000; Conrad and Schober 2006; Schaeffer et al. 2010). It may be that minor diversions from the script and extra efforts to transmit the meaning and requirements of the questions to the respondent result in more accurate (i.e. valid) responses. The de-professionalisation of the situation to create one of a more intimate conversation may actually result in the respondent going the extra distance to help the interviewer. This may lead to longer and deeper engagement with the task (and therefore less satisficing) and may cause the respondent to operate in a more 'confessional' mode in which they are more willing to divulge their true opinions and attitudes, shorn of inhibitions about how they would be received in society more widely.

In a large, comparative, cross-national study like the ESS involving thousands of interviewers and multiple languages, a more conversational approach would be rather risky, and it would be impossible afterwards to trace what interviewers really asked. In this case a standardised approach seems highly preferable (Loosveldt 2008). This will of course be easier when questions are easy to understand and relevant in all countries, languages and cultures in a cross-national study. This again means that extensive pretesting is required.

The principles of standardised interviewing have been described by Fowler and Mangione (1990, p. 350; cited by Schaeffer et al. 2010, p. 457): '1. Read questions as written. 2. Probe inadequate answers nondirectively. 3. Record answers without discretion. 4. Be interpersonally nonjudgmental regarding the substance of answers'. Given that, the principal tasks of the interviewer are to: help respondents perform their role correctly in the interview; manage the question and answer process; record the answers provided, possibly on a computer; and edit the answers and send the data back to the main office of the survey agency (Groves et al. 2009). Additionally (Loosveldt 2008), the interviewer has to provide both general and specific instructions to the respondent in order to make sure they understand their role. This professionalises what would otherwise be a basic conversation. In relation to all the elements of interviewer behaviour—reading the question,

clarifying the question, probing and recording answers, Loosveldt stresses the importance of 'playing it by the book'.

One particular area where the interviewer can play a role is in those cases where the respondent is unable or unwilling to provide a valid answer within the main categories or along the scale offered. The task of the interviewer is to minimise these occurrences. Having cajoled the respondent to take part in the exercise as a whole, they now have to repeat that persuasive process for each item. There remains some debate about the status of 'don't know' (DK) as a response. The layout of the response categories usually indicates the leanings of the question designers. If a DK option is visible, this suggests that they tend to regard it as a valid answer; if the designers wish to discourage do not knows, it is often absent from the show card but allotted a 'hidden code'. For some, DK is a valid response, suggesting that the respondent is still undecided as to the merits of an issue, or feels ambivalent towards it. This is itself data. The absence of a DK category can also be frustrating for a respondent, if they lack the factual knowledge to answer a question. The absence of a DK option may result in respondents breaking off from an online survey. For others, DK is seen as effectively missing data as the respondent is unable to provide the required information. For yet others, at least some, '… DKs often result not from genuine lack of opinions but rather from ambivalence, question ambiguity, satisficing, intimidation, and selfprotection. In each of these cases, there is something meaningful to be learned from pressing respondents to report their opinions, but DK response options discourage people from doing so. As a result, data quality does not improve when such options are explicitly included in questions' (Krosnick and Presser 2010, p. 284–285). Different attitudes towards DK among interviewers, different ways of handling DK in CAPI programmes (and in CAPI and PAPI) and different meanings or connotations of questions in different countries may therefore result in different rates of missing values across countries.

Understanding and interference

Face-to-face interviews are usually seen as interactions between interviewers and respondents. In practice, the presence of others, and other disturbing interferences, can have a large effect on the presenting and understanding of questions and the providing and recording of answers. The extensive attention being devoted to an issue like *social desirability,* begs the question of what the effect would be of the presence of a spouse when answering questions on marital happiness or on domestic violence for that matter. Studies on the effect of the presence of others are limited; one example is Smith (1997).

In the ESS the interviewer has to answer a number of questions about the interview at the end of the interview. These questions cover the difficulty of interviewing and the presence of and interference from others. Figure 15.7 shows the percentage of interviews in each country in the fourth round of the ESS in which someone else was present who interfered with the interview. We have not analysed the relationship between these data and the values of particular variables, but the large difference in the presence of others suggests that this may have an impact on cross-national comparability. Hungary, Romania and Israel especially stand out, as others were present and interfered in the interview in respectively 23, 27 and 32 % of cases in these countries.

Even more striking are the interviewers' answers when asked if respondents understood the questions, how often they asked for clarification and whether they tried to answer questions to the best of their ability. On the left-hand side of Fig. 15.8 we see those countries where interviewers spotted problems in 5 % or less of their interviews (Finland, Norway, Germany). At the other end of the spectrum, we see the Russian Federation and Ukraine, where interviewers felt that 54 % or even 71 % of the respondents tried to answer questions to the best of their ability now and then, almost never or never. In Romania and Turkey 28 or 36 % of the respondents understood the questions now and then, almost never or never. The requests for clarification also vary greatly by country.

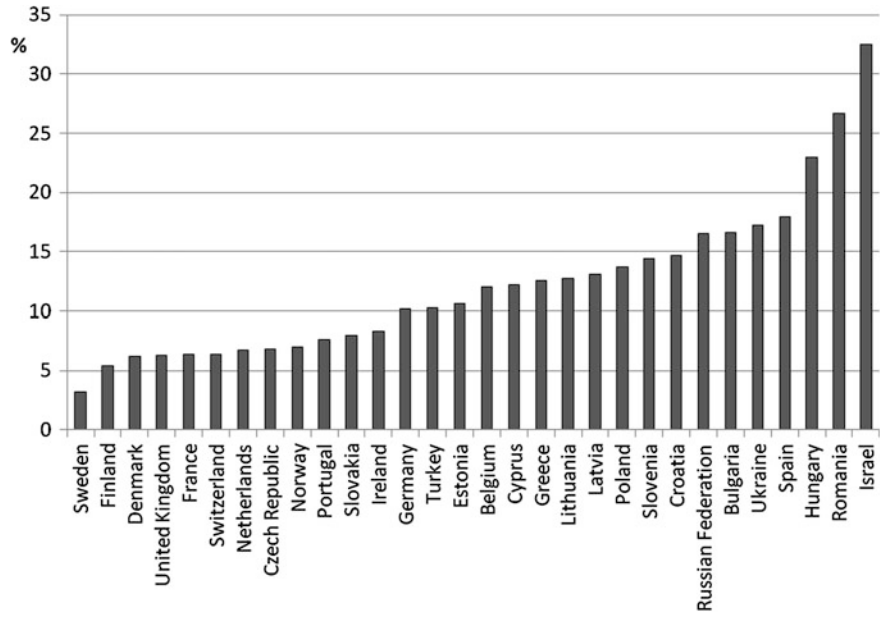


Fig. 15.7 Anyone present who interfered with the interview (ESS Round 4, unweighted)

We do not know whether these differences are real, reflect linguistic differences, interviewer expectations or other issues. It should be noted that, the correlation on a country level between the average assessments of the interviewer whether respondents tried to answer questions to their best ability and the average score on social trust of the respondents (on a scale from 'you can't be too careful' to 'most people can be trusted') is .54. The country level correlation between the average assessment of the interviewer whether the respondents understood the questions and social trust is even higher: 0.65. Figure 15.9 shows this relationship between social trust and the respondents' understanding the questionnaire according to the interviewers. Especially, striking is the difference between the high trust Scandinavian countries (Finland, Norway, Denmark and Sweden) in which interviewers assume that respondents understand all questions, and low trust Central European countries such as Turkey, Romania, Ukraine, Russia and Bulgaria in which interviewers cast doubt on the understanding of the respondents (Fig. 15.9).

So maybe these are to a large extent general cultural differences rather than reflections on the quality of the answers. Whatever the reason, it would be interesting to take the within-country differences into account when analysing the data.

Process, monitoring and control

Figures 15.7 and 15.8 make clear that there is more to 'interviewing' than interviewing. Interviewers can give feedback on the questionnaire when piloting, on the response process when establishing contact and interaction with the sample unit, during the interview when keeping the respondent involved, concentrated and motivated, and probing and clarifying when necessary, as well as after the interview when providing information on how the interview went.

Other tasks also have to be performed before, during and after the interview. Interviewers have to be monitored to make sure that all sample units are approached in time, that they are contacted a sufficient number of times and that sample units are used. Back-checks have to be performed to ascertain from respondents and nonrespondents whether the interviewer actually called. In fact, interviewer fraud (Biemer and Lyberg 2003; Groves et al. 2009) is difficult to exclude.

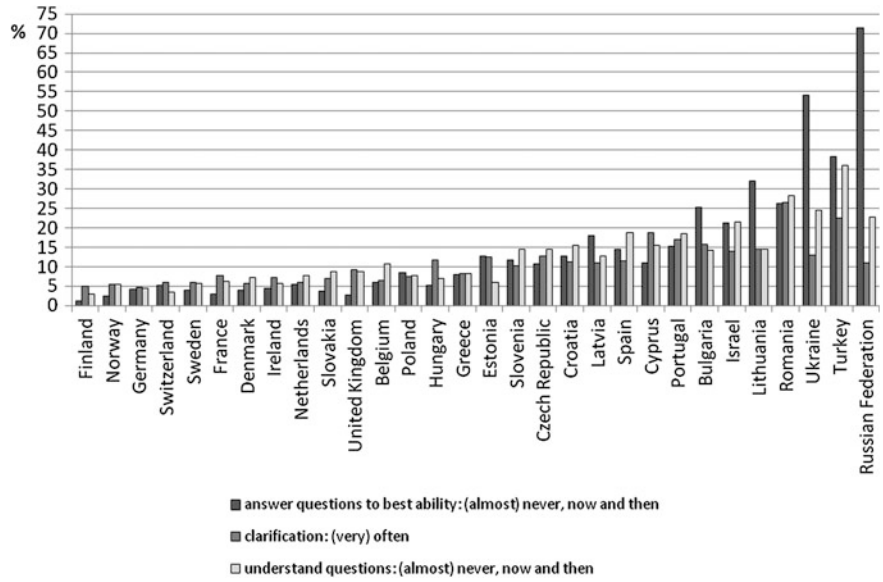


Fig. 15.8 Respondent understanding and effort in ESS Round 4 (unweighted)

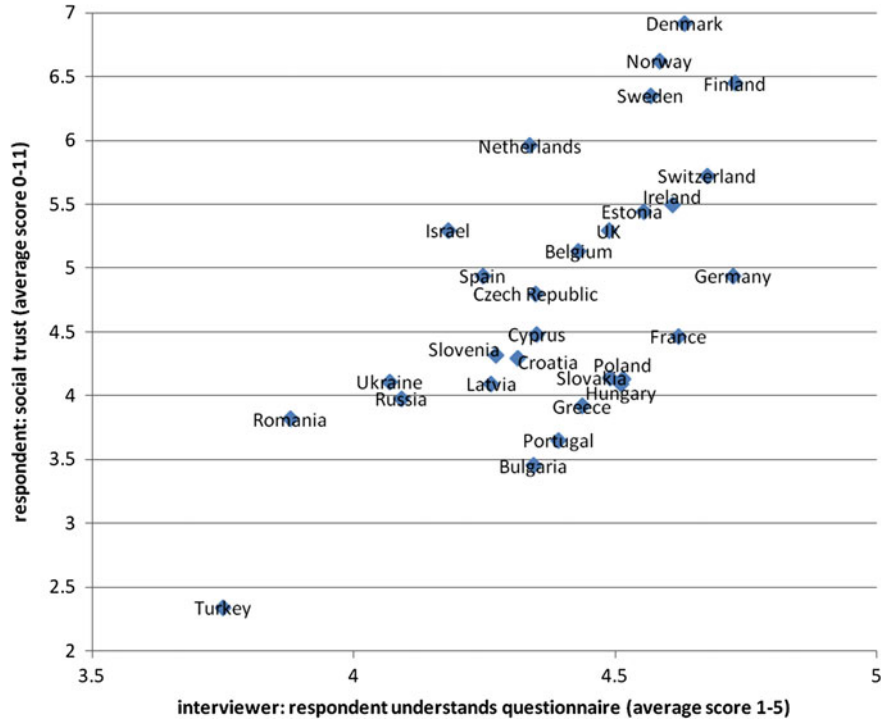


Fig. 15.9 Interviewer assessment of respondent understanding the questionnaire (unweighted) and social trust of respondents (weighted) in ESS Round 4

15.5 The Future of Face-To-Face Surveys

When thinking about the future of face-to-face-surveys we can follow three strands of thought. Firstly, the future of a particular face-to-face survey such as the ESS can be discussed. Here we again have to bring up the Survey Quality Life Cycle, which implies that a survey design will have to be adapted and improved over time. Adapted, because survey circumstances may change instruments or strategies that functioned well in the past but may not function well in the future. The default response to a survey request could become a 'No' (see Chap. 9), respondents may get used to incentives, fixed landlines may disappear, survey scarcity may turn into a survey overload and questions may become outdated or irrelevant. This means that survey design never stops, and that the design continually has to be evaluated and possibly adapted. This also means that each time a trade-off has to be made between comparability and consistency on the one hand and optimum design on the other.

Survey designs will also have to be improved over time because evidence from previous rounds may show that initial assumptions do not always hold or that the design or implementation has deficiencies, research results may show that a design can be improved or new instruments may become available. From this perspective, adaptation means that the environment has changed and we have to adapt and improvement means that we discovered new ways to achieve our own goals.

A second strand in thinking about the future of face-to-face surveys is to move (partly) to a different mode of data collection. Indeed, this is already happening in many surveys in many countries, where lack of funds and the growth of the online population have led from a move away from face-to-face surveys to other survey modes or mixed-mode data collection. Deploying other modes will of course have an effect on the recruitment phase of a survey and on the interview phase.

A third strand encompasses efforts to maintain the face-to-face character whilst eliminating interviewer travel or even getting rid of interviewers themselves. This is what could really be the face-to-face survey of the future. The first option could be achieved through some kind of video connection (or Skype), where the interaction between interviewer and respondent takes place via a computer screen (Anderson 2008). The second option goes a step further; here, an interviewer conducts the interview as an avatar in a 'Second Life' setting (Bell et al. 2008; Dean et al. 2009). And finally, the interview could be conducted by a robot on a computer screen (Conrad et al. 2011; Malakhoff and Jans 2011).

When evaluating these options, we have to reconsider what really makes a face-to-face interview (Sect. 15.1.3; also see Neuman, Chap. 14). How do the three factors distinguished by De Leeuw (2008) (interviewer effects, media-related factors, and factors influencing information transmission) operate when the interviewer is not physically present or when there is no real interviewer at all? It might be expected that Skype-based interviewing comes closest to real-life face-to-face interviewing, as interviewer and respondent can see one another and real interaction takes place. Also, individual interviewer behaviour, characteristics and strategies can play the same role as in face-to-face surveys, which can again result in interviewer effects.

An avatar in a Second Life setting seems an interesting option, but very little research has been carried out on this, which suggests that it is not a feasible option. The robot option (or Embodied Conversational Agent (ECA)) (Malakhoff and Jans 2011) seems most promising. The ECA can see and hear the respondent through a webcam and microphone and can react to body posture and spoken language. There are indications that respondents feel the same social pressure of the ECA as they do when a real interviewer is present. Conrad et al. (2011), for instance, showed that respondents preferred an interviewer of their own race. As in a 'real' interview, the robot can guide respondents through a standardised interview, probe and clarify. They might even be better at this than a real interviewer.

On the other hand, talking to your screen is not the same as talking to an interviewer. It is not clear whether a robot would be the same locus of control as a real interviewer, nor whether a robot can convey the same sincerity of purpose and communicate the same trustworthiness. Information transmission will also change when one is talking to a robot on a computer screen, if only because many people would feel very uncomfortable doing this in the presence of others.

Still, given the fact that interviewers can play an important role in recruiting respondents and navigating them safely through long and complex interviews, it seems logical that more research on video-interviews and robot interviews should be encouraged.

References

Anderson, A. H. (2008). Video-Mediated Interactions and Surveys. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95–118). Hoboken, NJ: Wiley.

Behr, D., Kaczmirek, L., Bandilla, W., & Braun, M. (2011). *Testing the validity of gender ideology items by implementing probing questions in web surveys*. Paper presented at the 5th MESS Workshop, Oisterwijk, The Netherlands.

Bell, M. W., Castronova, E., & Wagner, G. G. (2008). Virtual assisted self interviewing (VASI): An expansion of survey data collection methods to virtual worlds by means of VDCI. Berlin, University of Technology, DIW & Max Planck Institute for Human Development.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: Wiley.

Conrad, F. G., Schober, M. F., & Nielsen, D. (2011). *Race-of-Virtual-Interviewer Effects*. Paper presented at the 5th MESS Workshop, Oisterwijk, The Netherlands.

Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and non-use of clarification features in web surveys. *Journal of Official Statistics, 22*(2), 245–269.

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*, 1–28.

D'arrigo, J., Durrant, G. B., & Steele, F. (2009). *Using field process data to predict best times of contact, conditioning on household and interviewer influences*. Working Paper M09-12. University of Southampton.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233–255.

De Leeuw, E. D. (2008). Choosing the Method of Data Collection. In E. D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 113–135). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

De Leeuw, E. D., Hox, J. J., & Dillman, D. A (Eds.). (2008a). *International handbook of survey methodology*. New York: Lawrence Erlbaum Associates.

De Leeuw, E., Hox, J., & Scherpenzeel, A. (2011). *Emulating interviewers in an online survey: experimental manipulation of 'do-not-know' over the phone and on the web*. JSM Proceedings, Survey Research Methods Section (6305–6314). Alexandria, VA: American Statistical Association.

Dean, E., Murphy, J., & Cook, S. (2009). *Social presence in virtual world surveys*. International Society for Presence Research. Proceedings of the 12th Annual International Workshop on Presence.

Dept, S., Ferrari, A., & Wäyrynen, L. (2010). Developments in Translation Verification Procedures in Three Multilingual Assessments: a Plea for an Integrated Translation and Adaptation Monitoring Tool. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph Mohler (Eds.), *Cross-cultural survey methods* . (pp. 157–173). New York: Wiley.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley Co.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little R. J. A. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 3–26). New York: Wiley

ESF (1999) *The European Social Survey (ESS):A research instrument for the social sciences in Europe*. Report prepared for the Standing Committee for the Social Sciences (SCSS) of the European Science Foundation. Strasbourg, European Science Foundation.

European Social Survey (2002) Core questionnaire development. www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=62&Itemid=404 (downloaded 8 August, 2011).

European Social Survey (2011a) *ESS4-2008 Documentation Report*, Edition 4.0, The ESS Archive, Norwegian Social Science Data Services, Bergen.

European Social Survey (2011b) *Round 6 specification for participating countries*. Centre for Comparative Social Surveys, City University, London. Available at: www.europeansocialsurvey.org.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing minimizing interviewer-related error*. Newbury Park: Sage.

Groves, R. M., (1989). *Survey Errors and Survey Costs*. New York: Wiley, Inc.

Groves, R. M., Fowler, F.J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology, 2nd ed.* Hoboken, NJ: Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Häder, S., & Lynn, P. (2007). How representative can a multi-nation survey be? In R. Jowell, C. Roberts, R.

Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey* (33–52). London: Sage.

Hansen, S. -E., & Lee, H. J. (2010). Instrument Technical Design. In: Survey Research Center (2010) *Guidelines for Best Practice in Cross-Cultural Surveys* (Chapter IX). Ann Arbor, MI, Survey Research Center, Institute for Social Research, University of Michigan. Retrieved August, 8 2011, from www.ccsg.isr.umich.edu.

Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph Mohler (Eds.), *Cross-cultural survey methods (35–56)*. New York: Wiley.

Harkness, J. A. (2007). Improving the comparability of translations. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey* (79-93). London: Sage.

Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. Ph, et al. (Eds.). (2010a). *Survey methods in multinational, multiregional, and multicultural contexts (33–58)*. Hoboken, NJ: Wiley.

Harkness, J. A., Edwards, B., Hansen, S.-E., Millar, D. R., & Vilar, A. (2010b). Designing Questionnaires for Multipopulation Research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 33–58). Hoboken, NJ: Wiley.

Harkness, J., van de Vijver, F. J. R., & Johnson, T. (2003). Questionnaire Design in Comparative Research. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph Mohler (Eds.), *Cross-cultural survey methods (19–34)*. New York: Wiley.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population. differences in response quality. *Public Opinion Quarterly, 72*, 836–846.

Holbrook, A. L., Green, M., & Krosnick, J. A. (2003). Telephone vs. face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*, 79–125.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review, 78*(1), 3–20.

Jowell, R., Kaase, M., Fitzgerald, R., & Eva, G. (2007). The European social survey as a measurement model. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey* (1–31). London: Sage.

Kish, L. (1997). 'Foreword' In T. Lê and V. Verma, *DHS analytical reports No. 3: an analysis of sample designs and sampling errors of the demographic and health surveys*, Calverton, Maryland USA: Macro International Inc.

Koch, A., Blom, A. G., Stoop, I., & Kappelhof, J. (2009). Data collection quality assurance in cross-national surveys: the example of the ESS. Methoden Daten Analysen. *Zeitschrift für Empirische Sozialforschung*, Jahrgang 3. *Heft, 2*, 219–247.

Koch, A., Fitzgerald, R., Stoop, I., & Widdop, S. (2010). *Field procedures in the European social survey round 5: enhancing response rates*. Mannheim: European Social Survey, GESIS.

Kolsrud, K., Kallgraf Skjåk, K., & Henrichsen, B. (2007). Free and immediate access to data. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey* (139–156). London: Sage.

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (263–314). Emerald Group Publishing Ltd, 2nd Ed.

Krosnick, J. A., Narayan, S. S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in survey research* (pp. 29–44). San Francisco: Jossey-Bass.

Lesser, J., Eyerman, J., & Wang, K. (2008a). Interviewer Training. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 442–460). New York: Lawrence Erlbaum Associates.

Loosveldt, G. (2008a). Face-to-Face Interviews. In E. D de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 201–220). New York: Lawrence Erlbaum Associates.

Malakhoff, L. A. & Jans, M. (2011). Toward usage of avatar interviewers in web surveys. *Survey practice*, June. www.surveypractice.org.

Marsden P., & Wright, J. (Eds.). (2010). *Handbook of survey research*. Emerald Group Publishing Ltd, 2nd Ed.

Peytchev, A., Conrad, F. G., Couper, M. P., & Tourangeau, R. (2010). Increasing respondents' use of definitions in web surveys. *Journal of Official Statistics, 26*(4), 633–650.

Purdon, S., Campanelli, P., & Sturgis, P. (1999). Interviewers' calling strategies on face-to-face interview survey. *Journal of Official Statistics, 15*(2), 199–216.

Roberts, C. (2007) *Mixing modes of data collection in surveys: A methodological review*. ESRC National Centre for Research Methods. NCRM Methods Review Papers NCRM/008.

Saris, W. E., & Gallhofer, I. N. (2007a). *Design, evaluation, and analysis of questionnaires for survey research wiley series in survey methodology*. New York: Wiley.

Saris, W. E., & Gallhofer, I. N. (2007b). *Can questions travel successfully?*. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey* (53–77). London: Sage.

Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In: P. Marsden & J. Wright (Eds.), *Handbook of Survey Research* (437–470). Emerald Group Publishing Ltd, 2nd Ed.

Schober, M. F., & Conrad, F. G. (2008). Survey Interviews and New Communication Technologies. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning*

the survey interview of the future (pp. 1–30). Hoboken, NJ: Wiley.
Schuman, H., & Presser Stanley, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* Academic Press; 1996. Reprinted, with new Preface, New York: Sage.
Smith, T. W. (1997). The impact of the presence of others in a respondent's answers to questions. *International Journal of Public Opinion Research, 9*(1), 33–47.
Stoop, I. (2005). *The hunt for the last respondent*. The Hague: Social and Cultural Planning Office.
Stoop, I. (2007). If it bleeds, it leads: the impact of media-reported events. In: R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.),*Measuring attitudes cross-nationally – lessons from the European Social Survey* (95–111). London: Sage.
Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving survey response. Lessons learned from the European social survey*. Chichester:Wiley.
Survey Research Center (2010) *guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI, Survey Research Center, Institute for Social Research, University of Michigan. Retrieved August, 8 2011, from www.ccsg.isr.umich.edu.
West, B., & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance. *Public Opinion Quarterly, 74*(5), 1004–1026.
Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Further reading: *Survey Methodology and Quality*

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: Wiley.
De Leeuw, E. D., Hox, J. J., & Dillman, D. A. (Eds.). (2008). *International handbook of survey methodology*. New York: Lawrence Erlbaum Associates.
Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology, 2nd ed.* Hoboken, NJ: Wiley.
Marsden P., & Wright, J. (Eds.). (2010). *Handbook of survey research*. Emerald Group Publishing Ltd, 2nd Ed.

Questionnaire design

Harkness, J., van de Vijver, F. J. R., & Johnson, T. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph Mohler (Eds.), *Cross-cultural survey methods (19–34)*. New York: Wiley.
Harkness, J. A., Edwards, B., Hansen, S. E., Millar, D. R., & Vilar, A. (2010). Designing questionnaires for multi population research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 33–58). Hoboken, NJ: Wiley.
Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New York, Wiley: Wiley Series in Survey Methodology.

Interviewers

Lesser, J., Eyerman, J., & Wang, K. (2008). Interviewer training. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 442–460). New York: Lawrence Erlbaum Associates.
Loosveldt, G. (2008). Face-to-face interviews. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 201–220). New York: Lawrence Erlbaum Associates.
Morton-Williams, J. (1993). *Interviewer approaches*. Aldershot: Dartmouth Publishing.
Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (437–470). Emerald Group Publishing Ltd, 2nd Ed.

Cross-national surveys

Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. Ph. (Eds.). (2003). Cross-cultural survey methods. New York: Wiley.
Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. Ph (Eds.). (2010). *Survey methods in multinational, multiregional, and multicultural contexts (33–58)*. Hoboken, NJ: Wiley.
Survey Research Center (2010) *Guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI, Survey Research Center, Institute for Social Research, University of Michigan. Retrieved August, 8 2011, from www.ccsg.isr.umich.edu.

Costs and Errors in Fixed and Mobile Phone Surveys

16

Vasja Vehovar, Ana Slavec, and Nejc Berzelak

16.1 Methodological Issues of Telephone Surveys

16.1.1 Telephone Surveys and Telephone Use Trends

Telephone surveying remains one of the most popular modes to collect data for probability sample surveys, despite fast developments of newer and cheaper technologies (i.e. web surveys). One of the main advantages of telephone surveys is the relative simplicity of procedures to obtain probability samples of general population, with still acceptable response, and coverage rates.

When the first telephone surveys were systematically performed in the 1930s in the USA only 30% of population had a phone at home. The mode started gaining value with the increasing coverage rates in the 1960s that further exceeded 90% in the beginning of the 1970s. The detailed methodological studies examined the quality of telephone interviews and found no substantial error due to the mode effect and/or noncoverage (Groves and Kahn 1979). Another stimulus for the rising popularity of the mode was the development of the two-stage Waksberg–Mitofsky sampling procedure (Waksberg 1978), which enabled effective selection of random digit dialing (RDD) probability samples. In the 1980s, telephone surveying came to its full swing in the developed countries and become a replacement technology for personal interviewing, especially in marketing but also in academic, and official surveys. The breakthrough was enhanced with the development of computer assisted telephone interviewing (CATI), where computer-assisted questionnaires allowed smoother, faster, and less error-prone telephone data collection.

In Europe, the diffusion of telephone interviewing occurred later and was more specific. Many countries have very complicated telephone number systems, which makes it difficult to perform RDD and telephone surveys, so corresponding probability sampling procedures were developed only in the late 1990s, for instance in Germany (Gabler and Häder 1999) and United Kingdom (Nicolaas et al. 2000). On the other hand, Eastern European countries initially had problems with lower telephone coverage among the general population.

With the second generation of mobile phones in the 1990s the use of this technology started to increase and become a part of daily lives. In the 2000s, mobile telephony was sufficiently widespread to be considered by the survey industry as an appropriate data collection medium. With growing mobile phone penetration (as discussed further in this section) telephone surveying can

V. Vehovar (✉) · A. Slavec · N. Berzelak
University of Ljubljana Faculty of Social Sciences, Kardeljeva ploscad 5, 1000 Ljubljana, Slovenija
e-mail: Vasja.vehovar@fdv.uni-lj.si

A. Slavec
e-mail: ana.slavec@fdv.uni-lj.si

N. Berzelak
e-mail: nejc.berzelak@fdv.uni-lj.si

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_16, © Springer Science+Business Media New York 2012

now be performed on mobile phones (mobile CATI). The use of mobile phones offers also several other advantages. Text messages can be used for recruiting, inviting, and otherwise contacting respondents (Bosnjak et al. 2008; Steeh et al. 2007) for follow-up contacts (Virtanen et al. 2007), as well as a tool for quality control (Kuusela et al. 2007). SMS, WAP (Wireless Application Protocol) and mobile Internet possibilities allow collection of survey responses by the means of computerized self-administered questionnaires (i.e. mobile CSAQ), where respondents answer the questions presented on the screen of their mobile devices by themselves, without interviewer involvement (Fuchs 2007).

Although in Europe and USA virtually every adult has access to a telephone at home, there are various telephone use patterns that vary across countries and are rapidly changing in time. While the mobile phone penetration is steadily growing, the fixed phone coverage—on which telephone survey methodology was predominantly based for decades—is decreasing. The fixed and mobile phone coverage rates have been analyzed by the US National Health Interview Study (NHIS) and the Eurobarometer (EB) E-communication household survey (Box 16.1).

Box 16.1: Studies on household telephone coverage

The National Health Interview Survey (NHIS) has been observing the increasing percentage of mobile-only households in the US since 2003. At the end of 2005, there were less than 8% of adults living in mobile-only households which grew to almost 12% at the end of 2006 and to about 14.5% in 2007. The highest increase was in 2008 at the end of which 18.5% of adults and 20% of households were mobile-only (Blumberg and Luke 2009). The PEW studies show similar rates (Brick et al. 2006).

In the European Union, the telephone coverage rates have been studied by the 2006 Eurobarometer E-communication household survey. The research indicated that more than 95% EU households have at least one phone service: either fixed mobile or both (European Commission 2008a). Each type of access is typical for a certain group of countries. In total, 60% of European households have both fixed and mobile telephone access. The largest proportions (more than 80%) are to be found in Sweden, Luxembourg, Malta and the Netherlands, while Austria and Bulgaria have the lowest figure (about 40%). There are nine EU member states where the proportion of households with only mobile phones is higher than households with both phones and fixed phone only. The highest rate (more than 60%) is found in Czech Republic and Finland, followed by Lithuania (53%), Hungary, Portugal, Slovakia, Latvia, Estonia, and Austria (38%). Since mobile telephony is gaining more and more ground, households with only fixed phone are becoming rare. A relatively high percentage of fixed-only are Bulgaria, Germany, and France. While the number of households having fixed telephone access has stabilized and the percentage of dual-access households has declined (for 3 percentage points), the percentage of mobile-only households has increased (for 2 percentage points) since 2006. The increase was particularly intense (10 or more percentage points) in seven countries: Portugal, Finland, Italy, Austria, Czech Republic, Hungary, and Poland (European Commission 2008a).

The basic trend is that the mobile-only segment is slowly increasing, which creates a need for including mobile telephone sample in any representative survey. The figure below shows the case of Slovenia, which is a very typical country (Fig. 16.1).

The table below shows the structure of individual telephone users in the European Union (27 countries) in late spring 2009 according to the Eurobarometer Survey (Table 16.1).

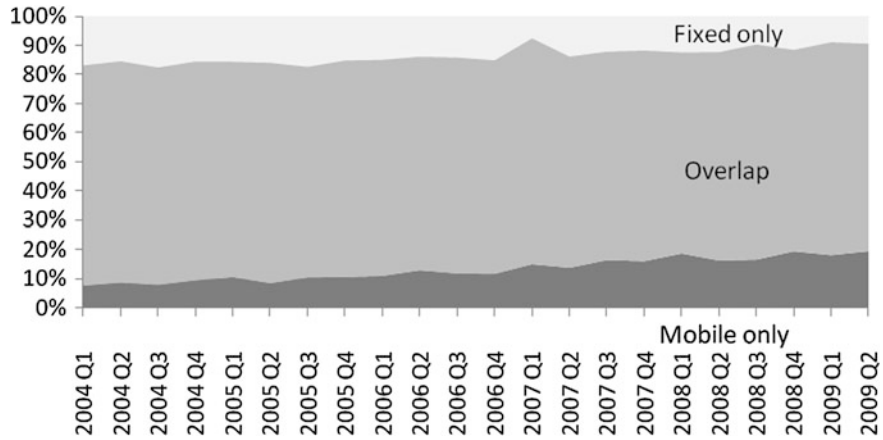


Fig. 16.1 The structure of telephone users in Slovenia (Labor Force survey 2004–2010)

Table 16.1 Mobile only population share (Eurobarometer May–June 2009)

Country	Fixed only	Both	Mobile only
Austria	10,1%	38,0%	51,9%
Belgium	8,8	59,1	32,0
Bulgaria	14,3	46,0	39,7
Cyprus	8,5	77,5	14,1
Czech Republic	3,0	18,9	78,0
Denmark	6,4	63,1	30,4
Estonia	6,9	47,2	46,0
Finland	2,5	30,1	67,4
France	14,6	72,0	13,4
Germany	15,7	77,1	7,2
Greece	17,4	64,6	18,0
Hungary	11,5	37,3	51,2
Ireland	5,9	66,3	27,8
Italy	7,0	58,3	34,7
Latvia	5,3	39,8	54,9
Lituania	6,5	36,4	57,1
Luxembourg	7,7	86,7	5,6
Malta	13,3	85,4	1,3
Netherlands	7,8	83,2	9,0
Poland	14,2	45,0	40,8
Portugal	11,3	42,6	46,1
Romania	16,1	34,4	49,5
Slovakia	5,7	38,6	55,7
Slovenia	8,2	76,8	15,0
Spain	20,2	46,3	33,5
Sweden	4,3	95,1	0,6
United Kingdom	12,2	71,8	15,9

16.1.2 Nonresponse Problem

With the rise of information–communication technologies and so-called postmodern societies, nonresponse is becoming one of the key issues in survey research. The growing volume of interpersonal communication in contemporary societies is additionally increased by uninvited personal communications of direct marketing and survey research industry. Message overload substantially decreases survey cooperation rates. These are affected also by a number of other features of information society: increased mobility, individualization, fragmentation, increasing complexity, family change, growing risks, privacy concerns, and so on.

This situation forces survey researchers to identify and utilize effective measures for increasing cooperation in surveys. These measures include improved presentations of research aims in introductory contacts (differentiation from telemarketing, sponsor reputation), refusal conversions, better monitoring of interviewers, and incentives.

The studies comparing response rates of mobile and fixed CATI have often provided contradicting results, from lower rates in mobile CATI (Steeh and Piekarski 2008; Link et al. 2007; Vehovar et al. 2004), through comparable results (Pew Research Center 2008) to higher rates in mobile CATI (Gallup Europe 2009). Moreover, the calculation of mobile survey response rates needs to be adjusted to different disposition codes (Callegaro et al. 2007).

In general, therefore, it seems there are only little differences in response rates between mobile and fixed telephones. Of course, both suffer from low cooperation. Today, only rare official surveys can achieve responses rates above 50%, typically when prenotification letters are used. On the other hand, even the best commercial brands (Tortora 2004) have to accept lower and decreasing response rates, while general telephone surveys (e.g. opinion polling) are typically below 20% or, more realistically, below 10%. Nevertheless, telephone surveys still seem to provide relatively reliable public opinion polling estimates, which are usually well confirmed at elections. The situation in Slovenia is perhaps very typical. Statistical office would receive response rate above 60%, academic research around 40%, the best commercial brands 10–20%, while others are below this benchmark. The situation is very similar also in other developed countries. Of course, noncoverage rate must be added to nonresponse rates, so to obtain the net reach.

According to Singer (2006), the research on survey nonresponse has since the mid1980s gone through three main phases: (1) study of the factors and comparisons, (2) investigation of field-work procedures, and (3) dealing with the relation of nonresponse rate and nonresponse bias. The last phase of research shows that the relation between nonresponse rates and bias is not straightforward (Groves and Peytcheva 2008), i.e. increased response rate does not necessarily result in decreased bias as it is often assumed. We therefore believe that the next, fourth stage of nonresponse research should focus on observing nonresponse within the complete environment of survey errors and costs. It need to be taken into account that cost savings can fund increased solicitation efforts and incentives that can compensate for the response bias arising from lower response rates. Survey costs are discussed later in this chapter (Sect. 2.4).

16.1.3 Specifics of Mobile Phone Surveys

The increasing number of households not having a fixed phone—but having at least one mobile phone—reduces fixed phone coverage, forcing a search for alternatives in telephone survey methodology. One idea is to incorporate mobile phones into survey designs, so CATI using mobile phones (i.e. mobile CATI) is becoming an increasingly frequent option complementing or even replacing fixed telephone interviews.

However, including mobile phones in survey designs is not an easy task. A pilot mobile phone survey by Vehovar et al. (2004) indicated that specific noncoverage and nonresponse problems

arise, but also that they are much more expensive than fixed telephone surveys (see Box 16.2). Furthermore, cell phone feasibility studies have reported to suffer from various methodological issues: sampling frames, un-listed numbers, pre-paid mobile phones, distinguishing mobile and fixed numbers, prefixes not necessarily indicative of geographic location, weighting procedures, measurement errors, possible mode effects, technological problems, lack of standardization, legal restrictions, and the personal nature of mobile phone device (Brick et al. 2007; Callegaro et al. 2007; Kuusela et al. 2007; Lavrakas et al. 2007). As regards the mode effects, on the other hand, as mentioned, little evidence of differences exist, so fixed and mobile CATI can be treated as one mode (McCutcheon 2009).

Box 16.2: Costs of mobile phone surveys

One of the most inconvenient characteristic of mobile phone interviewing is increased costs, especially due to expensive per minute charges. Based on telephone tariffs in Slovenia, the estimated cost per unit of mobile interviews is 3 times higher than for fixed units, but according to some studies it can be even up to 6 times higher (Belak 2007). Mobile CATI costs more than fixed CATI also in other European countries. For instance, in Finland the costs increased for more than 40% when they applied mobile phone interviewing for half of the sample (Kuusela and Simpanen 2002). Even in the USA where almost no difference between fixed and mobile phone tariffs exists, mobile surveys are more expensive (Link et al. 2007). In a PEW study, Kennedy (2007) roughly estimates that the price of conducting a mobile sample (including a 10$ incentive) is approximately 2.4 times higher than conducting a fixed telephone sample. Of course, with digitalization, we can expect increasingly smaller cost differences, at least for the part arising from telecommunication fees.

Mobile phone interviewing is thus believed to be a reliable data collection mode, at least when undertaking a proper methodological approach. In the future we can expect eliminated, or at least minimized, current country-specific barriers which today bring considerable complexity for the full implementation of mobile phones into the survey data collection process. This particularly relates to potential legal restrictions for random generation and selection of mobile phone numbers and corresponding technical verification of the eligibility of selected number (which is performed without knowledge and consent of the respondent).

Due to the above reasons, particularly due to the costs, as for now, the use of mobile CATI is typically limited to mixed mode settings which we explore in the next session.

16.2 Telephone Surveys in Mixed-Mode Designs

Mixed-mode surveys are one of the most promising answers of modern survey methodology to the key problems, mainly related to lowering response rates, and growing research costs. The use of several modes of solicitation and surveying at different stages of a survey project enables to reduce weaknesses of each individual mode and at the same time preserve its advantages. This is usually done in order to improve response and coverage rates to a degree that could not be achieved using one surveying mode only.

Within this context, the telephone—either fixed or mobile—plays an essential and indispensible role. In fact, telephone interviewing is the only mode, besides personal interviewing, where respondents are faced with live interaction with the interviewer, which provides dramatically higher response rates compared to self-administration. The telephone surveys are thus the component, which cannot be missed in almost any probability survey of the general population, because the only alternative to telephone interviews in mixed mode survey settings are face-to-face surveys, which are—as mentioned both by Neuman Chap. 14 and Stoop and Harrison

Chap. 15—painfully expensive and slow, and probability panels, which can give better response quality than telephone surveys (Chang and Krosnick 2009) but are in a very early stage and not ready for operational usage yet.

The telephone surveys are thus often the most essential component of mixed modes designs, of course, as long as we focus on probably based survey designs, which are the core scope of this chapter. Thus, to understand the modern role of telephone surveys, we need to understand the contemporary development of mixed mode surveys.

In order to discuss the range of possibilities offered in mixing survey modes, it is first necessary to distinguish survey solicitation from data collection activities (Vehovar and Lozar Manfreda 2008). Solicitation (i.e. invitation) to the survey, as a separated activity, can be performed together and with the same mode as the corresponding data collection (e.g. fixed telephone surveys, face-to-face surveys), or, increasingly often, with a different mode.

Telephone surveys can be introduced into mixed mode settings in various ways. For example, a mail prenotification to a telephone survey is a specific type of mixed mode where solicitation is performed using different mode than actual data collection. Furthermore, nonrespondents to a telephone survey can be interviewed face-to-face. Another frequently used option include telephone follow-up to web survey nonrespondents. Even combination of fixed and mobile telephone survey within the same survey project can be regarded as a mixed-mode approach, since there are several specifics related to each of them. Of course, numerous other possibilities are possible. These different mixed-mode designs are presented in the next section of this chapter.

The key problem of mixed-mode surveys is that different survey modes can cause differences in respondents' answers to otherwise equivalent questions (de Leeuw 2005). Each mode has its own specifics related to the medium of communication and the way of information transmission, factors that can cause mode-specific errors (de Leeuw 1992; Dillman et al. 2009). Such mode effect can produce additional differences among groups of survey respondents.

16.2.1 Mixed-Mode Surveys Classification

De Leeuw (2005) classified mixed-mode designs within three stages of a survey project: initial contacting, data collection, and follow-up of nonrespondents. In this section we present somewhat extended classification based on her work. It should be noted that this classification does not cover all possibilities that might be regarded as a mixed mode but focuses on the most common and feasible ones.

Initial contact:

1. Mode of survey prenotification differs from the mode of data collection. *Example: Mail prenotification to a telephone survey is sent to the target sample.*
2. Mode of solicitation, screening questionnaire, or selection of target persons differs from the mode of data collection. *Example*: *Screening telephone survey is used to identify web users and invite them to complete a questionnaire on the web.*

Data collection stage:

1. Different modes for different units from the same sample in one time point and with the same questionnaire. These mixed-mode designs can be further divided into:

a. Optional, where respondents can select the preferred mode by themselves. *Example: Respondents are offered to complete the questionnaire by telephone or on the web.*
b. Concurrent, where a researcher selects the most appropriate mode for individual respondent. *Example: those not having fixed-line telephone are surveyed by mobile phone, others by fixed.*
c. Sequential, where different modes are used at different stages of a survey project. *Example: nonrespondents to a mail survey are contacted and surveyed by telephone.*

2. Different modes for different parts of a questionnaire for all units from the same sample in one time point. *Example: Sensitive questions in a face-to-face survey are answered by a respondent's input of answer into interviewer's computer.*

3. Different modes in different time points for all units from the same sample (panel). *Example: Initial survey in a panel is performed face-to-face while all subsequent surveys on panel members are done by telephone*.
4. Different modes in different time points on different samples. *Example: A yearly performed survey (like* European social survey*) switches from face-to-face to telephone mode at one time point*.

Follow-up stage:

Mode of reminders to nonrespondents differs from the mode of data collection. *Example*: *Nonrespondents to a web survey are reminded by telephone to complete the survey on the web*.

One specific mixed mode type that should be added to the above classification by de Leeuw (2005) in the context of mobile telephone surveys is the use of different sampling frames within the same survey project. A part of the sample, for example, can be obtained from the telephone register of fixed-phone subscribers and another part by RDD on mobile phone numbers. Such dual frame designs are discussed further in the chapter.

As mentioned above, there might be many other possible mixed-mode designs, for example, different prenotification methods for different sample units, different modes of the screening questionnaire for different sample units, different modes of reminders for different sample units, and so on.

16.2.2 Response and Coverage Improvement

One of the key goals of using mixed-mode survey designs is often to increase coverage and response rates for a specific survey project. The former can be achieved because multiple modes allow reaching different groups of a target population. Furthermore, as proposed by Dillman et al. (2009), sending reminders in different modes and switching data collection mode for nonrespondents can be an effective way of nonresponse conversion. Another strategy of response rate improvement is incentives. Several studies suggested that these are by far most effective when provided in advance (Dillman et al. 2009), which can in telephone surveys be done, for example, together with a mail prenotification about the survey.

It should be, however, noted that various strategies of response rate improvement need to be carefully evaluated prior to their use. For example, mode effects when multiple survey modes are used can negatively impact the measurement quality of the obtained data even when response rates actually improve. Also it is necessary to keep in mind that nonresponse bias does not necessarily decrease with higher response rates as have been show by various meta-analysis (e.g. Gallup Europe 2009; Groves and Peytcheva 2008). While the exact relation between response rates and bias is usually unknown; however, conceptually we may think of three most frequent options: (A) Nonresponse bias decreases with the increasing response rate; (B) nonresponse remains unchanged regardless of the higher response rates until nearly all units have responded; (C) the nonresponse bias may even increase with increasing response rates. The latter situation was empirically shown by Vehovar et al. (2010) where the use of incentives significantly increased response rates in a mixed-mode survey, but at the cost of obtaining more and more specific segments of the target population thus resulting in substantially increased nonresponse bias.

16.2.3 Data Quality

The quality of survey data depends on many factors, including the definition of a target population, sampling procedure, definition and measurement of constructs, data collection processes, and data preparation. All these can impact survey data quality indicators such as precision, accuracy, validity, and reliability. Even further, they affect survey quality in a broader perspective which includes timeliness, clarity, comparability, coherence, and completeness of data (Biemer and Lyberg 2003). Fulfilling all these criteria requires a high level of methodological and statistical knowledge, as well as optimization of resources (time, personnel, and finances). Quality assurance is now

becoming increasingly difficult due to a declining willingness to participate in surveys and increasing survey costs.

A rich and various literature on how to approach and structure survey errors exist (Biemer and Lyberg 2003; Groves 1989, 2005; Groves et al. 2004; Lessler and Kalsbeek 1992; Weisberg 2005), with separation of errors of observation from nonobservation, random errors from biases, and sampling from nonsampling errors. Nonsampling errors are further decomposed into those related to specification, frame, nonresponse, measurement, and processing.

All these errors together contribute to the total survey error (TSE), which is the most commonly estimated *mean squared error* (Hansen et al. 1953; Biemer 2010). In this chapter we only focus on accuracy measured by the inverse of MSE, which is the most elaborated and essential component of survey error (Groves 1989). MSE expresses the expected magnitudes of difference between estimates and true population values. These differences may have systematic sources (consistent effects over replications of a particular survey design) and are referred to as *biases*, or random sources which are referred to as *variance* (Groves et al. 2004). Thus, the total survey error of a particular survey item takes the simple standard form summing the variance and the squared bias (Hansen et al. 1953; Deming 1950; Kish 1965). In the case of estimating the population mean, the MSE is expressed as:

$$\mathrm{MSE}(\bar{y}) = \mathrm{Var}(\bar{y}) + \mathrm{Bias}^2(\bar{y})$$

The bias component is defined as the difference between the expected value of the estimate and the true population value $\bar{Y}$:

$$\mathrm{Bias}(\bar{y}) = E(\bar{y}) + (\bar{Y})$$

While the population mean is the most frequently estimated parameter in surveys, these equations can be used to calculate MSE of any other parameter we are estimating.

There are various practical problems associated with using mean square error. First, the MSE can be calculated for one survey item at the time and it may behave differently for different items. Second, the true population value—needed to estimate bias—is rarely known and is usually estimated based on some auxiliary data (e.g. population registers, "gold standard" face-to-face surveys) with unknown accuracy. Third, MSE is often a combined result of different sources of error (sampling error, noncoverage, nonresponse, misspecification, measurement, and processing error) which are difficult to distinguish and to separate random errors from different components of bias. This is an issue especially when comparing alternative survey designs with different modes of data collection. Specific and complex experimental designs are needed to identify and separate the effect of mode of data collection from other error sources (Groves 1989).

On the other hand, taking into account different error sources is one of the key benefits of estimating survey data quality with MSE and is superior to approaches that focus on one or few sources only (e.g. comparing only sampling errors, response rates, or biases).

16.2.4 Cost Issues

Survey costs are important because savings can be used to offer incentives and improve solicitation activities. In contrast to survey errors, *survey costs* are much less elaborated in the literature (see Box 16.3 for an example). While numerous methodological nuances related to nonresponse, question wording and survey mode have been extensively discussed in survey methodology, hardly any systematic study about costs can be found, despite their crucial importance for establishment of the methodological approach, or a specific survey. Groves (2006) explains such lack of attention with the gaps in the understanding of production process, cost data, and organizational culture.

Box 16.3: Comparison of telephone and mail survey costs

One of the rare examples of studies examining survey costs was done by Link et al. (2007) who compared costs of telephone and mail surveys and considered

various costs for both modes: actual unit costs for material and supplies (based on pilot study), production statistics, and estimates of industry averages for direct hourly rates and indirect cost rates. Only costs assumed to be invariant to survey design were omitted (overall project management, survey design development, postdata collection weighting, and analysis). Despite higher costs of materials for mail surveys, their total cost were assessed 12% lower than telephone surveys. In fact, telephone surveys are much more labor intensive for the same number of completed interviews.

Another reason is the unfeasibility of establishing a universal cost model that would apply to most survey situations, organizations, and environments. It is difficult to provide suitable data input for complex cost models and such modeling usually brings little added value compared to simpler solutions (Kish 1965; Karr and Last 2006).

Nevertheless, a general cost model to compare different survey designs has been conceptualized by Vehovar et al. (2010). The model requires conceptual separation of survey solicitation and data collection activities. Solicitation activities are often repeated several times before actual data collection is performed, for example, several contact attempts are made or several reminders to nonrespondents sent. As noted in the discussion of mixed-mode surveys above, these activities are also becoming increasingly based on different modes (like mail prenotification to telephone survey). The conceptual separation is especially important in complex mixed-mode survey projects where numerous combinations of approaches to solicitation and data collection are possible.

In general, the costs can be estimated by summing fixed, solicitation and data collection costs (Vehovar et al. 2010):

$$C = c_0 + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(c_{0_{k,m}} + c_{k,m} \times n_{k,m}\right) + \sum_{k=1}^{K}\sum_{m=1}^{M}\left(a_{0_{k,m}} + a_{k,m} \times r_{k,m}\right)$$

There are up to K solicitation and data collection waves and within each of them there are M possible modes. Apart from general *fixed costs* (c_0), each wave and each mode within this wave may have its own additional fixed costs ($c_{0k,m}$, $a_{0k,m}$) that are independent from the number of units assigned at that level. We assume the fixed costs constant for a specific design but only for moderate changes in sample size, while radical increases in certain components of fixed costs are treated as different designs with different parameters so that the above equations can be still used. In addition, for each mode m within a certain wave k there are also specific *variable costs* for each unit in the process ($c_{k,m}$, $a_{k,m}$). The approach is illustrated by the example in Box 16.4.

Box 16.4: Cost calculation

Consider the initial sample of $n_1 = 1{,}000$ individuals. In the first wave ($k = 1$), all individuals are invited by mail ($m = 1$ for solicitation, so we have $n_{11} = 1000$) to complete a mail questionnaire ($m = 1$ also for data collection). The costs of the first solicitation wave equal $c_{11} * 1000$. If the response rate for the first wave is 30% ($x_{11} = 0.3$, $r_{11} = 300$), the costs of the data collection for this wave is $a_{11} * 300$. If having also 100 ineligible units identified in the first wave ($e_{11} = 0.1$, $u_{11} = 100$) we have $n_2 = n^*_{11} = (1-0.3-0.1) * 1000 = 600$ units to be allocated across modes in the second wave of solicitation.

The same logic can be applied to all subsequent waves and modes, even for complex mixed-modes design. For instance, the first wave (i.e. initial contact) can be a mail invitation to complete either a web or a paper questionnaire (two modes offered); in the second wave nonrespondents from the first wave can be contacted and interviewed by phone (either fixed or mobile); while in the third wave the remaining nonrespondents are personally visited by an interviewer.

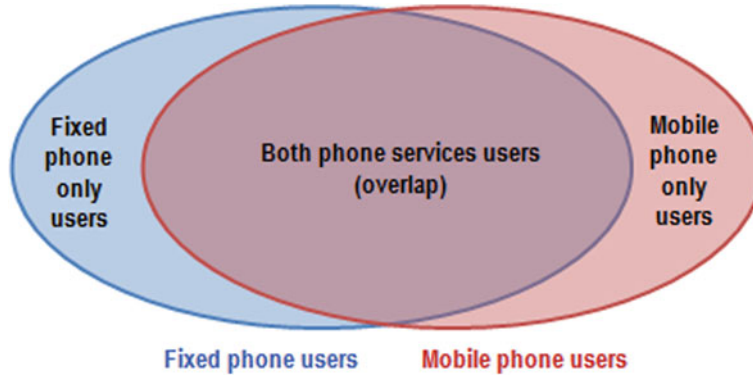


Fig. 16.2 Dual frame telephone survey (Vehovar and Slavec 2011)

In practical use of this approach, arbitrary estimates are needed for survey costs, especially in interviewer-administered survey modes when there is no clear separation between solicitation and data collection costs as both are conducted by the interviewer contacting a respondent. In telephone surveys, for example, the first minute of conversation is usually arbitrarily allocated to the solicitation (Vehovar et al. 2010) and the remainder to the data collection stage. Although the approach seems complex, experience shows that the calculation and separation of the costs of a particular survey project is practically feasible. In fact, the required data and estimates are already in use in the regular practice of research organizations (see earlier in this chapter Box 16.2). A cost benefit discussion is presented in a later chapter by Shin and Dulisse (Chap. 21).

16.2.5 Dual Frames of Fixed and Mobile Phones

One special type of mixed-mode surveys also rely on *multiple frame surveys*, which use two or more sampling frames (i.e. lists) to identify the elements of the target population. Such designs are useful for sampling rare or hard-to-reach populations and particularly for situations where a single frame does not provide sufficient coverage. Another reason for multiple frame designs can be cost savings over a single-frame design with comparable accuracy.

However, there are specific methodological problems associated with using multiple frame designs, in particular how to compensate for the overlap—units that belong to more than one frame—and how to optimally allocate sample size across frames. These issues are being researched in multiple frame literature that has emerged in three waves. The literature on multiple frame studies first emerged in the 1960s when researchers combined less costly but incomplete frames (e.g. telephone surveys) and expensive but more complete frames (e.g. area-sample frames with personal interviews) to compensate for drawbacks of single-frame survey types. Hartley (1962) was the first to systematically examine multiple frame estimation techniques. In the 1980 and 1990s, scholars further extended Hartley's method to improve combinations of telephone and personal visit surveys by determining the best allocation of sample size between the two frames, (e.g. Groves and Lepkowski 1982; Lepkowski and groves 1984; Lepkowski and Groves 1986; Bankier 1986; Traugott et al. 1987). Finally, recent multiple frame studies are related to issues raised by the diffusion of mobile phones (see Brick et al. 2006; Kennedy 2007; Keeter et al. 2007; Brick et al. 2011).

We thus have here *fixed phone numbers frame* and the *mobile phone numbers frame*. Each sampling frame has technical and administrative specifics that heavily depend on national environment. However, in almost all countries it is somehow possible to draw probability samples from these two sampling frames (Fig. 16.2).

In order to compensate for the duplication effect of the overlap, two design approaches are possible. First approach is to perform *screening*

of the mobile telephone sample before data collection to exclude those with both services and limit interviewing to mobile-only units, while for the fixed telephone sample no screening is needed. The second approach relies on appropriate *weighting* of the mobile subsample before merging it with the fixed telephone subsample.

By analyzing nonresponse bias in dual frame samples of mobile and fixed telephone numbers Brick et al. (2006) found out that mobile-only households are more likely to respond to mobile telephone surveys than households with both services. The units in the overlap accessed by mobile telephone might thus be abandoned without the interview (Brick et al. 2011), as they are in principle reachable also by fixed telephone. This may create a small specific additional coverage bias, but several studies show that it is not substantial (Brick et al. 2006; Blumberg and Luke 2009) or that it can be eliminated with proper weighting (Keeter et al. 2007).

Also Kennedy (2007) evaluated the effects of screening for telephone service in dual frame RDD surveys. Although not affecting coverage properties of a dual frame survey, screening could affect nonresponse, and other sources of error. Nevertheless, if data are properly weighted (based on telephone service), the effect of screening on bias and variance is minimal. Further evaluation of mobile-only screening should consider the relative costs of fixed telephone and mobile interviews and screen-outs in the mobile sample. Since the results showed no difference in accuracy between dual frame (with or without screener) and fixed telephone sample-only designs adding a mobile frame did not seem worthwhile (at least when the mobile-only share is below 20%, as in those studies).

Brick et al. (2011) examined nonsampling errors in dual frame telephone surveys, which were usually ignored in multiple frame literature. It was found that nonresponse and measurement error bias are substantial in specific conditions. A dual frame sampling and weighting method that emphasizes this importance is proposed to reduce this kind of bias. In fact, how the estimates from the two frames are combined is crucial. One of the main findings is that the choice of the compositing factor has only a small effect on variance but quite a significant effect on bias.

Some interesting empirical research was done by the Pew Research Center, analyzing data from 11 dual frame surveys conducted from 2009 to 2010. Various substantial topics were covered in these surveys: public policy issues, economic ratings, political views and attitudes, religious and social values, attitudes toward information–communication technologies, applications, and social networking services. Primarily, bias was estimated by comparing the fixed telephone and dual frame samples. Both were weighted to the same standard set of demographic parameters and to telephone usage (fixed only, mobile only, dual users). Although the number of mobile-only households increased, the amount of noncoverage bias is still small for most items. Despite being small, the differences are all in the predicted direction. This indicates that the estimates are biased, since most of the coverage problem is caused by the mobile-only households. The second comparison was made between fixed telephone samples and fixed telephone samples that additionally include mobile-only households. The differences are a bit lower than in the first comparison but have the same patterns (Leah et al. 2010).

To summarize, omitting the mobile-only segments, has had so far relatively small consequences for the estimates, at least in the situation of 20% (or lower) mobile-only units as in the US and in many EU countries. However, bias can have important substantive implications for some specific segments (i.e. younger, less educated, lower income, nonwhite) and for some specific variables (political attitudes, technology use patterns, some social issues). In addition, when the share of the mobile-only segment increases—which is an unavoidable trend—the biases may become substantial (Kuusela et al. 2007).

An issue that has not been addressed enough in dual frame surveys research is the definition and measurement of telephone use which is needed to weight data. At first glance we see

three key domains of telephone use: (1) fixed phone only users, (2) mobile phone only users, and (3) the overlap (users of both). These domains have to be equivalently identified on sample as well as on population data.

However, considerable ambiguity exists with respect to the wording of survey questions aimed at identification of telephone use domains. First, we have to specify the device (fixed/mobile) and possibly distinguish it from similar devices (e.g. fax machines) and determine if we are interested only in working or also in nonworking devices. Second, device ownership can be communal (fixed phones are often owned by the household) or individual which has to be distinguished also from business and company owned phones. Third, definition of phone use is the major ambiguity as mere possession is very different from actual availability of a person via the fixed or mobile telephone device.

Therefore, the overlap domain depends on the question wording, particularly to the distinction between phone *possessions* (person has a mobile/fixed phone) and *availability* (can be reached on a mobile/fixed phone). The overlap segment defined by availability is a subgroup of the overlap segment defined by possession. Within availability there are additional variations (availability in general or in a specific time frame, e.g. 'this moment'). See Box 16.5 for examples of possession and availability wordings.

Assessing availability is important since the overlap segment (according to possession) can be very heterogeneous in regard to phone use patterns. Data from NHIS 2008 indicate that about a quarter of the overlap receives all or almost all calls on their mobile phones (Blumberg and Luke 2009). Moreover, these "mobile-mostly" households are more typical for certain sociodemographic groups (the poor, those living in metropolitan areas, and having children). Similarly, "fixed-mostly" households can be defined as households that receive most calls on their fixed phone. About 30% of the overlap receive very few or none of their calls on their mobile phone according to CPS data (Tucker et al. 2007).

Box 16.5: Examples of question wordings for telephone possession and availability

Different survey organizations use different wordings to ask about telephone use possession and availability. Below the example of PEW and Eurobarometer surveys is given. For more examples of US survey's wordings see Appendix B of the American Association of Public Opinion Research considerations for telephone surveys with respondents reached via cell phone numbers (AAPOR Cell Phone Task Force 2010).

Questions about phone possession

PEW surveys

[2008/09] [Cell subsample] PC1. Now thinking about your telephone use … Is there at least one telephone INSIDE your home that is currently working and is not a cell phone?

[2008/09] [Landline subsample] PL1. Now thinking about your telephone use … Do you have a working cell phone?

[2006] [Cell subsample] Is the cell phone your only phone or do you also have a regular telephone at home?

[2006] [Landline subsample] Do you happen to have a cell phone or not?

EB and Flash EB

Do you own a fixed telephone in your household?

Do you own a personal mobile telephone?

[Mobile phone subsample] D11b. Do you have a fixed telephone at home?

[Fixed phone subsample]—D11a. Do you personally have a mobile phone?

Questions about availability

PEW surveys

[2008/09] P2. Of all the telephone calls that you receive, do you get … Almost all calls received on cell phones? Some received on cell phones and some on regular phones? Very few or none received on cell phones? [Rotate options—keeping SOME in the middle]

[2006] Thinking about all the phone calls you make, do you make more calls

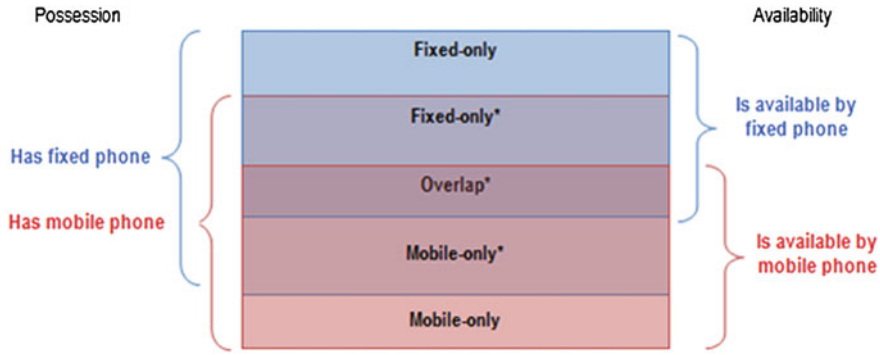


Fig. 16.3 Dual frame sampling model (Vehovar and Slavec 2011)

with your cell phone or more calls with your regular home phone?

[IF uses cell phone more] Would that be a LOT MORE or just a FEW more with your cell phone?

[IF uses regular phone more] Would that be a LOT MORE or just a FEW more with your regular home phone?

Flash EB

[Mobile subsample] D15. Could I have reached you just now on your fixed home phone?

[Fixed subsample] D14. Could I have reached you just now on your mobile phone?

[Both] D13. Thinking about all the phone calls you make, do you make more calls with your mobile phone or more calls with your fixed home phone? 1—A lot more with mobile phone, 2—A few more with mobile phone, 3—About equal, 4—A few more with fixed home phone, or 5—A lot more with fixed home phone.

[Both subsamples] D16. How often are you reachable on your mobile phone? Would you say…? 1—Always, 2—Most of the time, 3—Some of the time, 4—Rarely, or 5—Your phone is only switched on when you make a call.

When data for both criteria are at disposal, the overlap segment can be further dissected into three strata. Instead of referring to the new strata as fixed and mobile mostly, we label them with an asterisk: units that are only available by their fixed phone are "fixed-only*," while those available only by mobile phone are "mobile-only*." What remains of the overlap segment that cannot be further dissected is labeled as "overlap*." Thus, the total size of the sample consists of five strata (Vehovar and Slavec 2011):

1. Fixed-only (has only fixed phone);
2. Fixed-only* (has both services but is only reachable to fixed phone);
3. Overlap* (reachable by fixed and mobile phone);
4. Mobile-only* (has both services but is only reachable to mobile phone); and
5. Mobile-only (has only mobile phone (Fig. 16.3).

Usually only data about phone possession are at disposal via some large official surveys (e.g. LFS), but we do not know this information (i.e. which unit is in overlap) in advance to perform proper stratification. Thus, only poststratification can be used.

16.3 Optimizations of Errors and Costs

16.3.1 General Optimization Issues

All aspects of survey data quality in broadest sense as presented above can be optimized within given budgetary constraints. Literature dealing simultaneously with data quality issues and survey costs to compare different survey

design alternatives is surprisingly scarce—a review of empirical studies indicated that only 30% investigated also costs (Dolnicar et al. 2009). The object of optimization can be the entire data collection process, which is regarded a business process and as such nested in a broader context of business optimization and management, where standard quality concepts can be applied (such as cost-benefit analysis, total quality management etc.). Conversely, in survey practice, smaller components of the survey data process are also optimized, such as the workload of interviewers, the number of solicitation waves, response rates, calling strategy, and so on.

Here, we limit the discussion to optimization of the cost per unit of accuracy, which is measured by the inverse of mean squared error:

$$\frac{\text{COST}}{\text{accuracy}} = \frac{\text{COST}}{\frac{1}{\text{MSE}}} = \text{COST} * \text{MSE}$$

Minimizing the product of cost and errors is a standard approach, particularly in classic stratification, and cluster sampling literature (Deming 1953; Hansen et al. 1953; Kish 1965; Cochran 1978). However, these studies focus only on the product of costs and sampling variance. The extension to mean squared error and costs was theoretically introduced only recently (Lynn and Elliot 2000; Lyberg et al. 1998) and only rare empirical applications exist. Besides our evaluations of alternative survey designs (Vehovar and Lozar Manfreda 1998; Vehovar et al. 2001; Vehovar et al. 2010) there is also some work on double sampling nonresponse optimization (Harter et al. 2007; Deming 1953; Elliot et al. 2000) and some evaluations of alternative survey designs (Linacre and Trewin 1993). There are also some examples of dual frame survey optimizations (Deming 1950; Lepkowski and Groves 1984).

In classical sampling theory, optimization is usually performed a priori: the cost-error components are first explicitly modeled and then optimized analytically. On the other hand, standardized calculations of costs and errors of different designs can be evaluated on existing data from past surveys. Alternatives are ranked based on the product of cost errors metric derived from existing data from already implemented or simulated survey designs. Results of such comparison can be then applied to make cost-efficient survey design decision making.

16.3.2 Analytical Solution for Dual Frame Telephone Surveys

Dual frame telephone surveys require a more complex optimization procedure as accuracy and survey costs have to be simultaneously optimized over the two sampling frames (F—fixed, M—mobile. The mixture parameter x as a given share of the fixed telephone subsample in the total subsample should be modeled.

An analytical solution for x was developed by Vehovar and Slavec (2011), deriving from Hartley's (1962) expression for the estimator of mean across both domains, the estimate for the stratified mean for three strata (1—fixed only, 2—overlap, 3—mobile only) and then extended also to five strata (1—fixed only, 2—overlap but available only by fixed phone, 3—overlap (available by both phones), 4—overlap but available only by mobile phone, 5—mobile only).

In this approach, mean square error and costs are multiplied on the weighted stratified sample and the optimal solution computed by finding the global minimal value of the product on the interval from 0 to 1, which is the domain of definition of the mixture parameter x. The procedure was applied to Flash Eurobarometer 2008 dual frame data (Vehovar and Slavec 2011) (see Box 16.6).

Box 16.6: Case study: Analysis of Flash Eurobarometer 2008 data

Vehovar and Slavec (2011) applied the above approach to data from a 2008 Flash Eurobarometer dual frame telephone survey on Public attitudes and perceptions in the euro area for eight countries: Austria, Belgium, France, Germany, Italy, Portugal, Spain, and Slovenia (European Commission 2008b). The sample size was approximately 1,000 per country and 60–70% units were interviewed by fixed and the others by mobile phone.

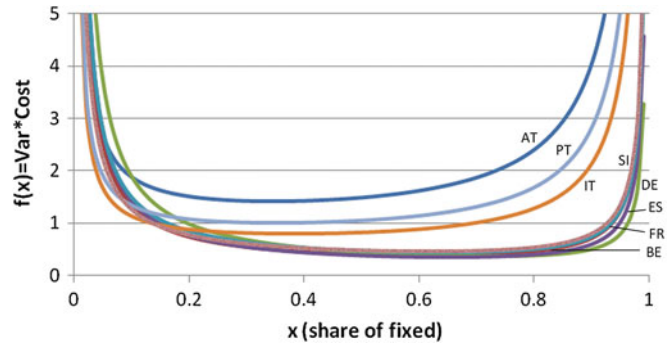


Fig. 16.4 Relations between errors-costs and share of fixed-line telephones in a sample for 2008 Flash Eurobarometer study (Vehovar and Slavec 2011)

Before optimization, the data were weighted to account for different demographic structures (gender and age, economic activity, region) and telephone use (five strata). The optimal mixture parameter was computed for ten different variables (socio-political issues) assuming different fixed to mobile cost ratios (e.g. 3 on the Fig. 16.4 below).

The results presented on Fig. 16.4 show that the optimal mixture parameter lies in a flat area, so errors, and costs do not increase substantially if we chose a higher or lower solution than suggested. When the mobile phone is three times more expensive than the fixed phone, the optimal mixture parameter for most countries is about 0.6–0.7 except for Austria, Portugal, and Italy where it is only about 0.35. This is due to the highest proportions of the mobile-only segment. The highest optimal share of fixed units was computed for Germany where the mobile-only segment is smaller than in other countries. The optimal design allocation is strongly correlated with the cost ratio. When there is no difference in costs, taking more mobile units is optimal, except for Germany where the fixed phone coverage is still very high. On the other hand, where the cost of a mobile interview is very high in relation fixed interviewing; taking more fixed units is optimal in all countries (Vehovar and Slavec 2011).

However, using this procedure we have to be aware of same problems. First, population strata estimates rely on sample estimates themselves. Second, there are no accurate estimates of the fixed to mobile cost ratio. Third, we took some problematic suppositions, in particular that there is no bias and that the response rate is the same for both subsamples. In the next section an alternative method of evaluating different survey designs is presented.

16.3.3 Post Survey Evaluation of Different Modes: Cost-Error Comparisons of Alternative Mixtures

It is possible to provide proper cost-error calculations for a set of *past* and predefined survey designs that were already implemented or simulated. There are three ways to evaluate the relationship between costs and errors when comparing survey designs:

1. The design providing a required fixed MSE of a target survey variable and having the lowest costs (used when having very clear requirements regarding the accuracy of survey estimates, e.g. unemployment rate in Labor Force Surveys);
2. The design with the lowest MSE of a target survey variable for a given fixed budget (most common approach where data quality follows and adapts to the available budgets);
3. The minimal product of costs and MSE for a target survey variable in each competing survey design option (most appropriate

approach when performing comparisons among different solutions at fixed initial or final sample sizes).

The third approach is the most illustrative insight into effects of alternative designs since calculating the product of MSE and costs enables us to rank alternative survey designs and identify the one with the lowest value of this product. Moreover, it enables comparisons with fixed final sample size. Because of its generality it can be applied to any survey design alternatives, particularly to those related to the inclusion of mobile phones—it can contribute metrics needed to decide about including mobile phones into a survey design.

Although this approach is inferior to the optimization algorithms such as the one presented above, it is useful for models that are too complex to be processed analytically. However, for simple cases (one mode, one solicitation wave, and sampling variance as the only MSE component) the optimal sample design structure can be analytically determined. Box 16.7 illustrates the use of this approach for comparison of alternative survey designs.

Box 16.7: Case study: Analysis of Slovenian statistical office survey on ICT usage among households 2008

In 2008, Faculty of Social Sciences (FSS) at University of Ljubljana performed an experimental study aimed at comparing various mixed-mode designs with face-to-face and telephone surveys (Berzelak et al. 2008) in Eurostat survey on ICT usage among households in Slovenia. The official face-to-face and telephone interviewing was performed by the Statistical Office of the Republic of Slovenia (SORS) (2008). Samples for both, official, and experimental parts were obtained from the Central Population Register (CRP), allowing obtaining true population values for key demographic variables. This was extended by inclusion of simulated scenarios of mobile phone surveying with different response rates, as well as a simulated dual frame scenario with a third of the sample surveyed by RDD mobile phone survey and two-thirds by conventional CATI. Parameters for these simulated designs were obtained from previous surveys conducted by FSS (Vehovar et al. 2004).

The obtained estimates for age of respondents in simulated scenarios and fixed CATI mode were compared to the true population values from CRP. Dual frame approach was found to be the most optimal among all approaches, despite slightly higher costs compared to fixed CATI and some mobile-only designs. Namely, the appropriate combination of fixed and mobile telephone samples allowed more accurate estimation of the population parameter to the degree which outweighed the increased costs.

16.4 Conclusions

Technological developments are rapidly changing the survey industry landscape. On one hand, the emerging new technologies can be increasingly used to effectively conduct surveys or to support current survey modes, so that survey researches have unprecedented alternatives to cost-effectively reach respondents, collect data, and then process them. On the other hand, with the expanding information–communication technology, our survey populations are overloaded with communication messages, especially for marketing purposes. The increased concerns with privacy issues are making not only marketing but also academic and official survey research seen as intruding, so the cooperation and response rates are declining. This is, in fact, a fatal problem for probability sample surveys—which are discussed here—while for nonprobability and quota samples this is not so much a critical issue.

In regard to telephone interviewing, the central phenomena are the rise of mobile telephony which is not only complementing but also replacing classic fixed telephone surveys. The latter were the dominant and best-elaborated

survey mode in past decades. Mobile telephone surveys bring changes and challenges not so much to the interviewing procedures—as there are no radical methodological differences (i.e. mode effect, response rate) between fixed and mobile phone interviews—but particularly to costs and to the way how to administer, manage, and sample the target units. There are typically no public registers of mobile phone surveys (the same is true for fixed telephones); however, serious legal issues and country specifics (not imposed for fixed telephone surveys) may exist for sampling, validating, dialing, and calling mobile phone numbers for surveying. Similarly, complicated costs plans and patterns may exist across countries, as well as different technical options, and expectations for incentives. In addition, people are using mobile phones in very different ways. Altogether, this makes mobile telephone surveys very different to the fixed ones.

As for now, mobile only surveys are not yet a generally feasible solution—except for specific countries (e.g. Finland)—but need to be mixed with other survey modes, most frequently they are combined with fixed telephone surveys. However, since most of the population uses both types of phone, the overlap issue has to be resolved before computing joint estimates.

One of the key challenges in telephone surveys is how to combine the two modes optimally, which is an aspect very relevant also for mixed mode settings in general. Here, we presented two applications, both using the optimization by minimizing the product of errors and costs. The first example is an approach how to determine the optimal mixture of fixed and mobile subsample. We showed how to analytically develop the product equation and compute the exact optimal mixture for a given setting. The results show that the mixing parameter is relatively robust and lies in a broad acceptable interval. In countries where fixed telephones still cover majority of population, the fixed telephone subsample should also dominate (e.g. 70%) over the mobile one.

The second example illustrates how to perform the postsurvey evaluations for mixed-mode surveys, where costs and errors are treated simultaneously. We believe that similar calculations will become increasingly frequent when deciding about right combination of the survey modes.

Of course, only the isolated issues of finding the optimal mixture were discussed here, without a separate treatment of differential nonresponse and/or mode effect, which would bring additional complexity.

In the future, we can expect that the difference between fixed and mobile phones will slowly diminish. However, given the existing lifestyles, cost ratios, legal restriction, and so on. we can expect that the specifics (existing or new) will be in existence for considerable time. Thus, research methodologists will have a continuous challenge for finding the optimal balance between data quality and resources.

References

AAPOR Cell Phone Task Force (2010) *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached* via *Cell Phone Numbers*. American Association for Public Opinion Research, from http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818.

Belak, E. (2007). *Vpliv uporabe mobilnih telefonov na anketno zbiranje podatkov. Magistrsko delo*. Ljubljana: FDV.

Berzelak, N.,Vehovar, V. & Lozar Manfreda, K. (2008). *What can we achieve with 5 Euros? optimization of survey data quality using mixed-mode approaches*. Paper presented at the Applied Statistics Conference, Ribno

Biemer, P. P. (2010). Total survey error: Design implementation and evaluation. *Public Opinion Quarterly, 74*(5), 817–848.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: Wiley.

Blumberg, S. J., & Luke, J. V. (2009). *Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2008*. National Center for Health Statistics, from http://www.cdc.gov/nchs/nhis.htm.

Bosnjak, M., Neubarth, W., Couper, M. P., Bandilla, W., & Kaczmirek, L. (2008). Prenotification in web-based access panel surveys: The influence of mobile text messaging versus e-mail on response rates and sample composition. *Social Science Computer Review, 26*(2), 213–223.

Brick, J. M., Dipko, S., Presser, S., Tucker, C., & Yangyang, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public opinion Quarterly,* 70(5, Special Issue), 780–793.

Brick, J. M., Edwards, W. S., & Sunghee, L. (2007). Sampling telephone numbers and adults, interview length, and weighting in the California health interview survey cell phone pilot study. *Public Opinion Quarterly, 71*(5), 793–813.

Brick, M. J., Cervantes, I. F., Lee, S., & Norman, G. (2011). *Nonsampling Errors in Dual Frame Telephone Surveys*. Survey Methodology *37* (1), 1–12.

Callegaro, M., Steeh, C., Buskirk, T. D., Vehovar, V., Kuusela, V., & Piekarski, L. (2007). Fitting disposition codes to mobile phone surveys: Experiences from studies in Finland, Slovenia and the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(3), 647–670.

Chang, L. C., & Krosnick, J. A. (2009). National surveys Via Rdd telephone interviewing versus the internet. Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*(4), 641–678.

Cochran, W. G. (1978). *Sampling techniques* (3rd ed.). New York, NY: Wiley.

de Leeuw, E. D. (1992). *Data quality in mail, telephone, and face-to-face surveys*. Amsterdam: TT-Publicaties.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233–255.

Deming, W. E. (1950). *Some theory of sampling*. New York, NY: Wiley.

Deming, W. E. (1953). On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse. *Journal of the American Statistical Association, 48*(264), 743–772.

Dillman, D. A., Smyth, J. D., & Christian, C. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken: Wiley.

Dolnicar, S., Laesser, C., & Matus, K. (2009). Online versus paper: Format effects in tourism surveys. *Journal of Travel Research, 47*(3), 295–316.

Elliott, M., Little, R. J. A., & Lewitsky, S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association, 95*, 730–738.

European Commission (2008a). Special Eurobarometer 293. *E-communication Household Survey*. TNS Opinion and social, from http://ec.europa.eu/public_opinion/archives/ebs/ebs_293_full_en.pdf

European Commission (2008b). Flash Eurobarometer 251. *Public attitudes and perceptions in the Euro area*. Gallup Europe, Brussels [Producer]. GESIS, Cologne [Publisher]. ZA4743, data set version 2.0.0. Description: http://info1.gesis.org/dbksearch13/sdesc2.asp?no=4743&db=e&doi=10.4232/1.10007 (July 29 2010).

Fuchs, M. (2007). Mobile web surveys: A preliminary discussion of methodological implications. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 77–94). Hoboken, NJ: Wiley.

Gabler, S., & Häder, S. (1999). Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland. *ZUMA-Nachrichten Heft, 44*, 45–61.

Gallup Europe (2009, February). *Presentation Technical workshop*. Paper presented at the Eurobarometer Technical Workshop, Budapest, Hungary.

Groves, R. M. (1989). *Survey errors and survey costs*. Hoboken, NJ: Wiley.

Groves, R. M. (2005, March). *Past, present, and future*. Washington, DC: Paper presented at the Total Survey Error Workshop.

Groves, R. M. (2006, April). *Survey budgets, cost models, and responsive survey designs*. Washington, DC: Paper presented at the Survey Cost Workshop.

Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.

Groves, R. M., & Lepkowski, J. M. (1982). *Alternative dual frame mixed mode surveys designs*. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly, 72*(2), 167–189.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory*. New York, NY: Wiley.

Harter, R., Mach, T., Wolken, J., & Chapline, J. (2007, June). *Determining subsampling rates for nonrespondents*. Paper presented at the Third International Conference on Establishment Surveys, Montréal, Canada.

Hartley, H. O. (1962). *Multiple frame surveys. Proceedings of the Social Statistics Section*, American Statistical Association.

Karr, A. F., & Last, M. (2006). *Survey Costs: Workshop Report and White Paper*.

Keeter, S., Kennedy, C., Clark, A., Tompson, T., & Mokrzycki, M. (2007). What's missing from national landline RDD surveys? *Public Opinion Quarterly, 71*(5), 772–792.

Kennedy, C. (2007). Evaluating the effects of screening for telephone service in dual frame RDD surveys. *Public opinion Quarterly, 71*(5), 750–771.

Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.

Kuusela, V., & Simpanen, M. (2002). *Effects of mobile phones on phone survey practices and results*. Copenhagen: ICIS Conference.

Kuusela, V., Callegaro, M., & Vehovar, V. (2007). The influence of mobile telephones on telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Javrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 87–112). Hoboken, NJ: Wiley.

Lavrakas, P. J., Shuttles, C. D., Steeh, C., & Fienberg, H. (2007). The state of surveying cell phone numbers in the United States: 2007 and beyond. *Public Opinion Quarterly, 71*(5), 840–854.

Leah, C., Keeter, S., Purcell, K., & Smith, A. (2010). *Assesing the cell phone challenge*. PEW research

centre, from http://pewresearch.org/pubs/1601/assessing-cell-phone-challenge-in-public-opinion-surveys

Lepkowski, J. M. & Groves, R. M. (1984). *The impact of bias on dual frame survey design. AMSE model. Proceedings of the Section on Survey Research*, American Statistical Association: http://www.amstat.org/sections/SRMS/ Proceedings/papers/1984_051.pdf

Lepkowski, J. M., & Groves, R. M. (1986). A mean squarred error model for dual frame, mixed mode survey design. *Journal of the American Statistical Association, 396*(81), .

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York, NY: Wiley.

Linacre, S. J., & Trewin, D. J. (1993). Total survey design—application to a collection of the construction industry. *Journal of Official Statistics, 9*(3), 611–621.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2007). Reaching the U.S. cell phone generation: Comparison of cell phone survey results with an ongoing landline telephone survey. *Public Opinion Quarterly, 71*(5), 814–839.

Lyberg, L. E., Japec, L., & Biemer, P. P. (1998). *Quality improvement in surveys—A process perspective*. Paper presented at the Survey Research Methods Section.

Lynn, P., & Elliot, D. (2000). *The british crime survey: A review of methodology*. London, UK: National Centre for Social Research.

McCutcheon, A. L. (2009). *Mode or Mensch? Respondent sensitivity to mode changes in data collection methods*. The 1st Alec Gallup Future of Survey Research Forum: Incorporating Mobile Phones in Social and Policy-oriented Surveys. Brussels.

Nicolaas, G., Lynn, P. & Lound, C. (2000). *Random digit dialing in the UK: viability of the sampling method revisited*. Paper presented at the International Conference on Social Science Methodology, Cologne.

Pew Research Center (2008, December 18). Calling cell phones in '08 pre-election polls. *News Release*. Retrieved February 20, 2009, from http://people-press.org/reports/pdf/cell-phone-commentary.pdf

Singer, E. (2006). Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 637–645.

Statistical Office of the Republic of Slovenia. (2008). *Community survey on ICT usage in households and by individuals—2008 [Survey data for Slovenia]*. Ljubljana, Slovenia: Statistical Office of the Republic of Slovenia.

Steeh, C., & Piekarski, L. (2008). "Accommodating new technologies: Mobile and VoIP communication." In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 423–448). New York, NY: Wiley.

Steeh, C., Buskirk, T. D., & Callegaro, M. (2007). Using text messages in U.S. mobile phone surveys. *Field Methods, 19*(1), 59–75.

Tortora, R. D. (2004). Response trends in a National Random Digit Dial telephone survey. *Advances in Methodology and Statistics, 1*(1), 61–84.

Traugot, M. V., Groves, R. M., & Lepkowski, J. M. (1987). Dual frames to reduce nonresponse. *Public Opinion Quarterly, 51*(4), 522–539.

Tucker, C., Brick, M. J., & Meekin, B. (2007). Household telephone service and usage patterns in the United States in 2004. *Public Opinion Quarterly, 71*(1), 3–22.

Vehovar, V., Belak, E., Batagelj, Z., & Čikić, S. (2004). Mobile phone surveys: The Slovenian case study. *Metodološki zvezki (Advances in Methodology and Statistics), 1*(1), 1–19.

Vehovar, V., & Lozar Manfreda, K. (2008). Overview: Online surveys. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The handbook of online research methods* (pp. 177–194). Thousand Oaks, CA: SAGE Publications.

Vehovar, V., & Slavec, A. (2011). *Costs Error Optimization for Cell-Landline Dual Frame Surveys*. Paper presented on the 66th conference of the American Association for Public Opinion Research.

Vehovar, V., Lozar Manfreda, K., & Batagelj, Z. (2001). Sensitivity of electronic commerce measurement to the survey instrument. *International Journal of Electronic Commerce, 6*(1), 31–52.

Vehovar, V., Berzelak, J., & Lozar Manfreda, K. (2010). Mobile phones in an environment of competing survey modes: Applying metric for evaluation of costs and errors. *Social Sciences Computer Review, 28*(3), 303–318.

Virtanen, V., Sirkiä, T., & Jokiranta, V. (2007). Reducing nonresponse by SMS reminders in mail surveys. *Social Science Computer Review, 25*(3), 384–395.

Waksberg, J. (1978). Sampling method for random digit dialing. *Journal of American Statistical Association, 73*(361), 40–46.

Weisberg, H. F. (2005). *The total survey error approach: A guide to the new science of survey research*. Chicago: University of Chicago Press.

Suggested Readings

AAPOR Cell Phone Task Force (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached* via *Cell Phone Numbers*. American Association for Public Opinion Research.

Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L, I. I., & Waksberg, J. (2001). *Telephone survey methodology*. Hoboken, NY: Wiley.

Lepkowski, J. M., Tucker, C., Brick, J. M., de Leeuw, E. D., Japec, L., Lavrakas, P. J., et al. (2008). *Advances in telephone survey methodology*. Hoboken, NY: Wiley.

Mail Survey in Social Research

17

Alana Henninger and Hung-En Sung

17.1 History and Purpose

Surveys are extremely versatile, efficient, and generalizable research instruments, making them a popular method of data collection (Bachman and Schutt 2007) for thousands of years (Erdos and Morgan 1970). The census is the earliest type of recorded survey. "Counting" the children of Israel is mentioned in the bible, while the first Roman census was conducted in 443 B.C. Surveys are also mentioned in early Babylonian records, and throughout history, such as the Domesday Book of 1086 and the surveys conducted by Napoleon in 1806. The first recorded mail survey in history was conducted in 1577 King Philip II of Spain in an effort to catalog his possessions throughout the New World. In the United States, mail surveys were first used for marketing research (e.g., an 1895 survey on advertising conducted by Harlow Gale at the University of Minnesota), and then for the Census beginning in 1960 (Erdos and Morgan 1970).

A. Henninger
The Graduate Center,
The City University of New York,
New York, NY, USA

H.-E. Sung (✉)
Department of Criminal Justice,
John Jay College of Criminal,
Justice, 899 Tenth Avenue,
New York, NY 10019, USA
e-mail: hsung@jjay.cuny.edu

Prior to the mid 1970s, however, mail surveys were considered a poor substitute for face-to-face interviews due to problems inherent to the method such as incomplete responses, self-selection bias, and response rate (Dillman 1978; Mangione 1995). Response rates were especially low in the early years of mail survey research. The average response rate for mail surveys in 1937 was 7%, with an overall range of less than 2% to a high of 30%. By 1955, mail surveys with response rates exceeding 43% were considered as having a high return (Erdos and Morgan 1970). Needless to say, mail surveys were not held in high regard. In the mid 1970s, however, the negative attitude toward mail surveys began to change. One reason for this change was that the response rates for face-to-face interviews fell from an average of 80–85% to an average of 60–65%, while the cost began to significantly rise. These declining response rates are thought to have occurred as a result of increased economic status and mobility in the general population. More households contained spouses who were both employed, more leisure activities were conducted outside the home, and there was a rise in the ability for more households to relocate. The price of conducting surveys then increased because interviewers were having a difficult time locating respondents (Dillman 1978).

Secondly, the individuals hired to administer face-to-face questionnaires are expensive. The hiring process can take a great deal of time and resources, and once hired, the interviewers require extensive training (Dillman 1978;

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_17, © Springer Science+Business Media New York 2012

Dialsingh 2008). They must be willing to work odd hours (such as nights and weekends) in order to reach the greatest amount of respondents, in neighborhoods that are not always safe, therefore security is often an added expense (Dillman 1978). More recently, the use of telephone surveys has become more difficult as well, due to the increasing use of cell phones by the general public. Coverage error for telephone surveys becomes a concern for two reasons. Firstly, the targeted population cannot be reached on landline telephones through Random Digit Dialing (Dialsingh 2008; Atkeson et al. 2011). This is then compounded by the fact that individuals without landlines are overwhelmingly non-white, renters, economically disadvantaged, and southern. This constitutes a major portion of the population who are absent from the telephone survey process (Atkeson et al. 2011). In light of these increasing costs and problems, alternative survey methods began to gain acceptance and more frequent use (Dillman 1978; Dialsingh 2008).

Today, the use of self-administered, mailed questionnaires for data collection has become a popular method in a variety of disciplines (e.g., criminal justice, psychology, marketing, political science) (Bachman and Schutt 2007). More specifically, the use of mail surveys has become especially prevalent by the government, school system (at both the primary and collegiate level), and businesses for a variety of research purposes (Dillman 1991). Mail surveys are unsupervised, self-administered questionnaires that are mailed to sampled respondents through the postal or other delivery service (Bourque and Fielder 2002; Bachman and Schutt 2007; Brill 2008; Dialsingh 2008; Wolf 2008).

Mail surveys are best used when (1) the sample covers a large geographical area; (2) the budget is restricted; (3) respondents are to be given ample time to consider their responses; (4) close-ended questions are to be used; (5) the sample is invested in the topic of the survey; (6) the objectives of the research are modest in length; (7) respondent privacy is important; (8) visual presentation is more effective than oral questioning; and (9) a limited staff is available to conduct the survey (Mangione 1995).

17.2 Design of the Mail Questionnaire

After deciding on a research question, the first important consideration before designing the mail survey is which population will be sampled (Dillman 1978). Sampling portions of a particular population allows researchers to obtain information about the characteristics of a smaller group of people and then generalize those characteristics to the larger population (Dillman 1991). Mail surveys are especially efficient because they have the ability to sample businesses and households, as well as specific individuals (Dialsingh 2008).

In general, the key to getting the most out of a mail survey is spending the time and resources designing high quality survey materials at every stage of the process. Information collected from a poorly worded and designed survey is often invalid and misleading, thus designing an efficient and well constructed questionnaire saves money in the long run (Mangione 1995). A mail survey should adhere to a five-part timeline, including the following mailings: (1) a pre-notification; (2) survey packet containing a questionnaire with cover letter and other instructional materials; and follow-up correspondence in the form of a (3) thank you postcard; (4) replacement questionnaire if necessary; and (5) a final contact (Dillman 2000; Brill 2008; Dialsingh 2008).

17.2.1 Pre-Notification

Respondents should first be sent a pre-notification informing them that they will be receiving a mailed questionnaire within the next few days, and that it would be greatly appreciated if they could take the time to respond (Dillman 2000; Brill 2008; Dialsingh 2008). If an incentive is to be offered, this should also be mentioned in the pre-notification (Dillman 2000; Bourque and Fielder 2002). This pre-notification should be briefly worded in a positive manner (Dillman 2000), and can be in the form of a letter or postcard (Brill 2008; Dialsingh 2008).

Dillman (2000) recommends the use of a letter rather than a postcard for several reasons. Letters, he argues, appear more professional and convey a greater feeling of importance than a postcard. In terms of the time it takes to commit information to long-term memory (approximately 20 s), letters are the better choice because they take longer to open and can provide more information than a postcard that takes mere seconds to flip over and read. Touches of personalization that have been shown to increase trust in the researcher (e.g., letterhead and signature) can also be included in a letter, but not in a postcard.

17.2.2 The Survey Packet

The actual mailed survey should be in the form of a packet containing a cover letter, the questionnaire (including detailed instructions explaining how to answer and return the questionnaire), the incentive (if one is being used), and a mode of returning the survey (Dillman 2000; Brill 2008; Dialsingh 2008). It is sometimes recommended that a brochure that provides information about the organization that is sponsoring the survey also be included in the packet (Brill 2008; Dialsingh 2008).

17.2.3 The Outer Envelope

As in many other areas of life, the first impression that the survey packet makes is crucial (Dillman 1978; Dialsingh 2008). Researchers should consider the demographics of the sample when designing the survey packet envelope, so that the desired reaction from the respondent is elicited (Dialsingh 2008). The envelope containing the survey should appear professional—using advertising gimmicks such as envelopes in unusual shapes, sizes, and colors have the tendency to turn the potential respondent off, and should thus be avoided. Additionally, stamping the envelope with instructions such as "dated materials enclosed" or "immediate reply requested" are not recommended (Dillman 1978). Sampled respondent are also more likely to open a survey envelope with a postage stamp (rather than a bulk rate stamp) because the parcel will appear more official, and less like junk mail (Dillman 1978; Dialsingh 2008).

If possible, the contents of the survey packet should be sent in a business envelope rather than a flat mailout. Flat mailout envelopes not only resemble junk mail, they also cost more in postage to send. Additionally, a smaller business envelope makes the survey packet appear less cumbersome, thus suggesting less of a time commitment on the part of the respondent. The address line should always be typed (Dillman 1978), and it is always best to use the name of the person who has been selected to respond to the survey on the envelope. If a business or household in general is asked to respond to a survey, however, generically addressing the envelope with an identifier such as "Research Household" is acceptable (Dialsingh 2008).

17.2.4 The Cover Letter

Every mail survey should be accompanied by a cover letter. A survey cover letter should serve as a precise, one-page introduction to the questionnaire (Dillman 1978, 2000; Dialsingh 2008) that is personalized with the use of letterhead, sponsorship information, individualized salutations, dates, and signatures (Mangione 1995; Bourque and Fielder 2002; Bachman and Schutt 2007). When writing the cover letter, the researcher should keep in mind that they are requesting a favor from the respondent and refrain from using an impersonal tone (Dillman 2000). The opening sentence should draw the reader in, and the overall writing style should motivate the respondent to participate using language that can be comprehended by all reading levels (Dillman 1978; Mangione 1995; Dialsingh 2008). Dillman (2000) suggests creating a "mental image of an acquaintance with an educational level a little lower than the average of the survey population from whom assistance is being requested."

Information should then be provided on the purpose and usefulness of the study (Dillman 1978; Bourque and Fielder 2002; Bachman and Schutt 2007; Dialsingh 2008). The researcher should first highlight why this study is important

to the respondent by focusing on a group to which the respondent belongs, thus making the survey appear to be a potentially rewarding experience (Dillman 1978; Bourque and Fielder 2002). This appeal to the membership of a respondent can be general (e.g., referencing "the needs of citizens throughout the state") or specific (e.g., referencing "helping your organization better serve the needs of its members"), but caution should be taken that an appeal is not made to an interest group that is not accepted by the entire sample (e.g., suggesting that "the survey is being done to help the Democratic (or Republican Party) better serve the needs of all Americans"). Furthermore, the researcher should avoid suggesting biased motives in the body of the cover letter, such as political affiliation or any personal gain that may result from conducting the survey (e.g., "This survey is integral to the completion of my dissertation") (Dillman 1978).

Respondents often feel that their opinions are not very important. Therefore, convincing the respondent that their specific response is valuable should be the focus of the second paragraph. The researcher can convey this importance by simply explaining that the survey is seeking to gain a representative perspective on a particular issue from a group of citizens that this issue may have an effect on. Confidentiality is then often addressed in the third paragraph. Respondents may fear that their responses will be misused, resulting in later repercussions, and addressing the issue of confidentiality in the cover letter may serve as a method of allaying these fears and increasing response rates (Dillman 1978). Standard confidentiality procedures should briefly be described (Mangione 1995; Bachman and Schutt 2007), including information such as the use of human subject boards for survey approval (Dillman 2000) and keeping completed surveys in a secure location, with access limited only to necessary research staff (Dialsingh 2008). It is also customary to offer to provide a detailed confidentiality policy to any respondent who requests it (Dillman 2000). Placing identification numbers, rather than names, on all survey materials will serve as further reassurance that the confidentiality policy will be upheld (Dillman 1978, 2000; Mangione 1995; Dialsingh 2008).

The final paragraph should briefly reiterate the purpose of the study (Dillman 1978), provide a date by which the questionnaire should be returned (Dialsingh 2008), and list the contact information for the researcher (either an email address or toll free telephone number) (Mangione 1995; Dillman 2000; Dialsingh 2008). A good cover letter will offer to answer any questions that the respondent may have (Dillman 2000), as well as offer to provide the results of the study if the respondent is interested (Dillman 1978). Many human subject review boards now require that the participant be notified that any response is voluntary (Mangione 1995; Dillman 2000), so it is often helpful at this point to ask the respondent to send back the questionnaire whether it has been completed or not. Researchers are then able to remove names from the follow-up mailing list once they receive blank questionnaires from nonrespondents, thus saving both time and money. Finally, the last paragraph should end with an expression of thanks and a mention of any token of appreciation (incentive) that may have been included in the packet (Dillman 2000).

An additional personalization of the cover letter can be found in the signature line. Hand-written signatures have been shown to increase response rates, so typed signatures should be avoided if possible (Dillman 2000; Dialsingh 2008). If a sample of a few hundred respondents is being used, every effort should be made to use a handwritten signature. It may not be possible, however, to hand-sign the cover letter in a sample of a thousand or more surveys. In this case, a signature stamp or a pre-printed signature in a color other than black (preferably blue) can be used (Dillman 2000).

17.2.5 The Questionnaire

Each respondent should receive the exact same questionnaire (Dillman 2000). The questionnaire should be clear, concise, and esthetically pleasing in order to invite respondent participation (Bourque and Fielder 2002). The goal is to

include all of the questions necessary for collecting the desired information without making the survey so lengthy that it becomes burdensome to the respondent. Two recommended formats for keeping the number of pages to a minimum, while maintaining a pleasing design, are the newspaper format and the booklet format. A newspaper format can be achieved by printing two columns per page (Dialsingh 2008). Booklets can be created by printing two pages of the questionnaire on an 8½“ × 14” piece of paper (double sided), and then stapling the spine (Dillman 2000).

When organizing the questionnaire, use different sections to separate topics in an attractive and logical order. Provide ample instructions, and make sure each question and answer choice is clearly numbered (Dillman 2000; Bachman and Schutt 2007). Questions should be adequately spaced, rather than cramped, with plenty of white space in between. Simple fonts should be chosen over decorative lettering, and the layout and styling of the format should remain cohesive throughout the entire questionnaire (Dillman 2000; Dialsingh 2008; Wolf 2008). All questions, and pages, should be numbered (Dialsingh 2008). Questions should be arranged in a logical and conversational manner, with groupings of similar topic areas. The survey should signal a change in topic both verbally and visually, using necessary symbols and graphics (Dillman 2000; Wolf 2008).

Designing a survey with a clear concept in mind can greatly help in choosing the best sampling frame for the research question as well as in designing concise questions. If a survey that will successfully test your research question already exists, then use it—there is no need to “reinvent the wheel” (Bachman and Schutt 2007). The following general guidelines should be followed in surveys where the researcher is writing their own survey questions. Dillman (1978) recommends writing the first few questions of the questionnaire at the same time that the cover letter is written. Linking the purpose of the study that was stated in the cover letter to the introductory questions should serve as a natural segue into the questionnaire (Dillman 1978). Careful attention should be paid to the wording of questions, which come in both open and closed formats. Both options have advantages and disadvantages, which should be weighed against the goals of the research project (Fowler 1988).

Open-ended questions do not provide answer choices, allowing respondents to answer the question in their own words. In giving respondents this freedom in answering, open-ended questions can be advantageous because researchers may be provided with answers that had not occurred to them. In this way, the true views of the respondent may be captured rather than answers expected by the researchers (Fowler 1988; Bachman and Schutt 2007). Open-ended questions may also encourage participation. It can sometimes be frustrating for respondents to be unable to let their voices be heard, and open-ended questions provide an outlet for expressing full opinions (Fowler 1988). The two disadvantages to using open-ended questions are that the length of answer may be overwhelming to the respondent, as well as time consuming for the researcher to code and interpret (Bachman and Schutt 2007).

Close-ended questions, however, are the preferred format for most researchers for several reasons (Fowler 1988). Close-ended questions provide response choices to respondents, thus limiting the amount of interpretation inherent to open-ended questions. Limiting choices also allows for analytic categories to emerge, rather than scattered answers that are highly inconsistent. A variety of types of close-ended questions are available, such as Likert scales or simply listing different answers to choose from (Fowler 1988; Bachman and Schutt 2007). Close-ended questions, therefore, are less ambiguous and easier to statistically analyze but may not collect the total possible range of answers since respondents are limited to specific answer choices (Bachman and Schutt 2007).

The following are general recommendations for the writing of questions. Straightforward and uncomplicated words and phrases are always best (Dillman 2000). Make sure that the questions are very specific, and not vague or confusing.

Complete sentences should always be used, in a conversational tone. Use mutually exclusive response choices and always provide exhaustive response categories (e.g., providing "other", "don't know", or "undecided" as an options). Finally, be careful to avoid using unusual abbreviations, double negatives (e.g., 'do you disagree that drinking and driving should not be a crime?'), and double-barreled questions (i.e., asking two questions at once like 'do you agree that our government should abolish capital punishment and build more prisons?') (Dillman 2000; Bachman and Schutt 2007).

17.2.6 Returning the Survey

Surveys can be returned to the researcher in a variety of ways, and instructions should be provided to the respondent in their questionnaire packet. The most traditional method used to return the survey is through the mail. In this case, an addressed envelope should be provided with pre-paid postage, which has been shown to increase response rates. Sometimes the responses to mail surveys can also be entered through a method in addition to paper and pencil, such as on the Internet or through a toll free telephone number (Dillman 2000; Dialsingh 2008; Atkeson et al. 2011).

Careful attention should then be paid to the questionnaires that are returned to the researcher. Returned surveys should be immediately opened, and the contents cataloged and dated. It cannot be assumed that return envelopes will contain completed surveys. Incomplete, or even blank questionnaires, are often returned to the researcher. In some cases, the incentive will also be returned. Keeping track of what was returned from whom is important not only for data collection, but also for planning future follow-up correspondence (Dialsingh 2008).

17.2.7 Assembling the Final Packet

Placing the survey materials into the envelope to be mailed may seem like a trivial activity that does not necessitate very much attention to detail, but this could not be farther from the truth (Dillman 2000; Dialsingh 2008). The presentation and order of the survey materials inside the envelope can be very important in influencing the respondents' decision to participate (Dialsingh 2008). The goal is for every item in the survey packet to be withdrawn from the envelope simultaneously, with nothing left behind. There are several strategies for ensuring the simultaneous withdrawal of all materials, depending on the type of envelope used. If a standard business envelope is being used in combination with a questionnaire in booklet format, the cover letter should be folded in the standard three part-manner (bottom folded up one third of the way with the top folded over the remaining third of the way, covering the bottom portion). The booklet can then be folded in half with the front cover facing out, and inserted into the center of the folded cover letter. If an incentive is being offered, that should be placed inside the folded cover letter in front of the folded booklet. The return envelope should be placed behind the booklet (Dillman 2000).

If a questionnaire is printed in a newspaper column form on 8½" × 11" paper, then all of the pages of the questionnaire should be stapled together in the upper left hand corner. The cover letter, which remains unstapled, should then be laid on top. The cover letter and questionnaire should then be folded in the style of an accordion (or "*z*-fold"), with the letterhead facing out. Both the incentive and the return envelope can then be placed into the folds of the *z*-fold and inserted into the envelope. In cases where the survey packets are too thick to fold, and are thus placed in flat mailouts, stickers should be used to attach the incentive to the cover letter because using metal clips runs the risk of tearing the envelope (Dillman 2000).

17.2.8 Follow-up Correspondence

A few days to a week after the survey packet has been sent, a thank you card should be mailed to each respondent. The thank you card can serve as both a reminder to the respondents who have yet to complete and return their questionnaire and a token of appreciation to the respondents who

have already returned their completed questionnaire (Mangione 1995; Dillman 2000). For respondents who have failed to return their completed questionnaire within two to 4 weeks after the initial mailing of the questionnaire, a third mailing containing a replacement questionnaire should be sent. This replacement questionnaire should serve as a third reminder and request for the respondent to complete the survey (Mangione 1995; Dillman 2000). A modified cover letter is often included, as well as a second return envelope with a postage stamp (Dialsingh 2008).

Finally, if the respondent has failed to return his or her completed questionnaire 2–4 weeks after the replacement questionnaire has been sent, one final contact should be made (Mangione 1995; Dillman 2000) via special delivery, Federal Express, or priority mail. The method of contact should be different from the original regular mailing to impress the importance of the correspondence upon the respondent. Another option would be to place a reminder telephone call, but a fourth mailing is usually preferred (Dillman 2000). The entire survey process, including follow-up correspondence, should take 8 or 9 weeks (Mangione 1995).

17.3 Strengths and Weaknesses

Perhaps the most important benefit specific to mail surveys is that they are far less costly than other survey methods (Linsky 1975; Kanuk and Berenson 1975; Dillman 1978; Fowler 1988; Mangione 1995; Bourque and Fielder 2002; Bachman and Schutt 2007; Brill 2008; Dialsingh 2008). Completed mail surveys are half the cost of telephone surveys and 75% less than face-to-face interviews for a variety of reasons (provided that the length and survey objective are comparable in all three methods) (Bourque and Fielder 2002).

Paying survey research organizations to carry out the research is no longer necessary because the design and implementation of mail surveys is relatively simple (Dillman 1991). This substantially lowers the cost of mail surveys because less staff is needed to carry out the research (Fowler 1988; Bourque and Fielder 2002). In fact, depending on the size of the sample, it would be possible for one researcher to manage the entire study (Bourque and Fielder 2002). The distinct advantage to a small staff is less training is involved. Additionally, the facilities necessary to conducting a mail survey is much less than other methods because interviews are not being conducted on site (Fowler 1988). Mail surveys are also spatially cost effective in that they can cover expansive geographical areas (Linsky 1975) (both internationally and domestically), allowing for a much greater sample size to be reached relatively quickly (Kanuk and Berenson 1975; Fowler 1988; Bourque and Fielder 2002; Dialsingh 2008). Surveying large sample sizes also increases the probability of obtaining a representative sample (Bourque and Fielder 2002).

Mail surveys have the potential to evoke more thoughtful answers from respondents because they are given more time to privately complete the questionnaire at their own convenience, as opposed to the convenience of an interviewer. Privacy and convenience are especially important when questions of a sensitive nature are being asked (e.g., illegal behavior), and can lead to higher response rates (Linsky 1975; Kanuk and Berenson 1975; Mangione 1995; Bourque and Fielder 2002; Dialsingh 2008; Wolf 2008). Finally, unsupervised surveys generally are highly consistent instruments because each respondent is presented with the exact same stimulus (Bourque and Fielder 2002).

There are, however, some weaknesses associated with mail surveys. Historically, mail surveys have been plagued by low response rates, response bias, and nonresponse bias (Kanuk and Berenson 1975; Dillman 1978). Potential respondents may be lost because of incorrect addresses (Fowler 1988). Due to their unsupervised nature, researchers using mail surveys are unable to control whether the person who is actually responding to the questionnaire was the intended target (Bourque and Fielder 2002; Brill 2008). Portions of the sample may

not have the necessary literacy level to understand the questionnaire (Fowler 1988; Dialsingh 2008), which could lead to the survey being returned incomplete or not at all. On the other hand, the more educated a respondent is, the more likely they are to respond. This leads to selection bias (Fowler 1988).

Measurement error is also a potential problem for mail surveys because the researcher is not present to ensure that the respondent understands the intent of the question. For this reason, it becomes very important for researchers to pay careful attention to the wording of questions as well as the design of the survey (Fowler 1988; Wolf 2008). Skip patterns may be confusing, leading to unanswered questions and missing data. Once the survey is mailed, corrections or changes become very difficult to reliably make. Pre-testing the questions becomes very important in such situations (Brill 2008).

17.4 Validity and Reliability

As with all social science research, reliability and validity are central concerns in the construction of high quality mail surveys. Reliability in research refers to dependability or consistency, suggesting that the same results will be reached each time a measure is used under the same conditions (i.e., the method and results can be replicated). Validity in research simply means that we are reaching accurate empirical conclusions about an attribute through our choice in measurement (Groves 1987; Fink 2005; Bachman and Schutt 2007).

17.4.1 Reliability

The simplest way to ensure consistency is to provide each respondent with the exact same set of questions, in the same order, and with the same formatting. In doing this, the researcher is attempting to ensure that any differences in answers are due to the varying attitudes and beliefs of respondents, rather than differences in the questionnaire. Questions should be carefully worded in order to provide researchers with reliable answers that are predictive in nature. Complete, concise, and unambiguous questions with properly defined terms should be constructed so that each respondent is experiencing each question in the same way. Using simple words that respondents at all education levels can understand is recommended so that confusion is avoided (Fowler 1988).

17.4.2 Pre-Testing

Any type of survey should be pre-tested before they are administered (Mangione 1995; Presser et al. 2004; Fink 2005; Bachman and Schutt 2007). Conventional pre-tests are similar to a dress rehearsal—they are the only way to evaluate in advance whether a questionnaire causes problems for respondents. The assumption of the pre-test is that problems with the wording of questions will be revealed in practice surveys with respondents (Presser et al. 2004; Fink 2005). The first step, however, is for the researcher to critically read each of their questions and revise as needed. After this first round of revisions, the questionnaire should be given to colleagues, students, and/or friends who are willing to help. Their feedback can be a useful method of determining how long the questionnaire will take respondents to complete, as well as the clarity of the questionnaire (Mangione 1995; Bachman and Schutt 2007).

If possible, formal pre-tests (i.e., a pilot study) should then be conducted. This pre-test can be conducted through the mail, like the actual survey will be, but administering the survey to a focus group is preferable. Being present as the group takes the survey allows the researcher the opportunity to receive a wide range of feedback on the questions, such as which items are unclear or are not conveying the proper intent (Mangione 1995). If it is not possible for the researcher to pre-test in a focus group, mailed out pre-tests can be scanned for indicators of problematic questions. For example, counting both the number of answers that were left blank and ‘don’t know’ responses can indicate which questions may have been unclear to the respondent (Presser et al. 2004).

17.4.3 Validity

There are four types of error related to the validity of a survey: measurement error, sampling error, coverage error, and nonresponse error (Groves 1987; Dillman 2000).

17.4.4 Measurement Error

Measurement error occurs when the answer that a respondent gives is inconsistent with their characteristics (Groves 1987), inaccurate, or too vague (Dillman 2000). In mail surveys, errors in measurement are most often attributed to the design and wording of the questionnaire, and mistakes made by the respondent (Dillman 2000). A great deal of research has focused on improving the quality of the questionnaire in order to reduce measurement error. Paying careful attention to the layout of the questionnaire can help the researcher maximize organization, thus avoiding overlooked questions and incomplete answers. Designing a simple and esthetically pleasing questionnaire can motivate the respondent to complete the entire survey (Dillman 2000).

In mail surveys, the question asked is the method of measurement; therefore the answers provided by respondents must be reflective of the intent of the question. There are differences in measuring the validity of objective survey questions and subjective survey questions. Objective survey questions explore matters that can be independently verified (e.g., through criminal or medical records). Subjective survey questions, however, are aimed at the discovery of opinions, beliefs, and attitudes, all of which are incapable of being validated objectively due to their highly individual nature (Fowler 1988). The best way to evaluate the validity of subjective survey questions, therefore, is through measuring a concept several times throughout the body of the survey, in a variety of different ways. Answers can then be evaluated to determine correlation, and thus validity (Groves 1987; Fowler 1988; Mangione 1995). As was discussed in reference to reliability, pre-testing the questionnaire is also a method of reducing measurement error (Mangione 1995).

17.4.5 Sampling Error

When a survey is deliberately provided to some members of a population and not others, *sampling error* occurs (Groves 1987). Sampling error can be reduced relatively easily by drawing a large and random sample (Groves 1987; Mangione 1995; Dillman 2000), as well as through statistical modeling (Groves 1987). A good rule of thumb to keep in mind is that if the sample size is quadrupled, then the sampling error can be reduced by one half (Groves 1987; Mangione 1995). Finally, as with coverage error, sampling error can be reduced by ensuring that the sampling frame covers all of the characteristics of the population. This can be done by checking that the list of potential respondents is unbiased and complete (Mangione 1995)

17.4.6 Coverage Error

The samples that mail surveys are drawn from are usually taken from lists of addresses in a particular area (Dillman 2000). If the list that is provided to the researcher does not encompass all of the elements of a specific population then *coverage error* occurs, meaning that not every person in a particular sample was included in the survey (Dillman 1991; Groves et al. 2000). For many surveys, coverage error is not a problem because the addresses of specifically targeted populations are easily available (e.g., members of organizations, homeowners, or individuals with a drivers license). When surveys of the general public are conducted, however, coverage error can be a major problem because all-inclusive lists of the total population do not exist (Groves 1987; Dillman 2000.) For example, individuals who are transient or economically disadvantaged are often difficult to locate and are thus unaccounted for in household surveys (Groves 1987).

There are two methods of reducing coverage error in mail surveys. One possible method is to use information on the distribution of race, age,

and gender in the population provided by the Census Bureau to make statistical adjustments to the data (Groves 1987). Another approach is to carefully evaluate the list of the members of a sample that has been provided. It is sometimes feasible to determine if there are gaps in the list, and then find alternate sources of gathering the missing information (Dillman 2000). It becomes very difficult to assert causal relationships within the data if certain elements remain unexplored in the sample, therefore every effort should be made to avoid coverage error (Groves 1987).

17.4.7 Nonresponse Error

Nonresponse error occurs when less than the total amount of surveys that were originally distributed are completed and returned (i.e., the survey did not measure everyone in the sample), making it the biggest weakness of mail surveys (Groves 1987; Dillman 1991; Mangione 1995). Nonresponse is a problem because there are often differences between the respondents who completed the survey and those who did not, thus creating biased results (Mangione 1995; Dillman 2000). The size of the bias depends on the number of nonresponders, and the amount of difference between the two groups. Determining the size of the bias is very difficult because the researcher has very little power in gathering information about nonresponders in order to compare those characteristics to individuals who did respond. Knowing these characteristics is especially important if the opinion of responders is significantly different from that of nonresponders because a biased population average will be reported. Low response rates also lower the apparent quality of the research (Fowler 1988; Dillman 1991; Mangione 1995).

17.5 Response Rates

Response rate is a special concern for mail surveys (Bachman and Schutt 2007) because the interviewer is not present to ensure that the questionnaire is completed and returned (Mangione 1995; Dialsingh 2008). Response rates are determined by dividing the number of completed and returned surveys, by the total number of possible surveys that were originally mailed out (Fowler 1988; Fink 2005). A response rate between 70 and 80% is the goal, with 70% being minimally acceptable and a response rate above 80% being the ideal, although highly unlikely. If the response rate falls below 60% it becomes very difficult to claim that the survey is representative of the entire sample because responses are missing from more than one third of surveyed individuals (Mangione 1995; Bachman and Schutt 2007). Information gained from surveys with response rates <50% are "not scientifically acceptable" (Mangione 1995).

Focusing on the improvement of response rates is very important. From a budgetary perspective, having a high response rates saves a great deal of money because less follow-up contacts are necessary. Methodologically speaking, having a large sample size increases the generalizeability of the results while nonresponse bias becomes less of a concern (Fox et al. 1988). Low response rates can, therefore, be addressed through a variety of methods at every stage of the survey process. The following techniques are frequently used approaches that have been proven to be empirically significant in raising response rates.

17.5.1 The Survey Packet

The first impression of the survey packet is a lasting one. Even the tiniest details, such as using postage stamps on the outer envelope of the survey packet, have been shown to increase response rates (Linsky 1975; Fox et al. 1988; Brill 2008). Personalizing all correspondence (e.g., using the respondent's name and a handwritten signature) has proven effective in raising response rates by five to eight percentage points (Dillman 2000). Simply providing a return envelope (Yammarino et al. 1991) can improve a response rate by 2–4% (Dillman 2000). Using a postage stamp (rather than using a business reply envelope without a stamp) increases response rates even more (Yammarino et al.

1991; Brill 2008). It is often suggested that placing a stamp on the return envelope subtly implies that money was spent buying postage stamps, and that money will go to waste if the questionnaire is not returned (Linsky 1975; Mangione 1995).

Using a concise and motivating cover letter that grabs the attention of the respondent (Mangione 1995; Brill 2008), while making an appeal for their cooperation, is extremely helpful in raising response rates (Yammarino et al. 1991). If the study is sponsored by another organization, mentioning the sponsorship in the cover letter also increases response rates (Fox et al. 1988). The most important element of the survey packet in increasing response rates, however, seems to be the design and content of the questionnaire (Mangione 1995; Brill 2008).

A variety of design techniques have been shown to lower response rates. Mail surveys that appear to be professional and scientific greatly increase participation (Fowler 1988; Brill 2008). Clear instructions that are easy to follow increase response rates because they eliminate frustration and confusion that lead respondents to fail to complete the survey (Fowler 1988; Mangione 1995). Close-ended questions with answer choices encourage more responses than open-ended questions (Brill 2008). Using pretests to eliminate questions that are unnecessary, redundant, unclear, difficult to answer, or cause excessive discomfort can result in the return of more completed questionnaires, thus raising the response rate (Mangione 1995; Fink 2005). Individuals who feel that their responses will remain confidential are more likely to respond. As was mentioned earlier, using identification numbers instead of names on all survey materials increases respondent confidence that confidentiality will be upheld (Mangione 1995; Fink 2005).

17.5.2 Questionnaire Length

Empirical evidence from published research is ambiguous on the effect of questionnaire length on response rate in mail surveys. Some earlier studies have concluded that survey length did not influence response (Brown 1965; Bruvold and Comer 1988; Mason et al. 1961), other studies suggest the opposite: the lengthier the questionnaire, the lower the response rate (Steele et al. 1992; Yammarino et al. 1991; Dillman et al. 1993; Jones et al. 2000; Sahlqvist et al. 2011). As a general rule of thumb, the shorter the survey, the higher the response rate will be (Linsky 1975; Groves 1987; Mangione 1995).

Properly trimmed down questionnaires had been found to increase the response rate by 10% (Jones et al. 2000) and to raise the odds of response by 50% (Sahlqvist et al. 2011). A recent word count analysis of postal survey indicated that questionnaires above the threshold of 1,000 words averaged a lower response rate than those below it (38.0% vs. 59.4%) (Jepson et al. 2005). Research investigating the number of pages most conducive to high response rates has been inconsistent, and therefore inconclusive, due to the myriad differences in sample sizes and demographics, research questions, and survey design and procedure. For example, Yammarino et al. (1991) found a 6.5% increase in response rates for using questionnaires that are four pages or less, while Dillman (2000) suggests that the length of the survey is not an issue at all if other techniques of increasing the response rate are used.

Background characteristics of respondents may also condition their reaction to survey length: For example, professionals working in corporate settings may be less responsive to lengthy mail questionnaires than home-based individual consumers (Jobber and Saunders 1993; Tomasokovic-Devey et al. 1994). The important thing to remember is that efficiently designed surveys will naturally be shorter (Mangione 1995).

17.5.3 Interest

While it is generally assumed that ensuring that the survey is of interest and value specific to the respondent will naturally raise response rates (Dillman 1978; Groves et al. 1992; Couper and Groves 1996; Greer et al. 2000; Bourque

and Fielder 2002; Kropf and Blair 2005), few studies have empirically tested this assumption. Heberlein and Baumgartner (1978) found that response rates almost doubled when the questionnaire was salient for the respondent. In a study comparing surveys using follow-up cover letters that discussed interest specific to the respondent and benefits to society as a whole, Kropf and Blair (2005) found that both increased response rates (implying that a discussion of interest in general does in fact increase cooperation). Cover letters enumerating the overall benefits to society resulted in response rates that were 13% higher than those discussing self-interest.

Some studies have suggested that the respondent's interest in the content is the most important technique for increasing response rates (Greer et al. 2000), especially when the respondents' beliefs, attitudes, and values are linked to the survey (Groves et al. 1992). Respondents may find it gratifying to demonstrate knowledge on a particular topic of interest, especially when they feel that they will personally benefit from information collected in the questionnaire (Couper and Groves 1992). Framing the appeal to cooperate as a rare opportunity for the respondent to "make their voice heard" on an issue of importance in their lives may also increase response rates (Groves et al. 1992).

17.5.4 Social Worthiness

Most surveys appeal to the altruistic nature of the respondent by highlighting value of the questionnaire to society, and the benefit that can be gained by recording the respondents' opinion (Linsky 1975; Dillman 1978). Empirical findings on the effect of the social worthiness of a questionnaire actually increases response rates are, however, inconsistent. Appealing to the respondents' sense of civic duty has been shown to increase response rates, especially when the sponsoring organization is a recognizably important social institution (e.g., government or university) (Couper and Groves 1996). Mowen and Cialdini (1980) found that by simply adding the phrase "it would really help us out" to the end of the description of the survey resulted in response rates increasing from 25 to 43.7%. Linksy (1975), however, found that making a social utility argument does not increase response rates at all.

Similarly, more respondents will cooperate with surveys that promote a sense of social cohesion between the respondent and the sampled population (Couper and Groves 1996). Providing social validation to respondents by appealing to their membership to a particular group, and mentioning that others within their group have already responded, can also increase response rates (Groves et al. 1992; Dillman 2000). In other words, "one should be more willing to comply with a request to the degree that one believes that similar others would comply with it" (p. 482), and it becomes the researcher's responsibility to convey that others have already responded (Groves et al. 1992).

17.5.5 Incentives

Monetary incentives serve as tokens of appreciation from the researcher to the respondent. Research has consistently shown that mail surveys that provide incentives have much higher response rates than those that do not (Linsky 1975; Kanuk and Berenson 1975; Yammarino et al. 1991; Dillman 2000; Trassell and Lavrakas 2004; Fink 2005). While there are a wide variety of methods for implementing incentives in a mail survey (e.g., small and inexpensive tokens like pens, notepads, or raffle tickets) (Bourque and Fielder 2002), cash appears to be the most commonly used incentive (Dillman 2000; Trassell and Lavrakas 2004; Dialsingh 2008). Incentives can be used in two different ways. One possibility is to send a reward after the completed survey has been returned. The promise of the reward for their cooperation is the motivating factor, therefore the reward must be something that the respondent will highly value (Mangione 1995).

The second option is to provide the reward before the survey is completed, in which case the incentive would be placed in the survey

packet. This method may seem counterintuitive, but is actually the more successful of the two incentive strategies. (Linsky 1975; Mangione 1995; Dillman 2000). A small noncontingent incentive provided with the pre-notification letter has consistently been found to increase response rates more than the promise of a more valuable future contingent incentive (Mangione 1995; Brill 2008; Dialsingh 2008).

The resulting high response rate using this method is often explained by social exchange theory, which purports that providing an incentive that is not contingent on the completion of the survey prompts a feeling of obligation to the researcher (i.e., that common courtesy mandates the return of the questionnaire because a gift was provided) (Linsky 1975; Dillman 1978; Trassell and Lavrakas 2004). For social exchange theory to work properly, the incentive amount must remain small. If the incentive amount begins to approach the actual value of the requested task, then the respondent will feel less obligated to the researcher because the token begins to look more like an economic exchange. In this case, more money will be expected for the time they spend completing the questionnaire (Dillman 1978, 2000).

The obvious disadvantage to providing the incentive prior to the completion of the survey is that respondents may choose to take the money without returning the questionnaire. This is an issue that has been extensively empirically studied because researchers want to know the lowest monetary incentive that is the most effective in increasing response rates (Mangione 1995). Varying results have been found when investigating the best monetary amount to use for the incentive. These inconsistent empirical results could be due to situational variables inherent to the specific design of each study (Yammarino et al. 1991), therefore the following suggestions for incentive amounts should be used as general guidelines and evaluated on an individual basis.

Using $1 or less as an incentive is the most popular recommendation (Linsky 1975; Fox et al. 1988; Dillman 2000; Dialsingh 2008), with a quarter being the smallest amount reported as successfully increasing response rates (Linsky 1975). In a study investigating the benefits of using different incentive amounts, Trussel and Lavrakas (2004) provide interesting guidelines for choosing which amount should be sent. They found no significant difference in using $1, or $2 to $4, so if the research budget allows only for incentives under $5, then $1 should be used. Sending incentives in amounts ranging from $5 to $8 result in the same response rate, suggesting that using $5 (rather than $6, $7, or $8) is the better strategy for budgetary purposes. A later study conducted by Kropf and Blair (2005) found similar results when comparing $1 incentives to $5 incentives, with a 7% point increase in response rate for the $5 incentive.

While using a $10 incentive does result in the highest response rate, and using $5 results in slightly higher response rates than using $1, these differences are not statistically significant. In sum, it appears that their results support the popular recommendation of using $1 (Trussel and Lavrakas 2004), because as the size of the incentive increases, a diminishing returns effect begins to occur (Fox et al. 1988). In other words, sending incentives up to the amount of $10 does increase response rates, but this increase may not be enough to justify the additional budgetary strain (Trassell and Lavrakas 2004; Kropf and Blair 2005, also see Toepoel Chap. 13).

17.5.6 Pre-Notification and Follow-Up Correspondence

In general, response rates improve with repeated contact with the respondent (Yammarino et al. 1991). Pre-notification letters have been shown to increase response rates by 8% simply by alerting the respondent that a questionnaire is on the way (Fox et al. 1988), thus increasing the chances that the questionnaire will be opened rather than mistaken for junk mail (Linsky 1975). Follow-up mailings are intended to address nonrespondents in an effort to remind and encourage them to complete and return the questionnaire, thus making them an important solution to an initial low response rate (Linsky 1975; Kanuk and Berenson 1975; Fox et al. 1988;

Yammarino et al. 1991; Fink 2005; Dialsingh 2008). Both postcards and letters have been found effective in the first two to three follow-up attempts. After two failed attempts, however, a different mode of contact is recommended. Using a special delivery mail service or telephone call as the final follow-up has been found to increase response rates more than the use of first class mail. Furthermore, if the researcher must choose between a pre-notification and follow-up mailing due to budgetary reasons, the follow-up correspondence should be chosen because it is more effective in raising response rates (Linsky 1975).

Each subsequent mailing results in diminishing returns, from both a cost and response rate perspective (Dialsingh 2008). In general, the response rate for all subsequent follow-up correspondence is usually half of the prior return rate (e.g., if 40% of the surveys were originally returned, you can expect to receive 20% back after the first follow-up letter, 10% after the second, etc.). For this reason, no more than three follow-up mailings should be sent for a response rate of 75% or higher (40 + 20 + 10 + 5%) (Mangione 1995).

17.6 Conclusion

Mail surveys provide an efficient, inexpensive, and relatively simple method of collecting large amounts of data from diverse populations in various geographic locations (Linsky 1975; Kanuk and Berenson 1975; Dillman 1978; Fowler 1988; Bourque and Fielder 2002; Dialsingh 2008). Although there are a variety of potential weaknesses in using mail surveys, mail surveys have recently proven to achieve results that are equal to, if not better than, face-to-face and telephone interviews when executed correctly (Brill 2008). Researchers in a variety of fields are constantly investigating methods through which the design and implementation of mail surveys can be improved (Dillman 1991), which will make mail surveys continue to be appealing to future generations of researchers.

References

Atkeson, L. R., Adams, A. N., Bryant, L. A., Zilberman, L., & Saunders, K. L. (2011). Considering mixed mode surveys for questions in political behavior: Using the internet and mail to get quality data at reasonable costs. *Political Behavior, 33*, 161–178.

Bachman, R., & Schutt, R. K. (2007). *The practice of research in criminology and criminal justice* (3rd ed.). Thousand Oaks, CA: Sage.

Bourque, L., & Fielder, E. (2002). *How to conduct self-administered and mail surveys*. Thousand Oaks, CA: Sage.

Brill, J. E. (2008). Survey Research Methods. *SAGE Publications*. 27 April 2011. http://www.sage-ereference.com/survey/Article_n281.html.

Brown, M. (1965). Use of a postcard query in mail surveys. *Public Opinion Quarterly, Winter, 29*, 635–637.

Bruvold, N. T., & Comer, J. M. (1988). A model for estimating the response rate to a mailed survey. *Journal of Business Research, 16*(2), 101–116.

Couper, M. P., & Groves, R. M. (1992). Interviewer reactions to alternative hardware for computer assisted personal interviewing. *Journal of Official Statistics*, *8*, 201–210

Couper, M. P., & Groves, R. M. (1996). Household-level determinants of survey nonresponse. *New Directions for Evaluation, 70*, 63–79.

Dialsingh, I. (2008). Mail Survey. *Encyclopedia of Survey Research Methods*. 27 April 2011. http://www.sageereference.com/survey/Article_n282.html.

Dillman, D. (1978). *Mail and telephone surveys: The total design method*. New York, NY: Wiley.

Dillman, D. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225–249.

Dillman, D. (2000). *Mail and Internet surveys: The tailored design method*. New York, NY: Wiley.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent friendly design, and a difficult question on response rates for occupant- addressed census mail surveys. *Public Opinion Quarterly, 57*(3), 289–304.

Erdos, P., & Morgan, A. J. (1970). *Professional mail surveys*. New York, NY: McGraw-Hill Book Company.

Fink, A. (2005). *How to conduct surveys: A step by step guide*. Thousand Oaks, CA: Sage.

Fowler, F. J. (1988). *Survey research methods*. Thousand Oaks, CA: Sage.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*, 467–491.

Greer, T. V., Chuchinprakarn, N., & Seshardri, S. (2000). Likelihood of participating in survey research:

Business respondents' perspectives. *Industrial Marketing Management, 29*, 97–109.

Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly, 51*(2), 156–172.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475–495.

Groves, R. M., Singer, E., & Corning, A. C. (2000). Leverage-salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly*,*64*, 299–308

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43*, 447–462.

Jepson, C., Asch, D. A., Hershey, J. C., & Ubel, P. A. (2005). In a mailed physician survey, questionnaire length had a threshold effect on response rate. *Journal of Clinical Epidemiology, 58*(1), 103–105.

Jobber, D., & Saunders, J. (1993). A note on the applicability of the Brurold-Comer model of mail survey response rates to commercial populations. *Journal of Business Research, 26*, 223–236.

Jones, R., Zhou, M., & Yates, W. (2000). Improving return rates for health-care outcome. *Psychological Reports, 87*(2), 639–642.

Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*, 440–453.

Kropf, M. E., & Blair, J. (2005). Eliciting survey cooperation: Incentives, self-interest, and norms of cooperation. *Evaluation Review, 29*(6), 559–575.

Linsky, A. S. (1975). Stimulating responses to mailed questionnaires: A review. *Public Opinion Quarterly, 39*(1), 82–101.

Mangione, T. W. (1995). *Mail surveys: Improving the quality*. Newbury Park, CA: Sage.

Mason, W., Dressel, R., & Bain, R. (1961). An experimental study of factors affecting response to a mail survey of beginning teachers. *Public Opinion Quarterly, 25*(Summer), 296–299.

Mowen, J. C., & Cialdini, R. B. (1980). On implementing the door-in-the-face compliance technique in a business context. *Journal of Marketing Research, 17*(2), 253–258.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., et al. (2004). Methods for testing and evaluating survey questions. *The Public Opinion Quarterly, 68*(1), 109–130.

Sahlqvist, S., Song, Y., Bull, F., Adams, E., Preston, J., & Ogilvie, D. (2011). Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: A randomised controlled trial. *BMC Medical Research Methodology, 11*, 62.

Steele, T. J., Schwendig, W. L., & Kilpatrick, J. A. (1992, March/April). Duplicate responses to multiple survey mailings: A problem? *Journal of Advertising Research*, *37*, 26–34.

Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organisational survey non-response. *Administrative Science Quarterly, 39*, 439–457.

Trassell, N., & Lavrakas, P. J. (2004). The influence of incremental increases in token cash incentives on mail survey response. *Public Opinion Quarterly, 68*(3), 349–367.

Wolf, J. (2008). Self-Administered Questionnaire. *Encyclopedia of Survey Research Methods*. SAGE Publications. 27 April 2011 http://www.sageereference.com/survey/Article_n522.html.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly, 55*, 613–639.

E-Mail Surveys

18

Gustavo Mesch

The increasing public adoption of the Internet has resulted in a growing segment of the population that has an e-mail address and frequently uses it for sending and receiving messages. Social research has taken note of this development and scholars and practitioners had explored in the past 20 years the possibilities involved in data collection using e-mail surveys. The purpose of this chapter is to discuss the advantages and limitations in the use of e-mail surveys and, to present an integrative and up to date summary of previous studies.

18.1 E-Mail Surveys Defined

E-mail surveys are defined as surveys in which a computer plays a major role in the recruitment of potential respondents, delivery, and response of the questionnaire, and the collection of data from actual respondents (Jansen et.al. 2006). In this sense, the central characteristic of an e-mail based surveys is that survey instruments are delivered to potential respondents through electronic mail applications over the Internet (Kiesler and Sproull 1986). Typically an e-mail survey is delivered either in the an e-mail body text or as an attachment file to an e-mail address and includes a letter of invitation to participate. The respondent needs to either introduce the answers to the e-mail body or to insert the answers in the attached file and send it back through the use of an e-mail message.

E-mail based surveys, compared to face-to-face, phone, and mail surveys, are perceived to be a cheaper method of data collection. A central reason for this perception is the lower cost involved in the administration and delivery of the survey, as e-mail surveys can be delivered to a very large sample of Internet users electronically. At the same time, it should be noted that other costs involved in any survey are not saved using an e-mail mode. As any other survey, e-mail surveys involve costs of survey design, sample recruitment, data entering, data coding, and analysis. Furthermore, similar to mail surveys, e-mail surveys require the use of follow-ups and incentives.

It is important to make a distinction between e-mail and Web-based surveys. Web-based surveys are defined as instruments that physically reside on a network server and that can be accessed through a Web browser only (Borkan 2010; Tse and Fan 2008). The role of e-mail in a Web-based survey is important only at the stage of recruitment of potential respondents. It is a common practice in Web mail surveys to send an e-mail message of invitation to participate with an explanation of the study goals. Then, the respondent is referred to a Web link and is asked to self-complete the survey. Web surveys share some of the advantages and weaknesses of e-mail surveys but they also have different

G. Mesch (✉)
Department of Sociology and Anthropology,
University of Haifa, Haifa, Israel
e-mail: Gustavo@soc.haifa.ac.il

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_18, © Springer Science+Business Media New York 2012

methodological advantages and pitfalls. For this reason in this chapter we restrict our discussion to e-mail based surveys only.

18.2 E-Mail Survey Applications

The use of e-mail for conducting surveys goes back to the early days of Internet adoption and use. At that time, the percentage of Internet users in the population was small and highly active in electronically-based forums. Online activity and computer-mediated communication were a central topic of research in the social sciences. Scholars were attempting to learn and understand the rules of computer-based communication, computer-based participation in work groups, and participation in virtual forums. An important study investigated how often personal relationship form in newsgroups (forums in which individuals participated reading, sending, and posting messages by e-mail). The study used a two-stage sampling procedure. In the first stage, 24 newsgroups were randomly selected from published lists of groups in each of four major Usenet newsgroup hierarchies. In the second stage, 22 people were randomly chosen from lists of those who had posted messages to these groups over a several-day period. E-mail surveys were then sent to prospective participants by direct e-mail with a response rate of 33.3%. The study reported that contrary to existing theories, personal relationships do develop online and migrate to face-to-face settings (Parks and Floyd 1996). Others studied non-normative online behaviors such as flaming, individuals privacy concerns as related to e-commerce, and perceived online social support (Sheehan and Hoy 1999). In the early period, this approach for data collection was limited to the collection of self-reported data from selected groups that shared a common interest and were using similar digital platforms, such as political attitudes surveys of newsgroup users, electronic forum, and virtual social support group members. Studies were mostly restricted to the study of the effect of computer-mediated communication features or the understanding of the role of virtual communities in social support. As access and use of the Internet expanded to more groups of the population and was adapted at the workplace, e-mail surveys are used for diverse purposes including health behavior, market research, consumer behavior, and organizational studies.

18.3 E-Mail Surveys Taking Advantage of Internet Characteristics

The use of e-mail surveys take advantage of the unique features of the Internet, mainly the ability to easily contact a large number of potential respondents that are geographically dispersed all over the world. In that sense, an e-mail survey facilitates to the researcher, through a single e-mail sent to a large list of e-mail addresses previously compiled, a letter of invitation and the questionnaire itself to reach a large number of potential respondents. Furthermore as the e-mail questionnaire is a self-completed instrument, the use of an e-mail survey reduces the costs of training interviewers, the use of interviewers, the costs involved in printing questionnaires in paper, and the payment of mail postage. The global reach of the Internet, makes e-mail surveys a well-suited methodology for collecting data from different countries.

According to the Internet World statistics it is estimated that in 2010 there were 1.9 billion users, this is, about 28% of the total world population (http://www.internetworldstats.com/stats.htm). The adoption rate differs by continent and countries, with more industrialized countries reporting higher rates. Taking into account this distribution of Internet access, e-mails can be used to gather data from a large number of respondents on values, attitudes, and behaviors from diverse countries and conduct comparative studies (Mehta and Sivadas 1995).

E-mail surveys are particularly appropriate for conducting studies on self-reported online behavior, patterns of use, and trust in websites, patterns of online information search, patterns of participation in online-based social networking sites and open forums. E-mail surveys are well

suited to reach an Internet sample and to inquire on a large variety of online behaviors, to describe it and to measure its effect on the effect of the Internet on social behavior (Elford et al. 2004; Ross et al. 2000; Li et al. 1999; Bhatnagar et al. 2000).

E-mail surveys are particularly appropriate for conducting studies on sensitive topics (drug use, sexual behavior, health conditions, and participation in sub-cultures). Research in computer-mediated communication has shown that characteristics of the online environment (such as perceived anonymity and the lack of socio-demographic identifiers) support the fast and accurate disclosure of personal and intimate information. It is very likely that e-mail studies on sensitive topics will produce more accurate and rich data than face-to-face and phone interviews.

Regarding the use of e-mail surveys for conducting qualitative studies, the initial approach of researchers is that e-mail is not well suited for these studies. In qualitative research, it is usually assumed that coverage, sampling, and non-response errors are not the important issues they are in quantitative research. The goal of qualitative research is more to understand a phenomenon than to make inferences to the population and researchers are more concerned with the quality of the information that is gathered than sampling issues. The main rationale was that e-mail is a lean medium unable to transmit important cues of communication such as gestures, intonation of voice, non-verbal expression of emotions, and gestures. For that reason, one of the main concerns over the use of e-mail survey for qualitative research was the quality of the data itself. The non-verbal cues, that cannot be fully transmitted through the Internet, are an integral part of collecting qualitative data. In addition, there is concern that online participants may find it difficult to set down their feelings or ideas in writing (Tse 1999).

More recently, (Coderre et al. 2004) argued that in e-mail qualitative studies, participants might be able to read and reread their responses; their answers are likely to be better thought out and may be less spontaneous. The inability of interviewers to capture nonverbal cues may limit the information elicited from participants, because a higher level of commitment is needed to complete an online-based interview. Of particular interest are the responses to open ended questions. Despite these concerns, studies have shown that e-mail recipients are more likely than mail recipients to respond to open ended questions and to add comments to their answers, to report their behavior in sensitive issues and to express their feelings (Bachman et al. 2000; Taylor 2000). In a study that compared 24 semi-structured interviews conducted face-to-face with the same number of semi-structured interviews conducted on the Web, Folkman-Curasi (Folkman-Curasi 2001) found that the online interview contained both the strongest and the weakest transcripts, they were less detailed and they were far more grammatically correct and seemed less spontaneous, that is the sentences had a clear beginning and end, with none of the "ummmhing" and "ahhing" associated with face-to-face interviews. Coderre et al. (2004) examined in their study 3 groups of 40 subjects selected from a panel of households. The first group had internet access and the other 2 groups were selected to match all the socio-demographic characteristics of the ones with internet access. Data was collected via the Internet through an e-mail survey, by mail and telephone. Respondents were asked to answer two open-ended questions on subject beliefs regarding a firm and the reasons for their evaluation. Answers to the open-ended questions were compared according to the number of items mentioned per item, the total number of different items, and the number of unique items. The results indicated that there were no major differences when the answers were compared. The only item in which there was a difference was the total number of different items. More items were mentioned in mail than e-mail and phone surveys. Furthermore, they used the data gathered from the three modes of qualitative data collection and checked their predictive validity. The results indicated that the predictive validity of the information gathered through the e-mail survey was equal or superior to that observed with other two data collection methods. Thus, from these studies, there is strong

evidence that the use of e-mail surveys for gathering qualitative data using open-ended questions is appropriate.

18.4 Steps in the Administration of an E-Mail Survey

As in any survey, the administration of an e-mail survey involves a series of stages. As in any survey the researcher need to define the relevant research questions and the items that will be used to measure the different theoretical concepts of interest. After these critical decisions have been made, there are a number of steps involved in the administration of an e-mail survey.

Step 1: Survey Pretest E-mail surveys include items that are self-completed by the respondent without the participation of a trained interviewer that can clarify the items of the survey. For this reason, typically e-mail surveys require that after they are written a pretest, with a selected group of respondents of the target population, be conducted. The goal of this step is to assess the validity and comprehension of the items. In addition, a pretest helps to get an estimate of the amount of deliverable surveys and, based on this data, to estimate the adequate sample size to achieve the response rate goals of the study (Sheehan and Hoy 1997).

Step 2: Sampling The difficulties involved in creating a sample frame depend on the target population. E-mail surveys had been widely used to collect self-reported data from populations that can be identified and their e-mail addresses accessed. A large body of research was conducted among college students, staff of health educational organizations, financial corporations, and firm customers (Couper et al. 1999; Jones and Pitt 1999; Dillman 1998; Sheehan and Hoy 1997).

In recent years, there is an increase in the attempts to use e-mail surveys to cover more representative samples of the total population. A recent approach is the use of online panels as an alternative for the creation of a sample frame from which many sub-samples of population groups with a known e-mail address can be drawn. Later on in this chapter online panels will be discussed in more detail (Goritz 2004).

Step 3: Solicitation E-mail surveys are sent by e-mail and nowadays run the risk of being classified as "junk mail" by the computer system or the user. For this reason, an important step in recruiting participants is to send a solicitation letter. The letter is sent usually a few days before the questionnaire itself is sent to the potential participants.

The solicitation e-mail provides potential respondents with the chance to opt out, that is, not receive the survey. Potential respondents who opted out can be replaced by other names from the sample pool (Mehta and Sivadas 1995).

Step 4: Administration Usually a week after the notification message, the actual survey is e-mailed to the sample. Instructions on completing the survey are needed including the option of returning the survey either via e-mail or postal mail. Because e-mail messages contain the name of the respondent included on the message, the researcher needs to inform the respondent that confidentiality (but not anonymity) is warranted. There is no significant evidence that the lack of anonymity causes a decrease in response rates (Kanuk and Berenson 1975).

Step 5: Follow-up It is a common practice in e-mail surveys to send follow-ups to those in the sample who did not respond within a week. A reminder message via e-mail, which included an additional copy of the survey, is sent. These follow-ups have been widely used in postal mail surveys, and tend to increase response rates.

18.5 Major Advantages of E-Mail Surveys

Research costs The costs involved in conducting large-scale surveys are high. In face-to-face and phone surveys a central component of the price are the costs involved in the recruitment and training of interviewers. E-mail surveys are similar in their administration and costs to mail surveys. After the design of the questionnaire, there are no costs involved in the administration

of the instrument as typically an e-mail survey involves the electronic sending of the instrument by electronic means (either as an attachment or in the e-mail body). In sum, one of the major advantages of using an e-mail survey for collecting data is the reduction in the costs involved in data collection.

Respondent convenience An important feature of e-mail surveys is the convenience for the respondent. Different from phone and face-to-face surveys that often require previous contact with the respondent to arrange for a day and time for conducting the interview, e-mail surveys are more convenient. Respondents may take as much time as they need to answer the questions. E-mail surveys can allow the respondent to skip answers and to go back to questions later on. Thus, instead of being annoyed at an inconvenient time with a phone survey or the visit of an interviewer, a respondent can take an online survey whenever he or she feels it is convenient (Hogg 2003).

Time framework Online surveys can be administered in a time efficient manner, minimizing the time that takes to get a survey into a field and the time needed for data collection. Regarding getting the survey in the field, after collecting the relevant e-mail addresses, and writing the questionnaire, there is no need for any further activities (such as interviewers recruitment and training) that constrain the timing of survey administration. Studies have found that once e-mail surveys are sent to the respondents, responses start arriving in a few hours after it has been sent to the potential respondents. As a matter of fact, comparative studies have shown that response time to e-mail surveys is faster than in regular mail surveys.

Speed of response A central issue that researchers face when conducting a survey is the speed of response, i.e., the time that takes to collect the study data. This time is usually measured from the day that the administration of the survey started until the day that the last completed questionnaire arrives to the researchers. Speed of response is important for various reasons. There are studies that attempt to gather data on attitudes or behaviors immediately after a central event. Speed of response is important as the closer the collection of the data to the event, the higher the likelihood of isolating behaviors that can be directly related to the influence of the event. In cases that an event does not play an important role, still there are reasons to expect rapid data collection, among them the possibility that some of the respondents will be influenced in their responses by a critical social event, which was not experienced evenly by all the respondents. Finally, rapid response allows the researcher to move on to the next steps of data coding and analysis increasing the ability of a timely report of the study findings.

The length of time it takes to field a survey is a function of the contact, response, and follows-up modes. Decreasing the time in one or more of these parts of the survey process tend to decrease the overall time in the field. However it is important to remember that the relevant measure is not average response time by maximum response time (or perhaps some large percentile of the response time distribution), because survey analysis generally does not begin until all the responses are in. Most studies tend to conclude, often with litter or no empirical evidence, that Internet-based surveys are faster than survey sent out by postal mail. This conclusion is usually based on the recognition that e-mail and other forms of electronic communication can be transmitted instantaneously; whereas postal mail takes more time. However, this conclusion naively ignores the reality that the total amount of time required for fielding a survey is more than just the survey response time.

A complete comparison must take into account the mode of contact, how long that process will take and the mode of follow-up, allowing for multiple follow-up contact periods. For example, if the e-mail addresses of respondents are unavailable and probability sample desired, the respondents may have to be contacted by regular mail. In this case, a Web survey saves time only for the delivery of the completed questionnaire but not for the contact and follow-up.

In e-mail survey, in which the presumption is that the potential respondents e-mail addresses are known and can therefore be used not just for

delivering the survey but also for prenotification and nonresponse follow-up, the time savings can be substantial. For example, a week of delivery time must be allowed when using the postal mail. With an advanced letter and single mail follow-up this 1 week delay telescopes into over a month because 2 weeks must be budgeted for initial survey delivery and response time. By comparison, in all electronic process the same operation has the potential to be completed in a few days or less. Yet even in an all-electronic environment it is not necessarily true that Internet-based surveys will require shorter time periods. In a comparison of response speed between e-mail and mail, Tse and Fan (2008) did not find a statistically significant difference in the time between the sending and receipt of an e-mail survey to university faculty and staff members and an equivalent survey sent by mail. Furthermore, to achieve sufficiently high response rates, it may be necessary to leave an Internet-based survey in the field for an extended period of time.

Some studies report a more timely response for e-mail surveys. Tse (1999) found a statistically significant difference in the average initial response time for those who received an e-mail survey compared to those who received a paper survey in the campus mail (1 day versus 2.5 days). Further, in Tse's experiment most e-mail survey recipients either responded almost immediately (within 1 day) or did not respond at all. Schaefer and Dillman (1998) also documented faster e-mail response rates. Seventy six percent of all responses were received in 4 days or less. Pealer et al. (2001) found a statistically significant difference in the average return time between their e-mail study (7.3 days) and their mail study (9.8 days). However the final e-mail survey was received after 24 days and the final mail survey after 25 days, a negligible difference.

Truell et al. (2002) conducted a study with the aim of compare timeliness and response rates of e-mail and mail survey. They used a sample from a directory of practitioners that contained both the residential and e-mail address of the practitioners. They splited the sample and sent to half the e-mail survey and to the other half a mail survey. In terms of response rates they found that 52% of the surveys were returned to the researchers, 51% of the e-mail based surveys, and 53% of the mail surveys. As to response speed, it took 9.22 days on average for the e-mail-based survey and 16.43 days for the mail surveys to be returned. In terms of response completeness, respondents to the internet version of the survey completed 22.51 of the 35 possible items, whereas respondents to the mail version completed 16.88. Sheehan (2001) conducted a meta-analysis of studies on e-mail surveys only.

There is evidence that the speed of response to e-mail surveys is faster from individuals who received pre-notification of the survey than from respondents that did not. According to Crawford et al. (2001) the e-mail invitation plays a disproportionately important role in eliciting responses to an e-mail survey.

In sum, based on the existing evidence it is reasonable to conclude that the delivery time of an e-mail based survey is faster than the delivery time of a survey by mail. It is also reasonable to conclude that completed e-mail surveys start arriving earlier than mail surveys. Yet, when measuring the time that takes to all the completed surveys to arrive to the researcher, differences between mail and e-mail surveys were not dramatic. Thus, it does not necessarily follow that the higher speed of response of e-mail surveys relative to mail, will translate into a significantly shorter survey fielding period. In order to decrease the time that the survey is in the field, the use of incentives and follow-ups to increase responsiveness and response rate is needed.

While there are clear advantages in the use of e-mail surveys, there are some issues that challenge data quality, response rates, and representativeness of the sample that need to be discussed and addressed.

A starting point in problems that might arise with the administration of e-mail surveys are variations in e-mail applications. Researchers need to be aware that computer users have different connections and use different e-mail applications. The format of the e-mail can vary according to the applications and the respondent might receive a questionnaire difficult to read and to complete. To overcome this problem,

before entering the field, surveys need to be pretested using different e-mail programs to make sure of their applicability and the ability of the respondent to understand the survey format (Evans and Matur 2005).

Data collection Data collection is dependent on the form of questionnaire distribution. In online surveys there are typically three ways to distribute the survey to respondents. One is sending to the respondents the entire questionnaire in an e-mail message. A second way is to send the respondents an introductory letter and ask their agreement to participate, and later on send the questionnaire as an attachment. Finally placing a general request in an electronic environment and asking respondents to participate.

Questionnaires sent by e-mail leave the respondents with the relatively simple task of completing and returning the document. In an e-mail version of online surveys, a questionnaire can be sent to respondents as text included within a conventional e-mail message. To complete the survey, respondents use the "reply" function and their e-mail systems, having inserted their responses into the text of the message returned. In order to ensure consistency in the returned e-mail with the questionnaire is strongly adviced to create the questionnaire in a HTML format or to send a formatted questionnaire in a standard word processing package as an attachment. Including the questionnaire in the e-mail message can create problems of understanding responses because of lack of consistency and variety of e-mail programs.

One problem that arises after the questionnaire is returned is the need to code the data. In addition, researchers need to be aware that even among Internet users, there is considerable variation in the skills and experience of users. To alleviate this problem, researchers need to provide simple instructions on responding to the survey and sending it back. To overcome this problem, providing a phone line or instant messenger address that can provide online help to the respondent can reduce difficulties (Evans and Matur 2005).

As any technique of data collection, the use of e-mail surveys is appropriate for some types of research and less appropriate for others. In this section we discuss central issues related to sampling, response rates, and reliability of the data. A central issue when discussing the use of e-mail surveys is the sampling frame. For specific studies, in which the population, the extent of Internet use, and their e-mail is known, sampling is not a major issue. For example, it should not be a problem to create a sampling frame for internal organizational studies. For profit, social welfare and governmental organizations interested in conducting a study of their employees or customers should not face serious problems in creating a sampling frame, drawing a representative sample, and conducting an e-mail survey. This is because the total population, their names, and addresses are known to the initiators of the study.

E-mail surveys are still limited in their ability to reach a random representative sample of the population. While Internet access and e-mail use are becoming more common in western societies, still access is not universal and can not ensure representativeness of the population. However, social research has many motivations and some of them can be met using an e-mail survey. For example, e-mail survey can be useful for studies on the use of the Internet and computer-mediated communication. Studies investigating Internet uses, participation in social network sites, forums, and bulletin boards attempt to reach a population of Internet users only and e-mail surveys can fill this function. An additional area that e-mail surveys may be useful is when the study attempts to reach respondents with deviant or non-normative behaviors. Their participation in electronic social spaces might help in gaining access to respondents normally difficult to reach and the Internet relative anonymity might produce high rates of survey participation. The differential access of social groups of the population to the Internet can be an advantage for the study of specific age groups. For example, youth higher rates of access and use of the Internet makes e-mail surveys useful to conduct studies on early and late adolescents, both high school and college students (Van Selm and Jankowski 2006).

However, sampling is a problem when attempting to conduct studies whose goal is to generalize to the general population. Achieving a random sample of Internet users is problematic. A major problem is the lack of a central registration of users on the Web. Unlike telephone numbers and home addresses, in which an almost complete list of respondents can be achieved, constructing a random sample from e-mail addresses seems impossible. Researchers who use online surveys generally invest a considerable effort into ensuring an acceptable level of randomness and representativeness of non-probability samples. A number of procedures can be implemented to achieve an acceptable level: random selection of e-mail or username addresses from forums, employment of sampling frame from list of users that have access to the Internet (Kaye and Johnson 1999).

Before a reliable estimate can be made about the distribution of a characteristic in a population based on as sample of a specified population, a number of sources of error must be overcome. Coverage errors are the result of situation in which all units of a population do not have an equal probability of inclusion in the sample that is drawn to represent the entire population. Sampling errors are a consequence of only surveying a portion of the sample rather than all members. In online surveys the sampling error cannot be calculated since the underlying assumption behind this calculation requires knowledge about the probabilities of selection. One solution is to include the entire population of given groups in order to avoid this type of error (Sills and Song 2002).

Online panels have become a popular solution to the sample frame problem for those instances in which there is no usable and complete list of e-mail addresses for the target population. For purposes of this study, an online panel is a data base of potential respondents who declare that they will cooperate with future online data collection if selected (International organization of standardization 2009). Probably the most familiar type of online panel is *a general population panel which typically includes* hundreds of thousands of members and is used for both general population studies, as well as for reaching respondents with low incidence events or characteristics such as owners of luxury cars or people suffering from pancreatic cancer. The panel serves as a frame from which samples are drawn to meet the specific needs of particular studies. The design of the study-specific samples may vary depending on the survey topic and population of interest.

Census balanced samples are designed to reflect the basic demographics of the population. The target proportions could be based on distributions as they occur in the larger population for some combination of relevant demographic characteristics.

A *specialty panel* is a group of people who are selected because they own certain products, are a specific demographic group, are in a specific profession, engage in certain behaviors, hold certain attitudes or beliefs, or are customers of a particular company. *A proprietary* panel is a type of specialty panel in which the members of the panel participate in research for a particular company. A car manufacturer gathers e-mail addresses of high end vehicle owners who have volunteered to take surveys about vehicles.

Targeted samples may select panel members who have characteristics of specific interest to a researcher, such as auto owners, specific occupational groups, persons suffering from same diseases, or households with children.

18.5.1 How Members are Recruited to an Online Panel?

Members are recruited through a constant recruitment campaign. The agency creating the online panel, typically appeals to some combination of the following motivations to complete surveys, a contingent incentive such as money or points or sweepstakes with a potential to win prizes. Another form is to appeal to the individual need for self expression. Agencies ask individuals that feel that it is important to express their opinion, to influence decisions to join the online panel. A further motivation is appealing for the sake of social comparison, asking participants to participate and promising

the diffusion of the results to the panel participants so they can learn not only on their behavior but to compare it to the behavior of others that participated in the panel. Panel companies usually recruit to an online panel using a variety of methods, from sending e-mails to company customers, displaying ads in various websites, and through banners.

Joining a panel is typically a two-step process. At a minimum most reputable research companies follow a double opt in process, whereby a person first indicates his or her interest in joining the panel. The panel company then sends an e-mail to the listed address; the person must then take a positive action indicating intent to join the panel and answer the questionnaire sent when he or she are chosen to be a participant in the sample. Upon indicating the willingness to participate in the panel companies will ask the new member to complete a profiling survey that collects a wide variety of background, demographic, attitudinal, experiential, and behavioral data that can be used later to select the panelist for specific studies.

Purposive sampling is the norm in online panels. The sampling specifications used to develop these samples can be very detailed and complex including not just demographic characteristics but also specific behaviors or even previously expressed opinions. With the sample drawn the company sends an e-mail of invitation to the sampled member.

Blasius and Brandt (2010) studied the representativeness of online surveys through stratified samples. In their study, they compared a stratified sample from over 20,000 participants in an online panel with a representative face-to-face data from the German Social Survey 2002. They compared the results of the two studies on the basis of more than 1,100 cases, each of which are equivalent in terms of age, sex, and education. They found that the stratified online sample was not representative of the entire populations. In the first step of the study, they compared the marginal of sex, education, and age groups from the online access panel with the micro-census data. They found that the online panel was biased in a way that almost reflects the different user groups of the Internet: females, elderly, and less educated were under represented. Under these conditions the probability of being selected is quite different. Blasius and Brandt (2010) further compared attitudes to society for individuals sampled from the GSS and the ones from an online panel, to find clear differences regarding items measuring attitudes. Consequently, they conclude that the online panel was biased and not representative, and that even the use of statistical weights did not correct the problem.

From this discussion, it can be inferred that sampling a representative sample of the population for an online survey is still a major weakness. Internet use in western countries is reaching saturation but is not universal. Internet users tend to be younger, more educated, with higher income and employed in white collar professions. In addition, there is no central registry of Internet users and their e-mail addresses. As a result in most cases the online sample reflects social differences in access and difficulties in reaching randomly e-mail addresses. Another central problem of online panels is panel attrition, e.g., panel members leaving the panel after responding to a few surveys. For this reason, companies created and developed incentive programs to combat panel attrition and to increase the likelihood that panel members will complete a survey. Panelists are offered compensation of some form. This can include cash, points redeemed for various goods or instant win games. In any case, panel maintenance is a challenge.

Online panels of respondents are currently being used in order to reduce sample selection bias. While there is considerable progress toward drawing representative sample from online panels, they are still limited as they still reflect the Internet population and statistical weights are still not a viable solution. For this reason, at this time it is more appropriate to use an e-mail survey when the total population is known and can be sampled, such as studies in organizations, studies of patterns of behavior of Internet users only, and studies that do not require a randomly selected sample of the total population.

A well-known problem with mail surveys is the low response rate that requires the researcher to send reminders to increase the survey response rate. As mail surveys, e-mail based surveys require sending reminders to the participants. An important feature of the use of e-mail surveys is the easiness of sending remainders to potential participants. Creating a simple data base of all the respondents e-mails facilitates the monitoring of responses and sending with a simple click e-mails reminding the respondents to fill out the questionnaire and sending it back.

Survey response rates are often affected by respondent's characteristics. E-mail surveys are either embedded or attached in e-mail notifications and respondents can return the surveys through e-mails. There have been concerns on the response rate of e-mail survey, partially because of the uneven access to e-mail technology for different target populations and their respective comfort level with the internet and e-mail methodology. Due to such concerns about uneven access to, and uneven familiarity with, internet in the general population, e-mail surveys technology unlike mail surveys target populations that have internet access (Mehta and Sivadas 1995), such as college students (Bachman et al. 2000). Studies aiming to compare e-mail and mail surveys tend to sample from the same population that can be contacted either by e-mail or mail around the same period of time. As a result, for the purpose of understanding survey mode differences, these comparative studies should provide us with better understanding about response rate differences between these two survey modes.

The importance of survey response rates, and a higher response rate, is viewed not only as desirable but also as an important criterion by which the quality of a survey is measured because a high response rate reflects less serious potential non-response bias. One central question is whether e-mail surveys produce comparable or even higher response rates than traditional mail surveys (Shih and Fan 2009).

A number of meta-analysis comparing response rates to e-mail and mail surveys have been conducted. These studies often focus on two questions: (1) what are the observed response rate differences between e-mail and mail surveys? Such studies usually examine a direct comparison of both e-mail and mail survey modes; and (2) how different factors could explain inconsistent response rates differences between e-mail and mail surveys? Tse and Fan (2008) conducted a meta-analysis based on 35 publications that compared e-mail and mail survey response rates. The subject of comparison included a broad range of disciplines. Comparisons were made using the most simple definition of response rates, the minimum response rate is essentially the number of returned completed surveys divided by the total number of surveys sent out, regardless of different reasons for non returns (refusal, non contact, unknown if household occupied).

Of the 35 comparative results the sample size varied considerably, with the smallest sample size being 43 and the largest 8,253. The average sample size was 1,519. The mean response rate in e-mail surveys was 33%, with a maximum of 85% and a minimum of 5%. The average response rate in mail surveys was 53% with a range from 11 to 85%. The variation in response rate was associated with 2 central features of the survey. About 36% of the variance was explained by populations type and 9% by follow-up remainders. The results indicated that there was almost no difference in response rate for e-mail and mail surveys when the populations were college students. However, when the populations were professional, employees or a general population the response rate for e-mail surveys was 28% less than for mail surveys. Regarding follow-ups, the results indicated that when there was one follow-up remainder the response rate was higher for mail surveys. Only under the condition of two reminders the average response rate was in favor of e-mail surveys by 2%.

A surprising effect of the meta-analysis was that little variation in the response rate differences between e-mail and mail surveys has been accounted by the study feature incentives for participation. The comparative studies that used some kind of incentive and those that did not use any incentive were statistically similar in terms of their response rate difference between e-mail

and mail survey methods. However, these results should not be taken against incentives. The meta-analysis did not compare e-mail surveys with incentive versus mail surveys without incentive or vice versa. Instead, the study examined studies that compared e-mail and mail surveys separated them into two groups the ones that used incentives versus those that did not.

How can response rate be increased? In recent years a mixed mode approach was developed and used (see later chapter by Shine and Dulisse Chap. 21). One approach is to use mixed mode surveys involving a combination of mail and e-mail. For example a letter is sent and individuals are asked to respond to an e-mail survey, or an e-mail is sent asking to answer a mail survey that is sent to their place of residence. Converse et al. (2008) conducted such a study comparing these two approaches. Participants were 1,500 teachers from pre-K to grade 12. The questionnaire contained 53 core items and additional 42 items that were conditional on the type of school they work in. Two versions of the questionnaire were created. First, paper-based version was created and next an online version was created. Respondents were randomly assigned to two groups. The first received an initial contact and questionnaire by mail and then a follow-up via an e-mail that directed them to complete an online questionnaire. The other group received an initial contact via e-mail that directed them to the online questionnaire and a follow-up by mail with the posted paper pencil questionnaire. The overall response rate was 76.3%, with 61.3% of the sample responding to the initial contact and 15% to the follow-up. There was a statistically significant difference in initial response rates with mail resulting in a higher response rate than e-mail.

18.6 Summary

Online surveys were frequently used at the beginning of the process of Internet adoption in society. At this early historical period online surveys were widely used to collect data on Internet behavior and later on the study of sensitive topics and the effects of computer-mediated communication on organizations and individuals. With the technological advances, mainly the introduction of Web survey methodology, e-mail surveys are used less often compared to Web-based surveys which tend to be the more preferred method nowadays. The future of e-mail surveys is unclear, as further technological developments may affect the basic infrastructure: an increasing percentage of the population that regularly writes and reads e-mail. With the advent of social networking sites, e-mail use is decreasing and it is difficult to predict the future of online communication. It is very likely that e-mail will be used more and more as part of a mixed method. Respondents will be contacted by e-mail and they will be able to choose if to answer the survey using an e-mail program or accessing a website with the survey or to answer using other method such as over the Internet phone call (e.g., Skype).

Assuming that e-mail surveys will continue to be used there are a number of actions that researchers need to take in order to improve the data quality that can be obtained from these surveys. The most important issue is to think on conducting an e-mail survey in close relation to the central question of the research and the population of interest. For conducting organizational studies e-mail surveys might continue to be a method of choice as the population is known, the e-mail addresses can be accessed and the response rates tend to be high. For conducting studies that involve the extraction of a representative sample of the population, researchers need to create more imaginative ways to access updated e-mail addresses. An approach that has been taken in recent years is replacing the search for a representative sample by relying on purposive recruited online panels of survey participants. This approach is promising but researchers need to be careful when taking this approach. In particular, they need to conduct research on the company that has created the panel, the demographic characteristics of the panel, and need to show that the sample taken from the panel is representative of the population. Thus, the use of a sample from an online panel does not necessarily indicate that it is representative and that the data is of high quality.

Quality of data from an online survey is dependent on the motivation of respondents to participate and the response rates. In order to increase the motivation for participation, a number of actions should be taken. For example, sending personalized e-mails to potential respondents. An e-mail that provides information on how participants were chosen, a short and engaging description of the research question and encouraging participation appealing to collective interests, such as the need to understand a phenomena, the ability of the participant to influence social processes expressing his/her views can help increasing the willingness of participation. Incentives such as credit points, discount coupons, and participation in sweepstakes have been shown to increase the motivation for participation. Response rates have been found to be lower in e-mail surveys than in other survey types. Beyond the techniques used for motivation to participate, response rates can be improved by following up the respondents and sending remainders.

E-mail surveys are a convenient and low cost methodology and its use is still an advantage for the study of Internet behavior. Technological changes shape the extent that they will be used and the patterns of its use in the future.

References

Bachman, D. P., Elfrind, J., & Vazzana, G. (2000). E-mail and snail mail face off in research. *Marketing Research, 11*, 10–15.

Bhatnagar, A., Misra, S., & Raoi, H. R. (2000). On risk, convenience, and Internet shopping behavior. *Communications of the ACM, 43*, 11.

Blasius, J., & Brandt, M. (2010). Representativeness in online surveys through stratified samples. *Bulletin of Sociological Methodology, 107*, 5–21.

Borkan, B. (2010). The mode effect in mixed mode surveys, mail and web surveys. *Social Science Computer Review, 28*, 371–380.

Converse, P. D., Wolfe, E. W., & Oswald, F. L. (2008). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation, 29*(1), 99–107.

Coderre, F., Mathieu, A., & St-Laurent, N. (2004). Comparison of quality of qualitative data obtained through telephone, postal and e-mail surveys. *International Journal of Market Research, 46*, 347–357.

Couper, M. P., Blair, J., & Tripleet, T. (1999). A comparison of mail and e-mail for a survey of employees in US statistical agencies. *Journal of official Statistics, 15*, 39–56.

Dillman, D. A., Torotra, R. D., Conradt, J., & Bowerk, D. (1998, 9–13 August). *Influence of plain vs. fancy design on response rates of surveys*. Dallas, TX: Paper presented at the annual meeting of the American Statistical Association.

Elford, J., Bolding, G., Davis, M., Sherr, L., & Hart, G. (2004). Web-based behavioral surveillance among men who have sex with men: A comparison of online and offline samples in London, UK. *Journal of Acquired Immune Deficiency Syndromes, 4*, 421–426.

Evans, J. R., & Matur, A. (2005). The value of online surveys. *Internet Research, 15*, 195–219.

Folkman-Curasi, C. (2001). A critical exploration of face-to-face interviewing vs c.m.c. interviewing. *International Journal of Market Research, 43*, 361–375.

Goritz, A. (2004). Recruitment for online survey panels. *International Journal of Market Research, 46*, 411–425.

Hogg, A. (2003). Web efforts energize costumer research. *Electric Perspectives, 28*(5) 81–83.

International Organization for Standardization. (2009). *ISO 26362:2009 Access panels in market, opinion, and social research-vocabulary and service requirements*. Geneva: Switzerland.

Jansen, K. J., Corley, K. G., & Jansen, B. J. (2006). E-survey methodology: A review, issues, and implications. In J. D. Baker & R. Woods (Eds.), *Encyclopedia of electronic surveys and measurements* (pp. 1–8). Hershey, PA.: Idea Group Publishing.

Jones, R., & Pitt, N., (1999). Health survey in the workplace. *Occupational Medicine, 49*, 556–558.

Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*, 440–453.

Kaye, B. K., & Johnson, T. J. (1999). Taming the cyber frontier—techniques for improving online surveys. *Social Science Computer Review, 17*, 323–337.

Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 402–413.

Li, H., Kuo, C., & Russell, M. (1999). The impact of perceived channel utilities, shopping orientations, and demographics on the consumer's online buying behavior. *Journal of Computer Mediated Communication, 5*, 2.

Mehta, R., & Sivadas, E. (1995). Comparing response rates and response content in mail versus electronic mail surveys. *Journal of the Market Research Society, 37*, 429–439.

Parks, M. R. & Floyd, C. (1996). Making friends in cyberspace. *Journal of Computer Mediated Communication, 1*, 4. Downloaded on February 1, 2011 from http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1996.tb00176.x/full.

Pealer, L., Weiler, R. M., Pigg, R. M., Miller, D. & Dorman, S. M. (2001). The feasibility of a web-based

surveillance system to collect health risk behavior data from college students. *Health Education & Behavior*, *28*, 547–559

Ross, M. W., Tikkanen, R., & Mansson, S.-A. (2000). Differences between Internet samples and conventional samples of men who have sex with men: implications for research and HIV interventions. *Social Science and Medicine, 51*, 749–758.

Schaefer, D. R., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly, 3*(62), 378–390.

Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication*, *6*(2).

Sheehan, K. B. & Hoy, M. G. (1997). Using E-mail to survey internet users in the United States: Methodology and assessment. *Journal of Computer Mediated Communication,* Downloaded on Februry 1, 2011 from *jcmc.indiana.edu/vol4/issue3/sheehan.htm*.

Sheehan, K. B., & Hoy, M. G. (1999). Flaming, complaining, abstaining: How online users respond to privacy concerns. *Journal of Advertising, 28*, 37–51.

Shih, T., & Fan, X. (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review, 4*(1), 26–40.

Sills, S. J., & Song, C. (2002). Innovations in survey research: An application of web surveys. *Social Science Computer Review, 20*, 22–30.

Taylor, H. (2000). Does internet research work? *International Journal of Market Research, 42*, 51–63.

Truell, A. D., Barlett, J. E, I. I., & Alexander, M. A. (2002). Response rate, speed, and completeness: A comparison of Internet-based and mail surveys. *Behavior Research Methods, Instruments, & Computers, 34*(1), 46–49.

Tse, A. C. B. (1999). Conducting electronic focus groups discussions among Chinese respondents. *Journal of Market Research Society, 41*, 407–415.

Tse, H. S., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods, 20*, 249–271.

Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality & Quantity*, *40*(3), 435−456.

Suggested Further Readings

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, M., Dillman, D, Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J. & Lavrakas, P. J. (2010). Research synthesis. AAPOR report on online panels. *Public Opinion Quarterly, 74*, 711–781.

Dillman, D. A. (2009). Internet, Mail, and Mixed Methods surveys: The tailored design method. Hoboken, N.J.: Wiley.

Increasing Response Rate in Web-Based/Internet Surveys

19

Amber N. Manzo and Jennifer M. Burke

19.1 Introduction

Web-based surveys have become increasingly popular. An unprecedented number of Americans currently have access to a computer making it possible to quickly and easily reach survey samples through the Internet. Despite this, response rates for Web-based surveys tend to be low. This may be part of a larger trend—response rates for all types of surveys have been falling in the United States over the last decade. Indeed, researchers have generally found that although we live in an increasingly connected age people are becoming harder to reach and are refusing to participate in surveys even when they can be reached.

This presents a real problem. Low response rates always threaten the validity of survey results because non-response might not be random—non-responders in a sample sometimes share similar characteristics. To the extent that this makes people in a sample who do respond different from the non-responders, the sample will be biased, as will the research results. In addition, statistically controlling for the impact of high non-response rates is tricky. Weighting techniques are complex and difficult to explain to practitioners who need information to formulate policies and recommendations. Accordingly, researchers use methods to try and proactively improve responses to traditional surveys instead of addressing deficiencies after the fact. Because many of these methods have been tested experimentally, there is some general agreement about which ones are effective for traditional types of surveys. But researchers are still exploring how to improve response rates for Web-based surveys.

Although researchers have yet to fully determine how to improve responses to Web-based surveys, this chapter outlines some methods that should be considered. Such methods can be roughly divided into two categories. The first category includes methods that were originally developed to improve responses to traditional surveys but have since been applied to Web-based surveys. Such techniques, discussed in Sect. 19.3, include incentives, pre-notifications, survey invitations, and reminders. The second category includes methods that are specific to Web-based surveys. These techniques, discussed in Sect. 19.4, include techniques that address some design and confidentiality issues that are unique to computers.

As a number of scholars have recently noted, however, a discussion of these different methods is incomplete unless attention is paid to the reasons why people elect or decline to participate in surveys. Examining human behavior in

A. N. Manzo (✉) · J. M. Burke
University of Cincinnati, 445 Dyer Hall,
P.O. Box 210389, Cincinnati, OH 45221, USA
e-mail: manzoan@mail.uc.edu

J. M. Burke
e-mail: Jennifer_M_Burke@hotmail.com

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_19, © Springer Science+Business Media New York 2012

this way makes it more possible to determine how any methods discussed here can be adapted for different populations and survey modes; and although much of the early research about how to improve survey responses was atheoretical in that it merely consisted of experimenting with different techniques, more contemporary work frequently reflects one of two theoretical approaches described in Sect. 19.2.

19.2 Theoretical Considerations

When analyzing why people respond to surveys, a number of researchers emphasize the role of reasoned decision-making. This approach, sometimes referred to as social exchange theory, posits that action is dependent upon three elements: rewards, costs, and trust. Rewards are the expected gains from an action while costs are whatever is given up to gain rewards. Trust refers to the expectation that rewards will outweigh costs at least in the long term. According to Dillman (2000), this means that to improve responses, survey designers must determine how to increase the perceived reward for responding, decrease the perceived cost of responding, and develop a potential respondent's trust that the rewards will outweigh the costs. If this theory explains why people choose to act, it could explain the effectiveness of some of the methods discussed below, such as incentives (increasing rewards) or reducing the amount of time it will take respondents to complete a survey (decreasing costs).

On the other hand, other researchers have noted that many people do not expend time or energy considering whether or not to participate in a survey. As a result, these researchers often devote attention to how people use psychological "short cuts" or "rules of thumb" to decide how to act in particular circumstances. By identifying these rules of thumb and how they lead to the decision to act, survey designers can create situations where the rules are likely to lead to decisions that favor participation.

Researchers have identified a number of such psychological shortcuts, including reciprocity and compliance with legitimate authority. Reciprocity refers to the belief that one may be obliged to offer something to a person who has conferred a benefit. This norm can be powerful. For example, Kunz and Wolcott (1976) sent Christmas cards to people they did not know. Most sent Christmas cards back to them; a number even sent the researchers cards in subsequent years. The belief that we must return a favor may be an alternative explanation for the effectiveness of some methods discussed below, including incentives, which are small and thus may not be enough to be perceived as a type of increased benefit described by proponents of the social exchange theory.

Another psychological shortcut, compliance with legitimate authority, refers to the fact that people are more likely to respond when a request comes from a respected source. For this reason, making surveys seem legitimate or indicating that they are sent from a high status source have been effective at improving response rates. Methods based on this are also discussed below.

Notably though, even if the decision to participate can be explained by the social exchange theory or by the use of psychological rules, it may be that both are operative. As a result, survey designers should not necessarily consider these competing explanations for decision making, but should perhaps keep them both in mind when analyzing how to improve response rates.

19.3 Application of Methods Developed to Improve Responses to Traditional Surveys to Web-Based Surveys

As noted, a number of methods have been developed to improve response rates for traditional types of surveys, including incentives, pre-notifications, survey invitations, and reminders. This section discusses their application to Web-based surveys.

1. **Incentives**

Incentives are frequently used to improve traditional survey response rates. Theorists sometimes argue that their effectiveness is based

on social exchange theory—people evaluate the costs and benefits of an action and act in a way that minimizes costs and maximizes benefits. Incentives are designed to maximize the perceived benefit of taking a survey. Although there is limited empirical research about the use of incentives to increase Web-based survey responses, what is available suggests that prepaid material incentives may be most effective. But when these are unfeasible, other types of incentives, including lotteries, might be effective if offered in the right way.

19.3.1 Material Incentives

Research about traditional surveys consistently shows that certain types of material incentives raise response rates. Specifically, pre-paid incentives, delivered to potential respondents with mailed survey instruments, raise response rates (Church 1993; Edwards et al. 2005). On the other hand, postpaid incentives, delivered after survey completion, do not. A possible social exchange-based explanation for this is that potential respondents interpret postpaid incentives as compensation for taking a survey. Because the incentives are typically small, potential respondents may not feel that they adequately reimburse them for their efforts. Indirect support for this explanation can be derived from the fact that increasing the amount of postpaid incentives tends to increase response rates to mailed surveys (James and Bolstein 1992). The monetary amount needed to create an increase in response rates is quite large, however, and thus it is typically impossible to increase response rates in this manner. In contrast, pre-paid incentives may work better because potential respondents interpret them as gifts. Gifts may trigger a sense of duty or obligation to complete a survey based on a norm of reciprocity.

Based on a meta-analysis, Göritz (2006) concluded that incentives also increase response rates for Web-based surveys. Across studies, people receiving incentives were 19% more likely to respond and 27% more likely to complete Web-based surveys than people who did not receive incentives. Moreover, the effects did not depend upon incentive type, including whether incentives were pre-paid or not. But that could be because only a small number of studies in Gritz's meta-analysis analyzed the effect of different incentive types as opposed to simply comparing the impact of an incentive with the absence of one on response and completion. Indeed, there is a dearth of research regarding whether pre- or postpaid incentives are more effective for Web-based surveys. Nonetheless, if incentives are effective for other types of surveys because of their gift-like status, there is no reason to believe that they would operate differently if distributed in a way that makes them seem similarly gift-like.

a. Pre-paid incentives

Distribution of pre-paid material incentives may be more difficult for Web-based than traditional surveys, but it is possible. If cash is mailed to potential respondents, it can be included with a pre-notification that, as discussed below, can also be an effective means of increasing survey response rates. Money can also be sent to PayPal or other types of online accounts. This option may be particularly pertinent if a sample is drawn from a population that maintains accounts with the same institution that provides the sample frame of e-mail addresses. For example, a sample of college students who may have some type of credit account at their University that allows them to make purchases. Alternatively, e-mails inviting potential respondents to participate in a survey can include a coupon or gift certificate that can be printed out and redeemed, although steps will need to be taken to prevent duplication by having the coupon or certificate appear in a pop-up window that can only be opened once or by having the e-mail invite include a code that expires after one use for a discount or free gift after logging into a separate website. Such coupons or codes can be set up for use at borderless retail sites like Amazon.com, which may be helpful when samples are widely dispersed.

Notably though, some research suggests that pre-paid coupons or certificates are not as effective at improving response rates as pre-paid

cash incentives (Birnholtz et al. 2004). This may be because coupons or certificates for small amounts, such as $2.00, may not be as valuable as the same amount of cash. A $2.00 certificate must typically be combined with other certificates or cash while the same amount of cash is readily spendable. In addition, given that low response rates are common for Web-based surveys, researchers may not want to award prepaid incentives to all potential respondents if it is anticipated that responses will be low despite such efforts. In that case, or when pre-paid incentives are not feasible for other reasons, surveyors might consider using a postpaid incentive because it is probably better to offer some incentive than no incentive at all. Again, coupons or gift certificates can be made available upon survey completion through e-mails that include codes or printable vouchers. Cash could once again be sent through mail or to PayPal or other accounts.

b. Post-paid incentives

Many Web-based surveys that use postpaid incentives rely on lotteries instead of guaranteed awards. In these lotteries, often called sweepstakes or prize draws, respondents are offered a chance to win cash or prizes after completing a survey. When won, the awards are often mailed to an individual long after he or she has completed a Web-based survey. Research regarding the effectiveness of lotteries at improving responses to Web-based surveys suggests that they are less effective than other forms of incentives. For instance, Porter and Whitcomb (2003) conducted an incentive experiment with a group of college students who were sent surveys after they sought information from a website. Students were randomly assigned to one of five groups and offered a chance to win Amazon.com gift certificates valued at $50, $100, $150, $200, or no incentive if they completed a survey. Overall, the response rate for students who were offered the chance to win a gift card (15.2%) did not dramatically differ from the response rate for students who were offered no incentive (13.9%). In a similar vein, Göritz (2004) found that response rates were higher when respondents were offered guaranteed "loyalty points" that could be redeemed for discounts and rebates on selected purchases (82.4%) than chances to win money (78.0%) or other "trinkets" (78.6%) after completing a survey. These findings suggest that when using a postpaid incentive, even a small guaranteed award may be more effective at increasing response rates than offering the mere possibility of winning a more valuable prize.

On the other hand, Tuten et al. (2004) compared response rates for people responding with a banner-advertised Web-survey in Croatia. Before the survey was administered, respondents were assigned to one of three groups. Group one was told that they would be immediately informed if they had won a lottery prize after completing a survey—their response rate was 76.6%. Group two was told that they would be informed later if they won the lottery—their response rate was 70.6%. Group three was not offered a lottery incentive at all—their response rate was 62.3%. Given the better response rate for the group that was told that they would learn if they won immediately, it may be that postpaid lottery incentives can improve response rates if administered in particular ways, including by limiting the amount of time that survey participants must wait to learn if they will receive an incentive. More research regarding lottery style incentives is warranted. If they can be used effectively, they might be an optimal choice because they can be easily administered and distributed via the Internet.

c. Caution about the use of material incentives

Although it is generally agreed that material incentives can increase response rates and thereby reduce error, some have voiced concern about how such methods may have the effect of creating a group of mercenaries who can use technology to find and respond to Web-based surveys purely for profit. Indeed, a quick Internet search reveals dozens of websites that solicit survey responders and encourage them to sign up for multiple surveys that will "pay cash" or benefits "just for offering their opinions" (e.g., http://getpaid4youropinion.com/ and http://www.paidsurveysonline.com/). Some of these same websites provide special software that allows respondents to automatically fill out surveys "even faster" by

inputting information into the program that is then used to automatically populate survey response fields so respondents can take more surveys in less time and receive more money (e.g. http://www.surveynet101.net/faqs.html). The possibility of people participating in surveys simply because they represent a chance to "earn" money raises issues about answer accuracy, including whether answers to preliminary screening questions are being answered falsely simply so that a respondent can be eligible to take a survey and receive an incentive. Researchers should be mindful about such issues particularly when administering opt-in surveys. Work may also need to be done to determine how to detect such respondents in a sample so they can be excluded.

19.3.2 Non-Material Incentives

To date, there has been even less research about the effect of non-material incentives on Web-based survey response rates. However, in three online experiments, Göritz (2005) found that response rates improved and drop-out rates decreased when respondents to a Web-based survey were told that upon completion, they would have a chance to review a summary of the results of the survey they had just taken. Again, more research about such incentives is likely warranted given that curiosity may be a powerful but underexplored motivator.

2. **Pre-notifications**

Pre-notifications are also frequently used to improve traditional survey responses. It has been shown that mailing an advance letter to survey respondents prior to administering a telephone survey is a cost-effective way of increasing telephone survey response rates (Hembroff et al. 2005; De Leeuw et al. 2007). One possible reason for this is that advance letters printed on official stationary with personal signatures lend legitimacy to telephone surveys (Groves and Couper 1998). On the other hand, at least one researcher has queried whether mailed pre-notifications increase response rates because they are special or different from the telephone surveys that they precede or simply because they create an additional contact with potential participants and, as discussed below, multiple contacts tend to increase response rates (Dillman 2000). Regardless of precisely how pre-notifications affect responses to telephone surveys, only a handful of studies have examined the effect of pre-notifications on Web-based surveys. The few findings available suggest that although e-mail pre-notifications may not be effective at increasing response rates, pre-notices that are sent in a different (non-Internet) medium may be most helpful when administering Web-based surveys.

To begin with, it appears that pre-notification do increase response rates for Web-based surveys. In an experiment involving college students, Kaplowitz et al. (2004) found that 31.5% of students who received a pre-notification letter and 29.7% of students who received a pre-notification postcard responded to a Web-based survey, while only 20.7% of students who received no pre-notification responded.

Moreover, these response rates appear to improve when pre-notifications for Web-based surveys are sent in a different (non-Internet) medium. For instance, Harmon, Westin and Levin (2005) tested the effect of three types of pre-notifications sent to people who applied to a government grants program and were selected for a Web-based survey sample. The first group received an e-mail pre-notification that included an attachment with a scanned letter from the survey sponsor. The second group received an e-mail pre-notification with a scanned copy of a letter from the data collection agency. The third group received a hard copy of a pre-notification letter from the sponsoring agency through the U.S. mail. The response rate for the group that received the hard copy letter (69.9%) was greater than the response rates for the other two groups (63.3 and 64.4% respectively).

Other researchers have also found that using a different mode for pre-notification may be beneficial. Crawford et al. (2004) randomly assigned a sample of college students to receive a mail or e-mail pre-notification one week before they were e-mailed an invitation to participate in a Web-based survey. The response rates for students that received the mail pre-notification (52.5%) were significantly higher

than the response rates for students that received the e-mail pre-notification (44.9%). In another experiment involving German college students, Bosnjak et al. (2008) found higher response rates for students who received pre-notifications through text messages (84%), as compared to students who received pre-notification through e-mail (71%) or no pre-notification at all (72%).

In sum, pre-notifications may be effective at improving response rates for Web-based surveys, particularly when they are delivered in a different (non-Internet) medium. As with traditional surveys, one reason for this could be because a letter or postcard, or even a text message, may seem official thus making a Web-based survey appear more legitimate, which may be important in an era where people are rightfully wary of Internet scams. An added advantage of mailing pre-notification letters is that they can be used to transmit pre-paid incentives, which are probably the most effective type of incentives for improving response rates. The disadvantage of using a different medium to send pre-notifications is that in addition to knowing a sample member's e-mail address, a researcher will need to obtain additional contact information, such as a mailing address for letters or postcards, or a telephone number for texts. Moreover, although texts and postcards are not exorbitantly expensive, they are more costly than e-mailed pre-notifications while mailed letters are undoubtedly the most costly way to send pre-notifications. Nonetheless, depending upon the availability of additional contact information, whether a pre-paid incentive will be used, the size of the sample, and sample characteristics, pre-notifications letters sent through the mail may be worth the additional expense.

3. **Invitations**

Researchers have developed a number of ways to use telephone and mailed survey invitations to increase response rates for traditional surveys. But invitations to participate in Web-based surveys are often sent via e-mail. Due to technological constraints, researchers may have less control over the design of these e-mail invitations than they might have over more traditional types of invitations. Nonetheless, attention should be paid to the headers that recipients will view in their e-mail in-box as well as the content of e-mail invitations.

19.3.3 E-Mail Invitation Headers: Sender, Recipient, and Subject Line

E-mail headers refer to the elements of an e-mail that are visible in a recipient's e-mail in-box: the sender, the recipient, and the subject line. Researchers will want to use headers to capture recipients' interest, reassure them that an e-mail invitation to participate in a survey is legitimate and not spam or an illicit attempt to gather personal information, often referred to as "phishing." Once again, there is a dearth of research examining how headers can be effectively used to accomplish these tasks, but some points can be made.

The sender portion of a header is typically found in a "From" field that contains the sender's e-mail address. At least one study found that differences in the title attached to a sender's e-mail address had little effect on response rates. For example, Smith and Kinorski (2003) found that instead of sending an e-mail invites to members of the Harris Interactive Panel from Harris_Poll@hpol.gsbc.com, the standard sender address, researchers sent e-mail invitations from the following e-mail addresses:

(1) Research@hpol.gsbc.com; (2) HIPoints@hpol.gsbc.com; (3) Jesse@hpol.gsbc.com; (4)Terri@hpol.gsbc.com (Smith and Kinorski 2003). They found that none of the different addresses achieved significantly more responses than the standard sender address, including the addresses that looked like they came from persons instead of generic e-mail accounts. However, all the addresses that were tested clearly identified the survey sponsor. It remains to be seen whether e-mail invites sent by data collection agents fare better or worse than those sent by sponsors. In addition, all the addresses came from the same domain (@hpol.gsbc.com). Again, no one has yet determined whether sender addresses from different domains, such as .gov, .edu, or .org,

might impact response rates. It is an issue worthy of exploring given that researchers may be able to use different domains.

The recipient portion of the header typically appears in a "To" field that contains the recipient's e-mail address or a group e-mail address or "list serve" that contains the individual recipients' e-mail address. Little research has been done about whether sending invites to individual addresses produces better response rates than invites sent to list serves. Individual addresses may, however, have the same positive effect of personalized greetings (Chesney 2006). Moreover, no research has been done about whether invites sent to individual addresses or list serves produce better responses than invites sent to a list of individual names. But researchers should consider that a list of individual names compromises recipient confidentiality and may alienate those who are protective of their privacy.

The subject portion of the header also appears in a "Subject" field. This field may be the one that is most likely to encourage recipients to at least open and review the contents of an e-mail invite. It is also probably the easiest piece of the header to manipulate. Nonetheless, only a handful of studies have examined how subject fields can be used to improve response rates. These studies suggest that the subject line matters.

For instance, Trouteaud (2004) sent e-mail invites to subscribers of an online newspaper. Recipients who received an invite that contained a plea for help in the subject line ("Please help" by providing "your advice and opinion") had slightly better response rates than recipients who received an invite that offered recipients an opportunity to share their feelings ("Share your advice and opinions now…"). Kent and Brandal (2003) found that invites with a subject line stating that the e-mail was about a survey had significantly better response rates than invites with a subject line stating offering a prize: "Win a Weekend for Two in Nice." Further, Smith and Kiniorski (2003) asked survey recipients about whether they would immediately open, delete, or do something else with e-mails containing different messages in the subject field. They found that respondents were more inclined to open or respond positively to invites that mentioned consumer products in the subject field than e-mails that mentioned surveys in the subject line. Notably, mentioning a survey in a subject line may not only reduce response rates, it could also create response bias if the subject line describes the survey (i.e. "We want to hear from cigarette smokers!"). This is because revealing this type of information may entice respondents who are interested in the subject matter to open e-mail invites while failing to capture others recipients' attention. Again, more research is warranted, but this issue should be considered.

19.3.4 The Contents of E-Mail Invitations

Due to the proliferation of junk e-mail and spam, e-mail invites may be easily ignored or deleted. For that reason, headers play an important role in convincing a recipient to at least open and review an invitation. But invitation content is equally critical to convincing a respondent to go forward with a survey. And although a number of survey researchers have offered advice about the contents of mail invitations or survey cover letters, few studies have been done about the content of e-mail invitations for Web-based surveys. The few findings suggest that attention should be paid to how the salutation, signature, contact information, and links to a Web-based survey are included in an invitation.

a. Salutations in Invitations

Depending upon the amount of information available, personalization may be relatively easy to accomplish in e-mail invitations given the software used by many Web-based survey programs. Personalization may be effective at increasing response rates under certain, but not all conditions.

For instance, Porter and Whitcomb (2003) tested the effect of personalized salutations and other items in e-mail invitations sent to a sample of 12,000 high school students who sought admissions information from a university but

never applied. They found that personalized salutations ("Dear John") did not generate better response rates than impersonal salutations ("Dear Student"). Similarly, Kent and Brandal (2003) found that personalized invitations only improved response rates by 2% when compared to invitations addressed to "Dear Customer."

Other studies have shown that personal salutations can improve response rates. Heerwegh (2005) found that recipients who received personalized salutations ("Dear John Doe") had significantly better login rates to a Web-based survey than recipients who received impersonal salutations ("Dear Student"). The study was conducted with a sample of 2,000 college freshmen who were asked questions about sensitive topics, including attitudes about marriage, divorce, and sexual behavior. Joinson and Reips (2007a) sent e-mail invitations to a sample of 10,000 Open University students asking them to participate in a survey about sensitive information. The students received invitations with one of the following salutations: "Dear Student;" "Dear Open University Student;" "Dear [First Name];" and "Dear [First Name Last Name]." They reported that the group who received the most personal salutation ("Dear [First Name]") had significantly higher response rates than the other three groups.

Subsequently, Joinson and Reips (2007b) found an interaction effect for salutation and sender status. Specifically, they varied invite salutation and the status of the invite sender and found that response rates were highest for recipients who received a personalized invitation from a high status sender. Response rates were lowest for recipients who received impersonal invitations from a low status sender.

These results suggest that personalized salutations may improve response rates, particularly if invitations are sent by persons with higher status. However, one study has found that personalized salutations might lead to socially desirable reporting when surveys seek information about sensitive topics (Heerwegh 2005). More research is warranted about this issue. Nevertheless, it appears that use of personalized salutations in invites may be helpful, particularly if invites can be sent from authoritative sources.

b. Signatures on invitations

Like sender status, sender signature may have a positive effect on response rates. Signatures may, at the very least, indicate the legitimacy of the survey invitation. Signatures on e-mail invitations can be done in a variety of ways. Digital signatures are images of actual signatures that can be attached to e-mail invitations. And although actual signatures are known to increase response rates for mail surveys, their use may be problematic in e-mail because some e-mail recipients will not be able to view or open digital signatures. Technical issues are addressed in other sections of this chapter, but it should be noted that recipients who cannot open e-mails, or portions of them, may fail to respond to a survey invitation. In addition, even if recipients can open images, their e-mail may not do so automatically.

It is also unclear whether variations in typed sign offs have a beneficial effect. Crawford et al. (2004) sent e-mail invitations to a sample of about 15,000 university students. The invite was signed in one of the following four ways: (1) No name used; (2) Carol J. Boyd, (3) Professor Carol J. Boyd, or (4) Substance Abuse Research Center. The response rate for the group that received the invite from Professor Carol J. Boyd (43%) was significantly higher than rates for the other groups; however, the magnitude of that difference was quite small. The difference between group 3 and the lowest group (the group 4) was only 2.9 percentage points.

On the other hand, Gueguen and Jacobs (2002) e-mailed survey invitations to 161 university students in France. The invitations included digital photos of the purported sender. The group that received the invitations with photos had a higher response rate (83.8%) than the group whose invitation did not include a photo (57.5%). Moreover, responses were highest when the invitation included a photo of a woman sender, although the gender of the purported sender had no impact on response rates when the invites did not include a photo. This study suggests that the sign off can be

manipulated to improve responses. However, as is the case with digital signatures, due to technical constraints images may trigger spam detectors or prevent people from opening the mail. Thus, such techniques may only be useful when invites are sent to a group, such as students at a particular university, who are known to possess e-mail accounts that have the same technical capabilities.

c. Links to Web-Based Surveys in Invitations

The most critical element of an invitation may be the information that allows a recipient to access the actual Web-based survey. Typically, this is done by including a URL, or Uniform Resource Locator. When invitations are sent via e-mail, the URL can be embedded so that the recipient can simply click on the URL and be taken to the survey. Clearly, there are advantages to this that may improve response rates—a recipient who opens an invitation with an embedded URL is already logged onto and using a computer with Internet access.

URLs can be designed to convey the identity of the sender or the sponsoring organization. For this reason, they can be used to establish or bolster the credibility or legitimacy of the survey. For example, the invitation in Fig. 19.1 includes an invitation to fill out a preference survey. Although the survey is hosted on Survey Monkey, an open-access website, the URL includes the name of the organization interested in collecting the information.

In addition to including a name that helps establish legitimacy, URLs should probably be short for a number of reasons. First, e-mail that is set up for rich text or HTML (Hypertext Markup Language) can accommodate lengthy URLs, but those that are strictly text-based cannot. Moreover, short URLs will be easier to use if a recipient wants to write down a URL and access it on a different computer for security or other reasons. If a URL is sent in a different medium, such as in a mailed invitation or in a mailed reminder that is sent after initial invitations, a short URL will be easier for recipients to retype correctly into a browser. Lastly, accuracy in the URL cannot be overstated. If an e-mail invitation is sent to an incorrect address, one sample member may be lost. If invitations include the incorrect URL, the entire sample may be unable to access the survey. It will be easier to proofread short URLs for accuracy.

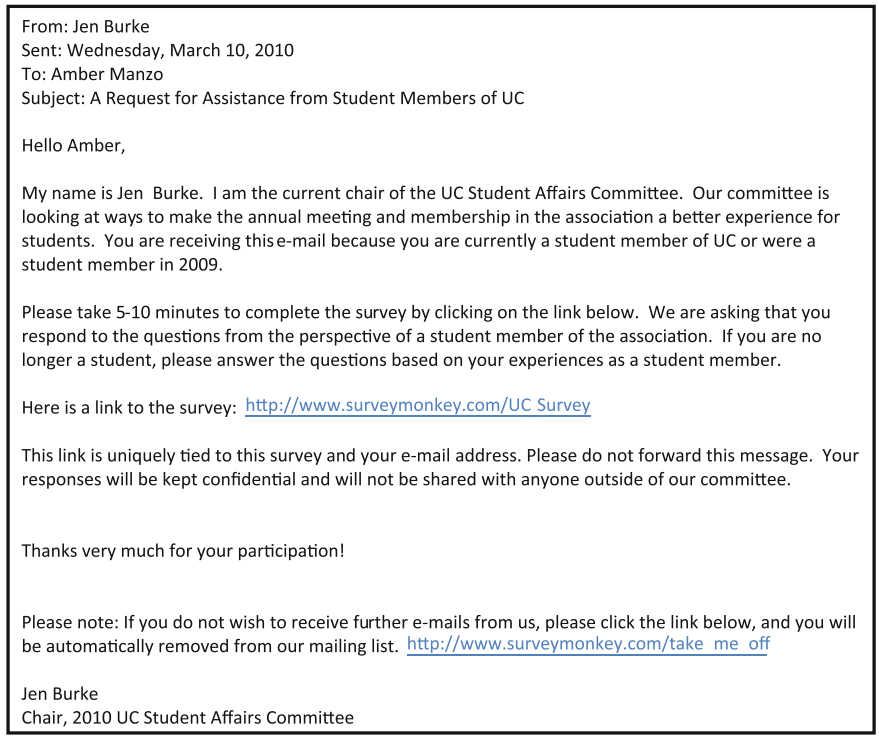
d. Reminders

Multiple contacts with potential respondents are one of the best ways to improve response rates for traditional surveys (Dillman 2000). This can be done by sending a pre-notification, a cover letter, or survey invitation with a copy of the survey or a URL to access the survey, and at least one reminder with another copy of the survey or URL. Multiple contacts may be successful because they remind people to fill out the surveys. However, they may also improve response rates by making surveys seem important or legitimate (Fox et al. 1988).

19.3.5 E-Mail Reminders

Multiple contacts can increase costs. However, increasing contacts with e-mail reminders is attractive because they can be virtually cost free. In addition, it generally seems that e-mail reminders can increase responses to Web-based surveys (Tuten 2005; Trouteaud 2004; Vehovar et al. 1999; Crawford et al. 2001). What remains unclear is how many and when e-mail reminders should be sent.

Research about how many reminders should be sent is sparse. Cook, Heath, and Thompson (2000) contend that response rates will be less than 30% if no reminders are sent at all. Some researchers contend that two follow-up e-mails are ideal (Couper et al. 2001, 2007) while others advocate use of four follow-up e-mails (Schaefer and Dillman 1998). But these suggestions should be tempered by certain considerations. If a targeted audience is likely to be comfortable with e-mail, then more, though not unlimited, reminders may be beneficial. But even with people who are e-mail savvy, there will be a point of diminishing return; researchers may want to consider switching to a different mode to send reminders, as discussed below. On the other hand, fewer reminders may be better for audiences that are likely to be uncomfortable



From: Jen Burke
Sent: Wednesday, March 10, 2010
To: Amber Manzo
Subject: A Request for Assistance from Student Members of UC

Hello Amber,

My name is Jen Burke. I am the current chair of the UC Student Affairs Committee. Our committee is looking at ways to make the annual meeting and membership in the association a better experience for students. You are receiving this e-mail because you are currently a student member of UC or were a student member in 2009.

Please take 5-10 minutes to complete the survey by clicking on the link below. We are asking that you respond to the questions from the perspective of a student member of the association. If you are no longer a student, please answer the questions based on your experiences as a student member.

Here is a link to the survey: http://www.surveymonkey.com/UC Survey

This link is uniquely tied to this survey and your e-mail address. Please do not forward this message. Your responses will be kept confidential and will not be shared with anyone outside of our committee.

Thanks very much for your participation!

Please note: If you do not wish to receive further e-mails from us, please click the link below, and you will be automatically removed from our mailing list. http://www.surveymonkey.com/take_me_off

Jen Burke
Chair, 2010 UC Student Affairs Committee

Fig. 19.1 E-mail invitation

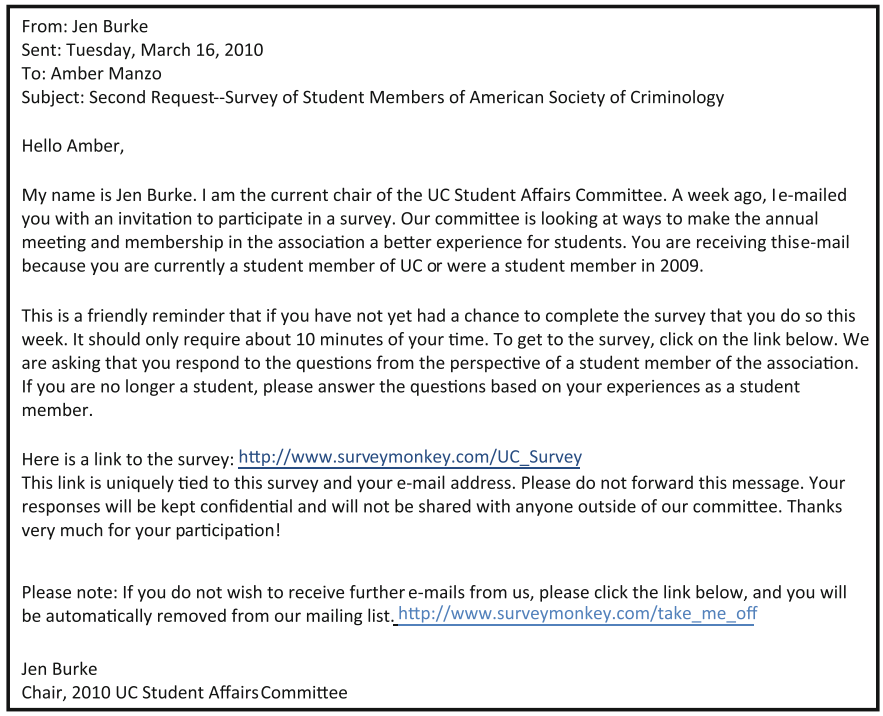
with e-mail because such groups may perceive even a limited number as an annoyance.

Although the timing of mail reminders for mail surveys has been fully explored and it is generally agreed that a first reminder should be sent about one week after a survey is mailed (Dillman 2000), little work has been done about the timing of e-mail reminders for Web-based surveys. What has been done suggests that e-mail reminders should be sent fairly soon after a survey. For instance, Crawford et al. (2001) sent half of the non-responders to a student survey an e-mail reminder two days after a survey was sent to them while the other half received an e-mail reminder five days after the survey was sent. The response rate for the two-day group (26.2%) was significantly higher than the rate for the five-day group (21.7%). In addition, when conducting market research, Deutskens et al. (2004) found that response rates were slightly higher for persons who were sent an e-mail reminder one week after a survey was sent (21.2%) than the response rate for persons who were sent an e-mail reminder two weeks after the survey was sent (19.5%), although this difference was not significant. That one study found a significant increase while the other could not be attributed to the sample type. College students may be more e-mail savvy than the general population and thus more sensitive to the effects of e-mail reminders (Fig. 19.2).

19.3.6 Mixed Mode Reminders

Assuming a pre-notification, survey invitation, or survey has been sent via e-mail, researchers may want to consider sending reminders in some other medium, such as regular mail, because mixing modes may make prior e-mail contacts seem more legitimate and less spam-like. Indeed, several studies have found that mixed mode reminders improve response rates.

Specifically, Kaplowitz et al. (2004) sent postcard reminders to the non-responders in a sample of college students 10 days after the initial e-mail invitation was sent. Some of the



From: Jen Burke
Sent: Tuesday, March 16, 2010
To: Amber Manzo
Subject: Second Request--Survey of Student Members of American Society of Criminology

Hello Amber,

My name is Jen Burke. I am the current chair of the UC Student Affairs Committee. A week ago, I e-mailed you with an invitation to participate in a survey. Our committee is looking at ways to make the annual meeting and membership in the association a better experience for students. You are receiving this e-mail because you are currently a student member of UC or were a student member in 2009.

This is a friendly reminder that if you have not yet had a chance to complete the survey that you do so this week. It should only require about 10 minutes of your time. To get to the survey, click on the link below. We are asking that you respond to the questions from the perspective of a student member of the association. If you are no longer a student, please answer the questions based on your experiences as a student member.

Here is a link to the survey: http://www.surveymonkey.com/UC_Survey
This link is uniquely tied to this survey and your e-mail address. Please do not forward this message. Your responses will be kept confidential and will not be shared with anyone outside of our committee. Thanks very much for your participation!

Please note: If you do not wish to receive further e-mails from us, please click the link below, and you will be automatically removed from our mailing list. http://www.surveymonkey.com/take_me_off

Jen Burke
Chair, 2010 UC Student Affairs Committee

Fig. 19.2 E-mail reminder

non-responders had previously received postcard pre-notifications while some did not. The postcard reminder increased the response rate for the students who never received a pre-notification by about 3.7% (from 20.7% before the reminder to 25.4% after the reminder). However, the response rate for students that received both a postcard reminder and a postcard pre-notification (28.6%) was lower than for the students who only received a postcard reminder (29.7%), though not significantly so.

Couper et al. (2007) also tested the effect of using different methods to send reminders. Half of the non-responders to a 12-month online survey were mailed reminders while the other half of the non-responders received telephone calls. The group that received mail reminders had a 55% response rate while the group receiving telephone calls had a 59% response rate.

Again, these results suggest that like pre-notifications, it may be beneficial to mix modes when sending reminders because the different mode makes the survey seem more official or legitimate. But it is subject to the same limitations. To use a different medium, researchers will need to have sample members' e-mail addresses and mailing addresses for letters or postcards, or a telephone number for calls or texts. In addition, although texts and postcards are not exorbitantly expensive, they are more costly than an e-mailed reminder while mailed letters are undoubtedly the most costly. Nonetheless, depending upon the availability of additional contact information, and if initial contacts have yielded low response rates, mixing modes may be an option to consider.

19.4 Web-Specific Methods of Improving Response Rates

The methods discussed so far were developed to improve responses to traditional types of surveys and then adapted for use with the Web. However, the difference between Web-based and traditional surveys can be so great that new considerations might be called for. This section examines some of the issues unique to Web-based surveys, including design and confidentiality issues, and how they might be addressed.

1. Improving response rates through design

19.4.1 Digital Images and Access Issues

With few exceptions, words have been the raw materials traditionally used by survey designers. Web-based surveys can, however, include a wide variety of visual elements such as pictures, shapes, and symbols (Witte et al. 2004). These are created by HTML, a program language that is continually evolving. Because of this evolution, designers are constantly being presented with innovative image options. As a number of experts have pointed out, however, such advancements require respondents to have more powerful computer systems, and more transmission time. Because transmission time increases the cost of taking a survey, it may negatively impact response rates. The information in this section may only be valid if researchers are trying to reach a population that primarily uses dial-up Internet (an option that is fading away in many countries around the world with the recent advances of high-speed Internet connections).

Dillman et al. (1998) report that in an experiment conducted by the Gallup Organization, advanced features led to a reduced response rate. Gallup sent one group of potential respondents a "fancy" version of a questionnaire that used bright colors, changing backgrounds, tables for questions, and answer categories on the far right. The fancy version took 11 min and 22 s to download on a modem. Gallup sent a second group a "plain" version of the same questionnaire that only used black text on a white background, no images, and placed answer categories on the far left as they are traditionally placed on paper surveys. The plain version required 642 k less computer memory than the fancy version and only took 3 min and 45 s to download. The group that received the plain version completed more pages, filled in more answers, and were less likely to abandon the survey before completion than the group that received the fancy version. Indeed, 93.1% of the group that received the plain version completed the survey while only 82.1% of the group that received the fancy version completed. Notably, individuals in both groups who had less powerful browsing capabilities needed more time to download the questionnaire and had issues with disabled features and computer "crashes."

Dillman et al. (1998) also report that Nicholas and sedivi (1998) found that use of JAVA programming language in a U.S. Census Bureau business survey made it impossible for many respondents to access a survey. Only 68% of the group that received a Web-based version of a questionnaire using JAVA completed the survey while 84% of the group that received a paper version of the questionnaire completed it. At the time, JAVA was an advanced language and many respondents' computers were simply unable to accommodate that level of technology. However, many of the people who agreed to respond indicated that they thought they had sufficient computer capability to do so, indicating that respondents may not always be aware of what type of technology is needed to download and complete a survey.

On the other hand, download times and technical incompatibility may not be as big a problem as they were in the past. Transmission speeds have generally increased as technology has improved. Nonetheless, researchers may want to consider using complex images conservatively in case they do slow things down, frustrate respondents, and decrease response rates. Researchers may even want to consider offering respondents a choice of surveys to complete—perhaps offering versions with and without graphics.

19.4.2 Lay-Outs: Scrolling Versus Paging

Designers also have options when it comes to how a survey appears on a computer screen. One option is to have all the questions in the survey appear in a single frame that the respondent can look at by scrolling down through the frame. Once all answers are completed, a respondent clicks a button at the bottom of the frame that sends or submits all answers. This closely emulates the experience of a paper survey

because respondents can browse through all the questions, answer questions in any order, and go back to change answers. For that reason, Dillman (2000) advocates use of scrolling for Web-based surveys. Use of scrolling also has several other advantages. It is relatively simple in terms of computer programming, is compatible with most browsers, and can be downloaded pretty quickly.

Yet there are disadvantages to scrolling. Typically, scrolling requires respondents to answer all questions in a single session because a respondent cannot answer some questions, save those answers, and then return to finish the rest of the survey. If a respondent fails to hit the button that sends or submits all of his or her answers, then all answers are lost. In regard to skip patterns, respondents who scroll through and see all the questions may also tailor their answers to avoid having to answer additional questions. In addition, skips and routing patterns that are not controlled by the computer program are therefore subject to the same skip and routing errors made by respondents who answer paper surveys.

The second option is to use paging. In paging, one or several related questions are grouped so that they appear on a respondent's computer screen at the same time. After answering those questions, a respondent clicks on a button that sends or submits the answers before they can see a new screen that contains the next question or the next series of related questions. Paging offers several advantages. For instance, skips and routing can be automated. Information from partially completed surveys can be saved and included in results. Respondents can also be told if they have failed to answer certain questions and reminded to do so. But the downside to paging is that it requires more complex programming and technical interaction because answers are constantly being sent and saved. In addition, because respondents cannot scroll through all questions, they may not be able to tell what the survey is about, how long it will take them to complete (though this may be addressed as discussed below), or choose the order in which they want to answer questions.

Little research has been done about the effect of scrolling or paging on response rates. What has been done suggests that there may be no difference. Peytchev et al. (2006) conducted an experiment with a survey sent to 20,000 University of Michigan students who received a survey divided into five topical sections that could be scrolled or a paging version. Response rates for both groups were comparable—about 43% of each group competed the survey—although on average, the paging version took less time (22.93 min) to complete than the scrolling version (23.53 min). Some of this time difference is likely because the paging version performed automatic skips that eliminated nonpertinent questions from respondents' view while the scrolling version allowed respondents to use hyperlinks to do skips but still allowed respondents to view all questions. Notably, although the students who received the scrolling version were provided with hyperlinks to facilitate skips, many did not use them and scrolled down instead. The few students who did use the hyperlinks on the scrolling version finished the survey more quickly than the students who scrolled through the survey.

For this reason, some experts have recommended that researchers should use scrolling or paging under different circumstances. Couper (2008) asserts that scrolling may be best for short surveys that have few or no skips, where missing data is not a major concern, where a respondent's ability to scroll through and review all questions is desirable as a priming mechanism, and where the order in which respondents provide answers is not an issue. Similarly, he recommends that paging be used for longer surveys that include skips and where the sequence that questions are answered is important. Further, when paging is used, Schonlau et al. (2001) recommends that each page contain as few questions per screen as possible to eliminate scrolling in a paging survey. In addition, paging surveys can and should be set up so respondents can start a survey, but then stop and restart it again later, particularly when surveys are lengthy.

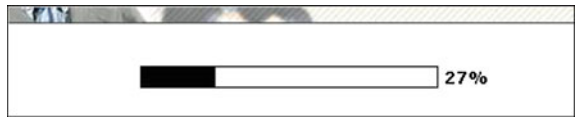


Fig. 19.3 Progress indicator

19.4.3 Announcing Completion Time and Progress Indicators

Survey length and completion time has always been a concern for those administering traditional surveys. Shorter surveys are less costly to a respondent, and therefore, lead many to believe that shorter is generally better. Of course, the cost of completing a longer survey might be offset by an increase in perceived benefit or interest in a survey subject. Ideal length is, therefore, highly context dependent.

Despite this, a number of experts make recommendations about Web-based survey length with estimates centering around 10 min (e.g., Farmer 2000; Rivers 2000). But few experiments have been done to test survey length on responses to Web-based surveys. Those that have been done suggest that providing an estimate of completion time or indicating how much progress has been made may have some interesting impacts.

In a paper survey, respondents can flip through and determine how lengthy a survey is before starting it. But when a Web-based survey uses paging instead of scrolling, a respondent cannot estimate how long a survey is before starting. For that reason, Crawford et al. (2001) administered a Web-based survey that took approximately 20 min to complete for two different groups. One group was told that the survey would take 8–10 min to complete while the other group was told that it would require about 20 min. The group that was falsely told that the survey would only take 8 min was both more likely to start the survey and more likely to abandon the survey before completion. As a result, both groups ended up with very similar response rates.

In addition, during a paper survey, respondents can flip through and estimate how much longer it will take to complete. But again, when a Web-based survey uses paging instead of scrolling, a respondent cannot determine how much progress has been made. The solution to this issue has been to provide progress indicators. Progress indicators (see Fig. 19.3) can take many forms, including pie charts or bar graphs showing the percentage of questions that have been completed. They can also be included on every screen or kept hidden until a respondent chooses to click on a button that makes them visible. Couper (2008) reports that progress indicators do not generally improve response rates and can, in some instances, increase abandonment. Summarizing the results of several experiments, he notes that having a progress indicator visible at all times can generally increase break-off rates for respondents, presumably because they may make a survey seem like it will take longer to complete than a respondent may assume in the absence of any information. He cautions against the use of an ever present progress indicator.

2. Improving responses by addressing confidentiality concerns

Confidentiality and anonymity are growing concerns for survey participants (Cho and LaRose 1999; Rotocki and Lahri 2003). And as Internet use has skyrocketed in the previous decade, people have grown more wary of Internet issues such as identity theft and the amount of personal information that can be accessed via the Internet. Respondents providing answers to sensitive questions will especially want to make sure that their information and identification are kept confidential and not released to websites that monitor outside information (like the FBI) or are reviewed by corporate IT departments (such as when an employer monitors employee browser use) (Fenlason et al. 2002). There are, however,

security steps that surveyors can take to address apprehensions and reassure respondents that their responses will be kept in confidence.

Research shows that concerns regarding confidentiality can have a negative effect on response and completion rates for Web-based surveys. For example, Pallack et al. (2010) conducted a Web-based survey regarding the intimate partner violence (IPV) experiences of working women and women's use of employee assistance programs (EAP). Due to the sensitive nature of the survey, they reported that women had concerns about the confidentiality of the information they were giving. Specifically, they did not want their employers to find out that they had used the EAP for an IVP situation.

Similarly, Mertler (2003) sent teachers who did not complete a Web-based survey an e-mail survey asking why they failed to complete the primary survey. They received a 21% response rate. Out of that, 24% expressed difficulties accessing the survey and 4% said they felt concerned for their security and confidentiality of their answers. This may, however, underestimate the actual concerns of the individuals who did not respond to the second survey because confidentially concerns might have prevented the non-respondents from filling out the second survey as well.

Fenlason et al. (2002) cautioned that Web-based surveys containing a comment section may generate very few comments if the respondents feel that their answers will lead back to them and be misused later on. Thompson and Surface (2007) examined this issue by surveying computer savvy employees working at a military headquarters. They found that the employees were not as comfortable taking online surveys as they were with other mediums of surveying. Additionally, employees who did complete the survey later stated that they felt uncomfortable expressing their true opinion; therefore showing that security concerns with Web-based surveys may affect accuracy as well as responses.

For researchers, Web surveys are not any riskier than traditional surveys in terms of keeping the data secure. The surveyor should take precautions to keep data protected against hacking, accidently leaking, or careless disclosure for both remote hosted solutions and home hosted solutions (Cho and LaRose 1999; Kraut et al. 2004). To do so, surveyors can keep any identifying information separate from the data, use password protections or encryption, use multiple servers to hold different parts of the survey, and install firewalls and virus protection on computers (Cho and LaRose 1999; Kraut et al. 2004; Fan and Yan 2010). When transferring information, they can also send the answers as a string of numbers or values that cannot be easily tied back to individuals or use an outside vendor so the survey is accessed through the Internet and not through an internal system (such as intranet) (Fenlason et al. 2002). Another effective way to keep data safe is to only allow use of it on a computer that does not have Internet access (Fan and Yan 2010).

19.5 Conclusion

In sum, improving response rates to Web-based surveys is a topic that is ripe for further research. Nonetheless, some steps have been taken to identify issues related to responses to such surveys and steps taken to figure out how those issues can best be addressed. A number of these methods have been adapted from research regarding how to improve responses to traditional surveys. Such techniques include use of pre-paid material incentives and mixed modes for multiple contacts such as pre-notifications and reminders. As is the case with traditional surveys, research also suggests that attention should be paid to how e-contacts are constructed with care given to e-mail headers that contain sender, recipient, and subject line information as well as the contents of the e-contacts. In addition, some issues require techniques that are Web specific including addressing how respondents must proceed through survey pages that appear on a screen as well as whether and how to inform respondents about completion time estimates given that they cannot flip through pages and estimate their own progress. Moreover, although confidentiality has always been a

concern, the use of Web-based surveys raises different issues given how information is obtained, retrieved, and stored. As a result, special care might be required to both reassure Web-based survey respondents about the privacy of their responses and to honor that obligation.

References

Birnholtz, J. P., Horn, D. B., Finholt, T. A., & Bae, S. J. (2004). The effects of cash electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents. *Social Science Computer Review, 22*(3), 355–362.

Bosnjak, M., Neubarth, W., Couper, M. P., Bandilla, W., & Kaczmirek, L. (2008). Prenotification in web surveys: The influence of mobile text messaging versus e-mail on response rates and sample composition. *Social Science Computer Review, 26*(2), 213–223.

Chesney, T. (2006). The effect of communication medium on research participation decisions. *Journal of Computer-Mediated Communication*, *11*(3), Article 10. Retrieved from http://jcmc.indiana.edu/vol.11/issue3/chesney.html

Cho, H., & LaRose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review, 17*, 421–434.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly, 57*(1), 62–79.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web—or internet based surveys. *Educational and Psychological Measurement, 60*(6), 821–836.

Couper, M. P. (2008). *Designing effective web surveys*. New York, NY: Cambridge University Press.

Couper, M. P., Traugott, M. W., & Lamas, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly, 65*(2), 230–253.

Couper, M. P., Peytchev, A., Stretcher, V. J., Robert, K., & Anderson, J. (2007). Following up nonrespondents to an online weight management intervention: Randomized trial comparing mail versus telephone. *Journal of Medical Internet Research, 9*(2), e16.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review, 19*(2), 146–162.

Crawford, S. D., McCabe, S. E., Saltz, B., Boyd, C. J., Freisthler, B., & Paschall, M. J. (2004). *Gaining respondent cooperation in college web-based alcohol surveys: Findings from experiments at two universities*. Presented at the American Association for Public Opinion Reserach (AAPOR) 59th Annual Conference.

De Leeuw, E. D., Callegaro, M., Hox, J. J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: A meta-analysis. *Public Opinion Quarterly, 71*(3), 413–444.

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet based surveys: An experimental study. *Marketing Letters, 15*(1), 21–36.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. New York, NY: Wiley.

Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). Principles for constructing web surveys. SESRC Technical Report, 98-50. Retrieved from http://www.websm.org/uploadi/editor/Dillman_1998_Principles_for%20Constructing.pdf

Edwards, P., Cooper, R., Roberts, I., & Frost, C. (2005). Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires. *Journal of Epidemiology and Community Health, 59*(11), 987–999.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*, 132–139.

Farmer, T. (2000). Using the internet for primary research data collection. InfoTek Research Group, Inc. Retrieved from http://researchinfo.com

Fenlason, K. J., Waller, V., DeMay, C. C. (2002). Key issues and approaches to preserving anonymity and confidentiality in web surveys. *Data Recognition Corporation.* Retrieved from http://drctesting.com/survey/articles/anon.pdf

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rates: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*(4), 467–491.

Göritz, A. S. (2004). The impact of material incentives on response quantity, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research, 46*(3), 327–345.

Göritz, A. S. (2005). Incentives in Web-based studies: What to consider and how to decide. WebSM Guide No.2. Retrieved from www.websm.org.

Göritz, A. S. (2006). Incentives in web based studies: Methodological issues and a review. *International Journal of Internet Science, 1*(1), 58–70.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York, NY: Wiley.

Gueguen, N., & Jacobs, C. (2002). Social presence reinforcement and computer-mediated communication: The effect of solicitor's photograph on compliance to a survey request made by e-mail. *Cyber Psychology and Behavior, 5*(2), 139–142.

Harmon, M. A., Westin, E. C., & Levin, K. Y. (2005). *Does type of pre-notification affect web survey response rates?* Presented at the American Association for Public Opinion Research (AAPOR) 60th Annual Conference.

Heerwegh, D. (2005). Effects of personal salutations in e-mail invitations to participate in a web survey. *Public Opinion Quarterly, 69*(1), 588–598.

Hembroff, L. A., Rusz, D., Rafferty, A., McGee, H., & Ehrlich, N. (2005). The cost-effectiveness of

alternative advance mailings in a telephone survey. *Public Opinion Quarterly, 69*(2), 232–245.

James, J. M., & Bolstein, R. (1992). Large monetary incentives and their effect on mail survey response rates. *Public Opinion Quarterly, 56*, 442–453.

Joinson, A. N., & Reips, U. D. (2007a). Personalization, authentication and self-disclosure in self-administered surveys. *Computers in Human Behavior, 23*(1), 275–285.

Joinson, A. N., & Reips, U. D. (2007b). Personalized salutation, power of sender and response rates to Web-based surveys. *Computers in Human Behavior, 23*(3), 1372–1383.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), 94–101.

Kent, R., & Brandal, H. (2003). Improving E-mail response in a permission marketing context. *International Journal of Market Research, 45*(1), 489–506.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. P. (2004). Psychological research online: Opportunities and challenges. *American Psychologist, 59*(2), 105–117.

Kunz, P. R., & Wolcott, M. (1976). Season's greetings: From my status to yours. *Social Science Research, 5*, 269–278.

Mertler, C. A. (2003). Patterns of response and nonresponse from teachers to traditional and web surveys. *Practical Assessment, Research & Evaluation, 8*(22), 1–17.

Nicholas, E., & Sedivi, B. (1998). Economic data collection via the web: A census bureau case study. In *Proceedings of survey methods section*. American statistical Association Annual Meetings, Dallas, Texas

Pallack, K. M., McKay, T., Cumminskey, C., Clinton-Sherrod, A. M., Lindquist, C. H., Lasater, B. M., et al. (2010). Employee assistance program services for intimate partner violence and client satisfaction with these services. *Journal of Occupational and Environmental Medicine, 52*(8), 819–826.

Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. (2006). Web survey design: Paging versus scrolling. *Public Opinion Quarterly, 70*(4), 596–607.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*(4), 389–407.

Rivers, D. (2000). Fulfilling the promises of the web Quirk's Marketing Research Review, 0562. Retrieved from www.quirks.com

Rotocki, N., & Lahri, N. A. (2003). Is the applicability of Web-based surveys for academic research limited to the fields of information technology? In *Proceedings of the 36th Hawaii International Conference on System Sciences*. Los Alamitos, CA: Computer Society Press.

Schaefer, D. R., & Dillman, D. A. (1998). Development of standard e-mail methodology. *Public Opinion Quarterly, 62*(3), 378–397.

Schonlau, M., Fricker, R. D., & Elliot, M. N. (2001). *Conducting research surveys* via *e-mail and the web.* Retrieved from http://www.rand.org/publications/

Smith, R. M., & Kinorski, K. (2003). *Participation in online surveys: Results from a series of experiments*. Presented at the American Association for Public Opinion Reserach (AAPOR) 58th Annual Conference.

Thompson, L. F., & Surface, E. A. (2007). Employee surveys administered online: Attitudes toward the medium, nonresponse, and data representativeness. *Organizational Research Methods, 10*(2), 241–261.

Trouteaud, A. R. (2004). How you ask counts: A test of internet-related components of response rates to a Web-based survey. *Social Science Computer Review, 22*(3), 385–392.

Tuten, T. L. (2005). *Do reminders encourage response but affect response behaviors? Reminders in Web-based surveys*. Paper presented at ESF Workshop on Internet Survey Methodology.

Tuten, T. L., Galesic, M., & Bosnjak, M. (2004). Effects of immediate versus delayed notification of prize draw results on response behavior in internet surveys. *Social Science Computer Review, 22*(3), 377–384.

Vehovar, V., Lozar Manfreda, K., & Batageli, Z. (1999). *Design issues in WWW surveys*. Paper presented at Annual Meeting of the American Association for Public Opinion Research.

Witte, J. C., Pargas, R. P., Mobley, C., & Hawdon, J. (2004). Instrument effects of images in web surveys: A research note. *Social Science Computer Review, 22*(3), 363–369.

Building Your Own Online Panel Via E-Mail and Other Digital Media

20

Vera Toepoel

An important development in the social sciences in recent decades has been the increased use of online surveys and online panel surveys in particular. A panel survey makes similar measurements on the same individuals at different points in time. While survey collection was left to professionals only a few years ago, with the rise of internet, online surveys have become available to the masses. Online surveying is cheap, little effort is needed (especially when quality is not one of your concerns), multimedia can be easily added, and the potential range of respondents is enormous. Many people in education, business, and government see the opportunities of doing Internet surveys on their own. Unfortunately, not everyone has access to the Internet (for penetration rates per country see www.Internetworldstats.com), making it difficult to generalize conclusions based on online surveys to the general population.

Building your own online survey panel can be fairly easy. Just write a questionnaire and publish it on the web via online survey software, often available for free (e.g. thesistools, survey monkey). Make sure that a lot of people visit the questionnaire's website: create ads, publish banners on websites, tweet, text, or if you have a large list of e-mail addresses, send an invitation to that address list. Just make sure you ask for the email addresses from the respondents if you do not already have them. Then you own your own online panel. Just send a new e-mail to all your previous respondents for every new questionnaire. That is it in a nutshell. Of course you have to make all kinds of design choices. This chapter will discuss important design steps for recruiting online panel members.

Advantages and disadvantages of panels

Panels, where the same individuals are interviewed repeatedly at various points in time, have important advantages over independent cross-sections, such as efficiency gains in recruiting, reduced sampling variation in the measurement of change, and the possibility to analyze behavior at the individual respondent level (see, e.g. Baltagi 2001; Toepoel et al. 2008). They help, for example, to analyze changes at the microlevel, to disentangle permanent from transitory changes, and to distinguish between causal effects and individual heterogeneity (Lee 2002). In addition, panels may become useful in marketing settings when it comes to product testing and monitoring (Sikkel & Hoogendoorn 2008).

Two potential drawbacks compared to, e.g., independent cross-sections are attrition bias and panel conditioning (for a discussion see, e.g., Das et al. 2011a, b; Sharot 1991 or Trivellato 1999).

Before going into recruiting online panel members, a note has to be made on the difference

V. Toepoel (✉)
Leisure Studies, Tilburg University,
Warandelaan 2, P.O.Box 90153 5000LE
Tilburg, The Netherlands
e-mail: V.Toepoel@uvt.nl

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_20, © Springer Science+Business Media New York 2012

between client-side and server-side computer surveys. Client-side surveys are executed on a respondent's PC and include e-mail surveys and downloadable executables. The instrument is transmitted to respondents via the Internet. Respondents then answer the survey by either using the reply function in the e-mail client, by entering responses using a word processor, or by using software installed on their computers (Couper 2008). These client-side surveys are commonly completed offline. Server-side surveys are executed online on the survey organization's Web server. Answers to server-side surveys are transmitted to the server every time the "next" button is pressed. Web surveys are the most prevalent type of server-side surveys. Client-side surveys have fallen out of favor, largely because of technical limitations, security concerns, and the rise in accessibility of the Internet (Couper 2008).

This chapter will focus on server-side surveys and discusses ways of recruiting online panel members. The emphasis lies on the recruitment of panel members; methodological issues associated with working with online panel survey data are not taken into account. For more information about this issue see other chapters in this book or, e.g., *International Handbook of Survey Methodology* (De Leeuw et al. 2008) and *Social and Behavioral Research and the Internet* (Das et al. 2011a, b). Section 20.1 discusses types of online surveys. Section 20.2 handles representativeness of online panel surveys. Sections 20.3–20.6 discusses different methods of (online) panel recruitment, while in Sect. 20.7 different strategies for online panel recruitment are summarized. This chapter ends with some best practices and a view on the future of online panel surveys.

20.1 Types of Online Surveys

More than any other mode of data collection, online surveys have led to a large number of different data collection uses, varying widely on several dimensions of survey quality. Couper (2000) predicted more than a decade ago that the rapid spread of electronic data collection methods such as the Internet would produce a division between high quality surveys based on probability samples and surveys focusing more on low cost and rapid turnaround. However, online surveys have become more of a fragmentation than a bifurcation, where there is a wide array of approaches representing varying levels of quality and costs. Couper (2000) identifies eight different types of online surveys, as can be seen in Table 20.1.

Couper (2000) makes a distinction between probability and non-probability online surveys. In non-probability surveys, members of the target population do not have known non-zero probabilities of selection. As a result, generalizations to the population are difficult to make. *Polls* are a type of non-probability survey. Polls are mostly done for entertainment purposes and typically do not lead to generalizations beyond reflecting the views of those who responded. The second type of non-probability surveys are *self-selected Web surveys*. This approach often uses open invitations on banners, websites, etc. These surveys sometimes have no access restrictions and no control over multiple completions. Polls and unrestricted self-selected surveys are often used for independent cross-sectional surveys and not for longitudinal studies, or panels. Companies who have a lot of media attention often use these types of online surveys. The last form of a non-probability survey is a *volunteer opt-in panel*. This approach creates a volunteer panel by wide appeals on well-traveled websites. The panel character results in multiple surveys for the same individuals. The first questionnaire normally consists of some basic demographics and requests the participant's e-mail address. In subsequent appeals this information is used to re-contact the participants and target the questionnaires. Access to these later surveys is often by invitation only and controlled through usernames and passwords in a panel management environment. This has been and probably will be the type of Web survey with the largest market share. Examples of this type of online panel surveys are easy to find. For example, go to www.money4surveys.com for a list of panels. A major concern with this type of panel is the fact that they often claim to be representative of the general population. Often, volunteer opt-in panels

Table 20.1 Types of Online surveys

Non-probability methods	Probability-based methods
1. Polls as entertainment	4. Intercept surveys
2. Unrestricted self-selected surveys	5. List-based samples
3. Volunteer opt-in panels	6. Web option in mixed-mode surveys
	7. Prerecruited panels of internet users
	8. Prerecruited panels of full population

Source Couper (2000)

use weighting techniques to compensate for biases in their (online) sample. For example, propensity score adjustments are made via parallel telephone surveys, estimating the propensity of being in the online sample, based on a vector of covariates measured in both the telephone and online mode (Couper 2000). The success of this approach depends on the quality of the benchmark measures (in this example, the telephone mode) and the choice of variables used for the adjustment. The choice of these variables is dependent upon the main variables of interest in a particular survey. Volunteer opt-in panel agencies often do not take the time to construct varying weights for each survey. In addition, how the weights are exactly constructed is often unclear. Therefore, generalizations from an online volunteer opt-in panel to the general population are difficult to make.

While probability-based surveys are not free of measurement biases on their own, they permit us to measure the sources of non-response. Given the fact that not everyone has access to the Internet, let alone have an e-mail address, and no frame of Internet users exists, there are essentially two approaches to achieving probability-based sampling. You can restrict the sample to those who have access to the Internet, restricting the population to those with Internet access. Or you use alternative offline methods such as telephone or address-based sampling. Those offline methods are then used to recruit members for an online survey panel. *Intercept surveys* and *list-based samples of high-coverage populations* are examples of probability samples where the population is restricted to those who have Internet access. Intercept-based approaches target visitors to a website, generally using systematic sampling techniques to invite every *n*th visitor to the site to participate in a survey. The sampling frame is narrowed down to visitors of the website. Customer satisfaction surveys are often using this approach. Student surveys are an example of list-based samples of high-coverage populations. Examples of offline recruitment methods are *mixed-mode designs* and *prerecruited samples of the full population*. The first approach offers the Web as just one alternative among several in a mixed-mode design. This approach is often used when non-response should be kept to a minimum and measurement error effects of varying mode are not large (Couper 2000). The Web is offered to reduce survey costs or to minimize response burden, other modes are used for those who cannot or will not fill out the survey online. Prerecruited panels of the full population use probability sampling methods such as Random Digit Dialing (RDD) telephone interviews or address-based sampling techniques to identify those with Internet access and recruit panel members. This setup is similar to *prerecruited panels of Internet users*. The difference between prerecruited panels of Internet users and the full population lies in the provision of equipment to those who do not have Internet access in the latter case. Examples of companies that provide equipment to make it possible to fill out surveys online are Knowledge Networks in the U.S. and CentERdata in the Netherlands. Prerecruited samples of the full population are the only method having the potential for obtaining a probability sample of the full population, making it possible to generalize conclusions to the full population (Couper 2000). Volunteer opt-in panels (non-probability-based) and prerecruited panels (of the Internet population

Table 20.2 Top 10 countries with the highest Internet penetration rate

No	Country or region	Penetration (% population)	Internet users latest data	Population (2009 est.)	Source and date of latest data
1	Falkland Islands	100.0	2,483	2,483	ITU—Sept/09
2	Iceland	93.2	285,700	306,694	ITU—Sept/09
3	Norway	90.9	4,235,800	4,660,539	ITU—Sept/09
4	Greenland	90.3	52,000	57,600	ITU—Mar/08
5	Sweden	89.2	8,085,500	9,059,651	ITU—Sept/09
6	Saint Kitts and Nevis	87.2	35,000	40,131	ITU—Dec/09
7	Netherlands	85.6	14,304,600	16,715,999	ITU—Sept/09
8	Denmark	84.2	4,629,600	5,500,510	ITU—June/09
9	Finland	83.5	4,382,700	5,250,275	ITU—Sept/09
10	New Zealand	83.1	3,500,000	4,213,418	IWS—Dec/09

Source www.Internetworldstats.com

or the full population) are types of online surveys suitable for building an online panel due to their longitudinal character.

20.2 Representativeness of Online Panels

As mentioned earlier, representativeness is a major concern for online panels. The theory of survey sampling is heavily based on the probability sampling paradigm, which holds that by selecting random samples, probability theory can be applied, making it possible to quantify the accuracy of estimates (Scherpenzeel & Bethlehem 2011). Since online panels often are not based on probability sampling, but rather on self-selection, representativeness of the sample cannot be estimated. Non-probability panels often claim to produce representative and therefore generalizable results. The term representative is often used vaguely. Some people seem to think or claim that a large sample size and/or high response rate is an indicator of survey quality. However, a large sample size of self-selected respondents or a high response rate among cooperative panel members is unlikely to guarantee reliable outcomes.

One of the major drawbacks of Web surveys is the fact that not everyone has access to the Internet and as a result, under coverage is prominent. Under coverage occurs when elements in the target population do not appear in the sampling frame. Results of online panels may therefore only apply to the Internet population and not to the target population as a whole (with the exception of prerecruited Internet panels of the full population where Internet access is provided to panel members).

Table 20.2 shows the top 10 countries with the highest Internet penetration rates.

As you can see, countries in the north-west of Europe have high penetration rates. Only the Falkland Islands have an Internet penetration of 100%. The Falkland Islands are therefore the only region in the world where under coverage due to Internet penetration is not a problem. Table 20.3 shows Internet penetration rates per continent.

Table 20.3 shows that North America has the highest penetration rate (78%). While we saw in Table 20.2 that the top 10 countries with the highest Internet penetration rate are mostly based in the north-west of Europe, Europe as a whole has a lower Internet penetration rate than North America (58%) because of large differences between European countries. While Africa has the lowest penetration rate, the growth of Internet users in the last 10 years is by far the largest. In the whole world, 30% of people have access to the Web.

Undercoverage bias is determined by two factors: the relative size of the population without Internet to the full population and the contrast

Table 20.3 World Internet usage and population statistics (March, 2011)

World regions	Population (2011 est.)	Internet users Dec. 31, 2000	Internet users latest data	Penetration (% population)	Growth 2000–2011 (%)	Users % of table
Africa	1,037,524,058	4,514,400	118,609,620	11.4	2,527.4	5.7
Asia	3,879,740,877	114,304,000	922,329,554	23.8	706.9	44.0
Europe	816,426,346	105,096,093	476,213,935	58.3	353.1	22.7
Middle East	216,258,843	3,284,800	68,553,666	31.7	1,987.0	3.3
North America	347,394,870	108,096,800	272,066,000	78.3	151.7	13.0
Latin America/ Carib.	597,283,165	18,068,919	215,939,400	36.2	1,037.4	10.3
Oceania/ Australia	35,426,995	7,620,480	21,293,830	60.1	179.4	1.0
WORLD TOTAL	6,930,055,154	360,985,492	2,095,006,005	30.2	480.4	100.0

Source www.Internetworldstats.com

between the Internet population and the non-Internet population (Scherpenzeel & Bethlehem 2011). An increased Internet penetration rate will reduce coverage bias but the contrast does not necessarily decrease as Internet coverage grows. In contrast, the difference between the haves and have nots, also known as the "digital divide", is increasing. Specific groups are substantially underrepresented, such as the elderly, low educated people, and ethnic minorities. The disappearance of undercoverage is therefore not inevitable in the future. Adjustment weighting techniques are often applied in attempt to reduce undercoverage (and/or self-selection) bias. For a discussion of weighing techniques see Scherpenzeel & Bethlehem (2011).

20.3 Traditional Methods of Panel Recruitment

Traditionally, panel recruitment methods have been face-to-face (doorstep method), telephone, or paper-and-pencil (mail). These offline methods can be used for panel recruitment of online panels, making it possible to draw probability samples of the general population. Face-to-face recruitment is very costly, and only big organizations with large survey budgets can use this type of recruitment. In addition to face-to-face recruitment, telephone surveys make use of persuasive abilities of interviewers. Telephone surveys are often used when the budget is limited. Scherpenzeel & Toepoel (forthcoming) experimented with face-to-face (Computer-Assisted Personal Interviewing—CAPI) and telephone (Computer-Assisted Telephone Interviewing—CATI) methods for recruiting a prerecruited internet panel of the full population. Due to over surveying, the response rates for telephone surveys have fallen in the last decades. In addition, the percentage of people with landline telephone numbers has decreased. The experiment was done to determine whether using the telephone for the first contact would lower the response rates. Scherpenzeel & Toepoel (forthcoming) also wanted to determine whether it was worthwhile to contact the whole sample face-to-face instead of using a combination of CATI (for those who have a known, regular telephone connection) and CAPI (for those who do not). Responses in recruitment are presented in Table 20.4. Responses were significantly lower for households without a known regular telephone connection. People in the CAPI mode with a known phone number were significantly more likely to express their willingness to participate in the panel. However, when analyzing the number of registered panel members (people

Table 20.4 Response in recruitment, in different interview modes. Percentage of the total gross sample

	CATI[a] Households with phone number (%)	CAPI Households with phone number (%)	CAPI Households without phone number (%)	Pearson chi-square
Not usable[b]	2	2	4	6.54*
Not reached	7	6	15	21.61**
Refusals	25	31	34	9.35*
Central questions only[c]	7	5	3	6.61*
Complete recruitment interview	59	56	43	22.86**
Willing to participate in panel	45	52	41	10.65**
Registered panel member	38	40	30	10.70**
Total gross sample	410	405	386	

[a] After 15 unsuccessful CATI contact attempts, households were transferred to the CAPI fieldwork. These households, 37 in total, are counted in the CATI response statistics

[b] Not usable includes, among other things, non-existing or non-inhabited addresses, companies, long-term infirm or disabled respondents, language problems

[c] The interviewer proposed to ask only one to three central questions if respondents refused the complete interview. If this was successful, the questions were followed by the request to participate in the panel

Significance * $p < 0.05$, ** $p < 0.01$

Source Scherpenzeel & Toepoel (forthcoming)

could agree to join the panel but fail to register as a panel member) the difference between CATI and CAPI (with phone number) is not significant. This suggests that CAPI (face-to-face) results in more socially desirable responses without actual registration and participation in the panel. With equal panel registration rates, the authors conclude that telephone recruitment works equally well and is cheaper than face-to-face recruitment when telephone numbers are available.

Paper-and-pencil surveys are less persuasive than face-to-face or telephone surveys since no interviewer is present. Yet, this remains one of the most effective methods of panel recruitment since response rates are relatively stable over time and costs are lower than face-to-face or telephone surveys.

A dramatic increase can be seen in the use of websites and e-mails as methods for panel recruitment in the twenty-first century. Sending out invitations via e-mail entails little cost and is very time-efficient. Success rates of e-mail recruitment vary greatly. In general, response rates to recruitment via websites are lower than response rates to e-mail invitations due to the personalization of the e-mail invitation (Lozar Manfreda & Vehovar 2008). Recently, attempts have been made to use social network sites and other digital media for panel recruitment, although the effectiveness of these new innovative methods is still unclear. In the next sections new means of panel recruitment will be discussed in more detail.

20.4 Recruitment Via Websites

Sending out personal invitations is far more effective than general invitations via websites. Yet, when no e-mail addresses are available, a

researcher has to make most out of the available options. For well-visited sites, recruitment via websites can be effective. Many research agencies form a strategic alliance with well-visited sites such as public transport OR television websites, in order to generate responses. Lozar Manfreda & Vehovar (2008) argue that the success of recruiting via websites depends on where (on which websites), on how many places, and when the invitation is posted. In addition, the invitation can be built in different ways, for example a banner ad or hyperlink, a pop-up window, or a simple hypertext link. A pop-up window draws most attention, but pop-up blockers in visitors' survey software make its effectiveness difficult to predict. A banner or hyperlink on the other hand may be easily overlooked. When possible, try to recruit via personal e-mail.

20.5 Recruitment Via E-Mail

There are pros and cons to using e-mail addresses for recruitment. Sending out e-mail invitations is cost-effective and fast. Hundreds of potential participants can be reached by one push on the "send" button, and responses usually return within a couple of days (with the major bulk being returned in a single day). The downside of using e-mail addresses for recruitment invitations lies in the frequent change of e-mail addresses, the fact that a mail cannot be addressed when there is (only) one letter misspelled, and spam-filters that may block the delivery of the email. In addition, many email users get an enormous amount of email in their inbox each day and the invitation might be easily overlooked or thrown away without being read. Also, since conducting online surveys is so easy, and many people take the opportunity of doing it, potential participants may receive many requests for competing surveys online and do not find the time or motivation to fill out every request.

Couper (2008) devotes almost an entire chapter on the design of the e-mail invitation. His discussion of the e-mail invitation is briefly described here. For more information see his book on "Designing Effective Web Surveys".

As illustrated in Fig. 20.1, the first task is to get the recipient of the e-mail to open the message. At this point the only information available to the recipient is the sender's identity (e-mail address and/or name) and the subject of the message. The amount of information that can be conveyed within these two fields is quite limited; yet, this is an important step in the sequence of events or decisions that lead eventually to the recipient of the invitation starting the survey (Couper 2008, p. 306). Once the recipient opens the e-mail invitation, more information can be conveyed in the body of the message. The options available to the researcher increase, but the relative importance of the information in terms of encouraging response may decline as fewer recipients make it to this step. Even fewer people make it to the part where the survey is introduced in the e-mail message. The e-mail invitation basically consists of four parts, the sender, subject line, body of e-mail, and the survey introduction, which can be designed by the researcher.

20.5.1 Sender E-Mail

Nowadays, Internet users are overloaded with messages from strangers, i.e., spam or phishing messages. Separating the legitimate messages from the illegitimate messages is difficult. E-mail invitations for surveys where no prior relationship exists between sender and potential respondent are likely to get little attention in these days of junk mail and spam. So, who will be the sender of the e-mail invitation? Some people argue that the impact of the invitation is bigger if the recipient of the e-mail invitation knows the sender of the message. Some also argue that is better if the recipient has a personal relationship with the sender, or if the sender is an authority or much respected person. Manipulations of the name of the sender are often limited. Yet, if it is possible, try to send the invitation from an individual who knows the recipient of the e-mail invitation or who can be

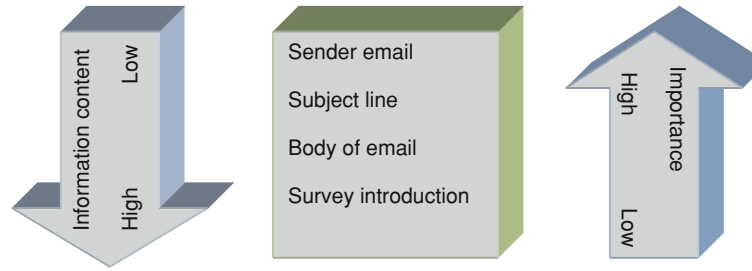


Fig. 20.1 E-mail invitation: information content and importance of the information. Adapted from Couper (2008)

viewed as an authority in some sense. In addition to the sender's name, you can choose the e-mail address of the sender. Again, authorities that are well known and respected are assumed to be more easily trusted and e-mail from these figures is more likely to be opened. Therefore, sending an e-mail from Tilburg University (ilse.oomen@tilburguniversity.edu) is more likely to be opened than an e-mail from a relatively unknown research agency (ilse.oomen@leisurepanel.nl). Little empirical evidence exists, however, that the sender of the invitation has a significant impact on the outcome of the survey invitation.

20.5.2 Subject Line

Subject lines for e-mail invitations of online surveys fall mainly in the following categories:

1. Identifying the panel or survey request (keywords: panel, survey, research, opinion)
2. Identifying incentives for participating (keywords: win, dollar, prize, lottery)
3. A plea (keyword: help).
4. Reinforcement of the nature of the relationship between sender and recipient (keywords dependent on the relationship, i.e., previous consumer, contact, etc.)

Little is known about the relation between subject lines and response rates. Some researchers experimented with the subject line, but results show mixed findings. For example, Zhang (2011) framed the subject line either as "to win" or "to help", and found a five-percentage point advantage for the "win" version. The effect became smaller at multiple contacts. Kent & Brandal (2003), on the other hand, found that a prize subject line produced a lower response rate than a subject line that stated that the e-mail invitation was about a survey. Toepoel (2011a, b) found no significant differences in responses to several types of subject lines (including all four types of subject lines mentioned above). The effectiveness of the subject line is probably dependent on numerous factors such as characteristics of the sample person, the sender, the survey, etc., making it difficult to predict the impact on the outcome of the invitation.

Note that the text in the subject line may depend on whether you are making a single request or one of many requests. The text in the invitation to recruit panel members may therefore be containing other information than the text that goes along invitations to fill out a survey where it is necessary to identify each survey request. Once participants have joined the panel, and a level of trust between the researcher and the panel member is established, clearly differentiating one invitation from the next may be important. On the other hand, revealing the subject of the survey may encourage only those panel members who are interested in the particular topic of the survey.

20.5.3 Body of E-Mail and Survey Introduction

While the knowledge on the effectiveness of the type of subject line is small and unclear, the knowledge we have on the body of the email text is even less developed. Part of this problem is related to the fact that there is a wide range of opportunities in the body text, and its effectiveness depends on so many factors that it is difficult to predict or experiment with while warranting

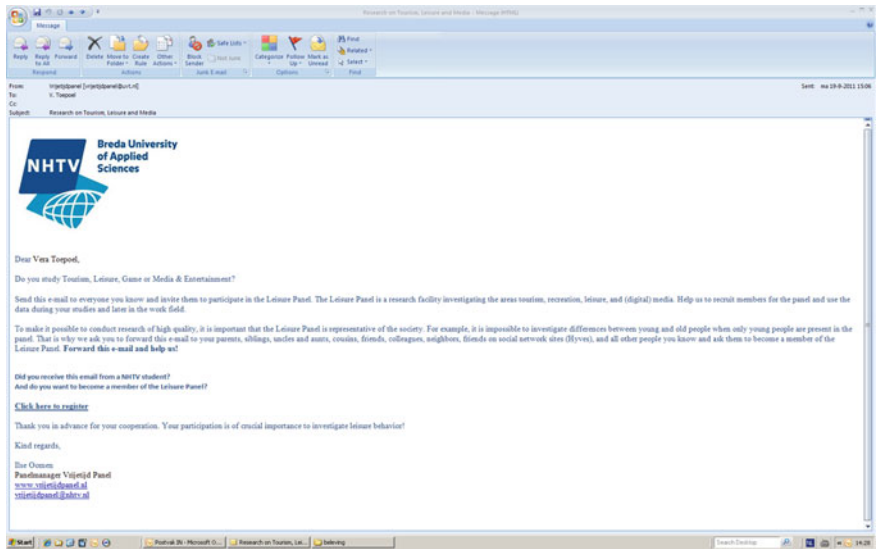


Fig. 20.2 Example of an email invitation

any general conclusions. You can send an email as plain text or as an HTML. HTML emails are more attractive and can include company logos. Many email systems either block or strip HTML emails, however. Figure 20.2 presents an example of an email invitation. The body of email contains different elements for which several design decisions have to be made:

1. Salutation: personal (e.g. Dear Vera) or impersonal (Dear Sir/Madam)
2. Content of the invitation: short or long, order of elements, email content versus welcome page content
3. Survey introduction including URL for the web survey
4. Authority of email signatory: high (dean) or low (administrative assistant)
5. Profile of the requesting office: high (Statistics Netherlands) or low (market research agency)

Some researchers have found that personalization increases response rates, while others have found no significant effect of salutation. In addition, some researchers argue that personalization increases social desirable answers. For questionnaires where social desirability is not a major issue, personalization could help to increase response rates.

The invitation should be short and contain all the crucial information. Avoid the necessity to scroll. Leave details to the Web survey welcome page or to accompanying website links. The elements should be ordered in a natural reading order following from general and important to specific or unimportant. The e-mail text should start with some sort of inviting and interesting statement. This should draw people's attention and motivate them to read the rest of the text. Then, the request is made. The survey has to be introduced, e.g., the goal of the survey, ways of recruitment, topic, researchers involved, etc. In addition, some kind of confidentially statement is warranted. Then the URL is provided. This should all be seen in a single screen, without requiring the respondent to scroll. Remember that the most important elements of the e-mail message are those that get respondents to the Web survey, and these elements should be prominently and clearly displayed. The goal should be to make filling out the survey as easy as possible.

The URL is the Web address of the survey. Clicking on the URL should launch the survey directly from the e-mail client. The URL should be short, understandable, and easy to retype. This is because long URLs may wrap over two

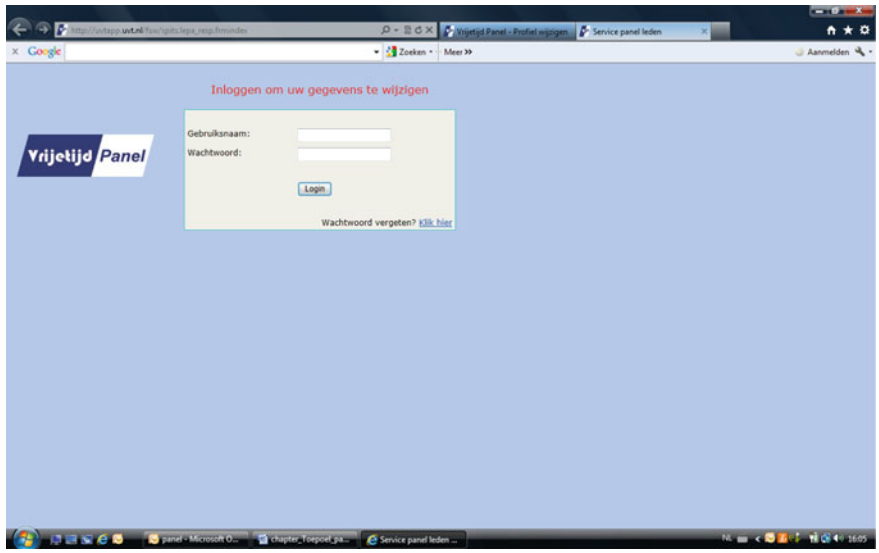


Fig. 20.3 Manual login with username and password required (in Dutch). For automatic login see Fig. 20.2

or more lines causing the link not to work or to be difficult to copy and paste. Also, some respondents might need to retype the URL if the link does not work. The URL should be placed right after the most important invitation information, so respondents do not need to scroll down in the message to see the link. Normally you want to restrict participation to those who are invited, and invite people to only fill out the survey once. Login processes are therefore important filters. Good survey software allows users to login multiple times (so that they can break off the survey and login again at their own time and pace), but once the survey has been finished the survey should be shut down. Automatic login procedures use a unique ID and/or password embedded in the URL, resulting in a unique URL for each respondent. Manual login approaches use an ID in the URL but require a password, or require that the respondent fills out both an ID (username) and password. For an example, see Fig. 20.3. Research suggests that automatic logins result in higher responses compared to manual logins, possibly explained by less effort that is needed from the respondent in the first case.

With regard to the sender's information details: contact information is always desirable, for example a phone number, personal mailing address, or additional information on an accompanying website. The norm of reciprocity and openness should outweigh any concerns about the sender of the invitation (Couper 2008). As mentioned earlier, signatures from authorities and high profile offices are believed to work positively on the outcome of the invitation.

20.6 Recruitment Via Digital Media Such as Social Network Sites

Apart from recruiting via e-mail, new methods of panel recruitment have become available with the advent of social network sites such as Facebook, LinkedIn, Hyves, Twitter, etc. These social network sites have an enormous spread of users. They can serve as a recruitment channel for online panel members because most network sites offer good targeting options (e.g. age, gender, location, interests), have low-entry level costs, have a global reach, and offer do-it-yourself systems. The reach of each network site differs per country. Table 20.5 shows the number of users and the population penetration of the top ten countries on Facebook. The United States has the highest penetration rate (49%),

Table 20.5 Top 10 countries on Facebook 2011

No	Country	Users	Change	(±%)	Population penetration (%)
1.	United States	150,800,200	−3,096,800	−2.01	48.61
2.	Indonesia	40,144,320	+4,969,380	+14.13	16.52
3.	India	35,623,220	+13,565,940	+61.50	3.04
4.	Turkey	30,545,000	+4,127,180	+15.62	39.26
5.	United Kingdom	30,168,540	+905,170	+3.09	48.39
6.	Mexico	28,963,320	+7,071,300	+32.30	25.75
7.	Brazil	26,829,900	+13,263,270	+97.76	13.34
8.	Philippines	26,493,340	+3,577,080	+15.61	26.52
9.	France	22,806,420	+1,769,080	+8.41	35.21
10.	Germany	20,951,460	+4,549,520	+27.74	25.46

Source www.socialbakers.com/facebook-statistics

although the number is slightly decreasing in the last year. Countries such as Turkey and Germany show a high increase in the number of users in the last year.

Most social network sites offer opportunities to send requests or place ads on people's personal pages. These systems are easy to use, making them available to the masses. Placing an ad on Facebook, for example, is fairly easy. Facebook has a lot of demographic information available that you can use to target your advertisement. You can select age groups, gender, and even region. Advertisements are shown according to the offer you make for placing the ad: the higher the payment, the more exposure the ad gets. Payment can be made per view or per click. In the case of panel recruitment, paying per click seems the best thing to do. Toepoel (2011a, b) experimented with recruiting panel members for an online research panel in the Netherlands via Facebook. Results are presented below.

Example: Recruiting via Facebook

Toepoel (2011a, b) placed an ad on Facebook users personal pages, as can be seen in Fig. 20.4. The advertisement is placed on the right-hand side of the page. By placing an ad, you can add a title ("Leisure Panel"), small text (Dutch universities want to know your activities when it comes to sports, culture, media, and holidays. Fill out the survey!), and picture (Taj Mahal). The picture of the Taj Mahal was chosen to draw people's attention.

A pilot was run to see how efficient recruiting via Facebook could be. Results are presented in Table 20.4. The fieldwork period was 3 weeks in April/May 2011. Facebook generated an offer per click of 60 eurocents. The actual payment was lower (probably due to low advertisement activity on Facebook) so the actual amount paid per click was 55 eurocents. The advertisement was placed more than 2.6 million times on Dutch people's personal Facebook pages. The age selection was 16 and older. Facebook provides no information on how many times individuals were shown the ad, but the ad was probably shown several times to the same individuals. Unfortunately, a researcher has no control over the number of times each Facebook user will be shown the advertisement. The more than 2.6 million views resulted in only 551 clicks, with the ratio clicks/view being only 0.0002%. Unfortunately, a large number of people only clicked on the ad, but did not start the questionnaire. The number of people starting the questionnaire was 54, which is about 10% of the people clicking on the advert. In the questionnaire, people answered a number of questions, including whether they wanted to be a panel member or not. Only one out of four agreed to be a panel member, resulting in 13 panel members. The cost per panel member was 28 €.

Analysis of the advertisement statistics show that almost 2 million of the 2.6 million views were placed on users aged 16 or 17, a group that

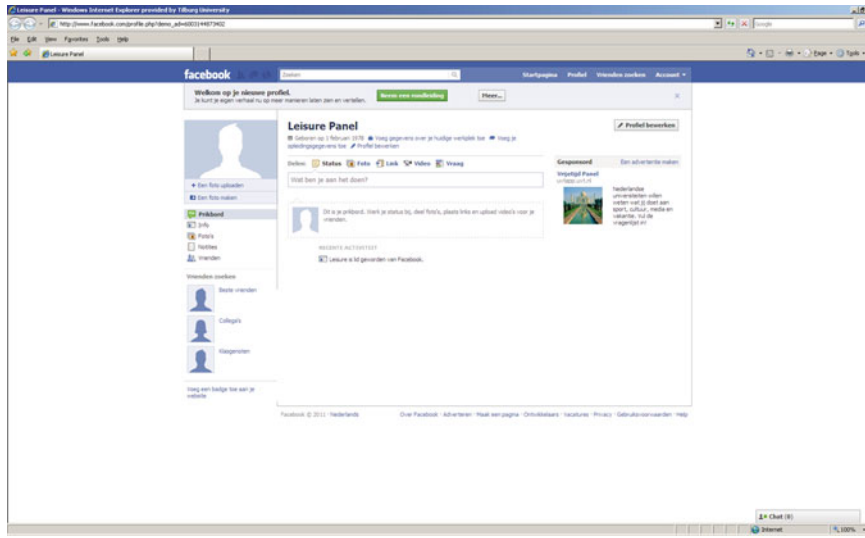


Fig. 20.4 Advertisement for panel recruitment on Facebook

might not be of interest to many other companies placing ads on Facebook. Although the advertisement was placed mostly on adolescents' personal pages, the older age groups seem to be more eager to agree to the panel request. The assumption that older people are more willing to join the panel was tested in another experiment. This time Facebook did not spend the budget, and for some unknown reason did not show the ad as often as in the first pilot (although the fieldwork period was extended a number of times and the offer per click was raised). The relation between age and willingness to become a panel member via Facebook could therefore not be tested. Toepoel (2011a, b) concludes that recruiting via Facebook cannot be recommended due to poor quantity, quality, and control over the recruiting process (Table 20.6).

The effectiveness of placing ads on Facebook as a recruiting method seems to be poor since the researcher has no control over the number of advertisement places (views) or where (for whom) the advertisement is shown. In addition, response quantity and quality seems to be poor. Using a snowball method for recruiting via Facebook or other social network sites might be more effective since it is based on the social relationship between users, but users might not be inclined to spam their friends and relatives with survey requests. Therefore, using social network sites as recruiting method has potential, but so far there is no evidence that it can work effectively.

20.7 The Optimal Strategy for Recruiting an Online Panel

Table 20.7 presents recruitment strategies and their expected effect on response rates. Paper-and-pencil, telephone, and face-to-face are traditional methods of panel recruitment. They are regarded as methods that are more difficult than new methods such as recruitment via websites, e-mail, and social media. Paper-and-pencil (mail) surveys are burdensome because respondents have to mail them back to the research agency. Telephone and face-to-face surveys are burdensome because respondents cannot fill them out at their own time and pace, but instead are forced to answer questions according to the interviewer's schedule. Traditional methods score better on legitimacy. Face-to-Face surveys are often associated with legitimate authorities, for example national statistical organizations. Paper-and-pencil surveys are also often seen as

Table 20.6 Pilot experiment for panel recruitment on Facebook

Fieldwork	April 21th–May 11th (3 weeks)
Budget	357 euro inclusive of tax
Offer per click	60 eurocents
Actual payment per click	55 eurocents
Number of views	2,607,823
Number of clicks	551 (.0002%)
Questionnaire started	54 (10%)
Agreed to be a panel member	13 (24%)
Costs per respondent (questionnaire)	6.61 euro
Costs per panel member (person agreed to join the panel)	27.46 euro

quite legitimate since the research agency took the trouble to print them and send an envelope with them, showing that they are prepared to incur costs for the survey research. Telephone surveys, on the other hand, are often associated with call centers and people trying to sell products via the phone, therefore putting a strain on legitimacy. Traditional methods are personal in the sense that an interviewer is available and a personal relationship can be established between interviewer and interviewee. In addition, they often make use of personal information such as name, telephone number, and address. Paper-and-pencil surveys that are not personalized are seen as less legitimate.

New methods such as recruitment via websites, e-mail, and social media are less burdensome than traditional methods since respondents can click on them or answer them at their own time and pace. They are often seen as less legitimate. Recruitment via websites is often associated with pop ups and banners of commercial companies trying to sell their goods and services. Recruitment via e-mail is often associated with spam. Little is known about recruitment via social media, although one could argue that people are not very keen on recruitment messages via a medium that is supposed to be regarded as "for friends only". Social media sites such as Facebook are often in the news because of security leaks and privacy violations; they cannot be classified as legitimate sources of panel recruitment. Social media scores well on personality, on the other hand, which can make then a good panel recruitment strategy if legitimacy is less of an issue. E-mail can also be regarded as personal; at least when the invited person knows the sender of the e-mail and the invite is personally addressed. Websites are not a very personal means of recruitment and are therefore less effective as a strategy for panel recruitment.

In all, strategies for panel recruitment involving e-mail and social media can be effective if the research agency can count on some sort of legitimacy. Legitimacy is apparent when a personal relationship between the potential panel member and the research agency is already established, for example because of prior contact, or because the research agency is well known and well regarded. If legitimacy is an issue of concern, it might be better to (also) use more traditional methods such as paper-and-pencil, telephone, or face-to-face contact. The latter two are more costly but interviewers could persuade potential panel members to reply in a positive manner and might therefore be very effective. A combination between new digital methods and traditional methods (mixed-mode) might work best, especially when making multiple contacts (Dillman 2007). For example, a mail announcement letter, e-mail invitation, and face-to-face reminder (to non-respondents) make use of an optimal combination of personality, legitimacy, and burden while being cost-effective.

Table 20.7 Recruiting methods and their hypothesized effect on response rates

		Personality	Legitimacy	Burden
Traditional	Paper and pencil	±	+	–
	Telephone	++	–	–
	Face to face	++	++	–
Modern	Websites	–	–	+
	Email	±	–	++
	Social media	+	–	++

20.8 Best practices (Do's and Don'ts)

20.8.1 Do

- Use different recruitment strategies, combining methods of recruitment.
- Send personal invitations.
- Establish a level of trust between you and the potential panel member.
- Send e-mails via legitimate authorities.
- Reduce the burden of completing a survey or panel request.
- Provide a simple link to the questionnaire for the recipient to use.
- Provide all relevant information to respondents while being brief.
- Use a probability sample (pre-recruited panel of the full population).

20.8.2 Don't

- Use general invitations, i.e., on websites.
- Use social media.
- Generalize conclusions to the full population if the only method of recruitment is online.

20.9 View to the Future

Online recruitment is an effective mode of panel recruitment when dealing with closed populations with high Internet penetration rates, when probability sampling or representativeness is not essential, when legitimacy can be relied upon, and when personalization is possible. Closed populations provide the optimal situation for online panel recruitment since they typically have existing sampling frames such as e-mail lists of employees, members, students, etc. Potential respondents will be familiar with the survey agency and may regard the survey as official business. In addition, the subject matter might be of interest and relevance to the potential panel members (Sue & Ritter 2007).

Online surveys do not allow for probability sampling of the full population. Although Internet penetration rates are increasing at a rapid pace throughout the world, the difference between the "haves" and "have nots" is larger than ever before, resulting in a sort of digital divide. Unless Internet penetration rates equal 100% and a sampling frame of e-mail addresses is available, online panels will never be without coverage bias. This is in addition to other forms of biases such as non-response bias, measurement bias, and sampling bias. When representativeness is not a major concern or when Internet penetration in the population is large, online panels can be an extremely effective means of data collection.

Technological barriers such as spam filters, pop-up blockers, and the like will still present challenges for online survey researchers in the future. Recruitment via e-mail has increased at a rapid pace in the last decade due to the fact that online surveying has become available to the masses. We can see drops in response rates to online surveys already, suggesting that recruitment via e-mail will show a similar pattern as telephone surveys: from a dramatic increase to a drop due to over surveying of the population. Since privacy and online security are a major concern with social media, the effectiveness of

recruiting via social media will not be much higher than recruitment via e-mail. In all, recruiting panel members via e-mail and social media has a lot of potential, but since there are also important drawbacks associated with these new forms of panel recruitment, survey researchers will still have to focus on optimal (mixed-mode) strategies in panel recruitment, and may create new recruitment strategies. These days everybody is posting information on the Internet about a vast range of topics. Recruitment may no longer be an issue. The information is already there. Just find out how you can get it and what the quality is. Questioning may not be interesting anymore. Listening (scanning) will be. Why ask some people when you can listen to all (or most) of them?

Questions

1. What are traditional forms of panel recruitment and what are new ways of panel recruitment?
2. What are the different types of online surveys and which of these types can be generalized to the full population?
3. What are the characteristics of the e-mail invitation and which design choices have to be made when designing an email invitation?
4. How can you recruit via social network sites, is this effective, and what can be done for increasing effectiveness?
5. Discuss how recruitment methods differ with regard to personalization, legitimacy, and burden and how this relates to response rates.

References

Baltagi, B. (2001). *Econometric analysis of panel data*. Chischester: Wiley.

Couper, M. P. (2008). *Designing effective web surveys*. New York: Cambridge University Press.

Couper, M. P. (2000). Web surveys. A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494.

Das, M., Ester, P., & Kaczmirek, L. (2011a). *Social and behavioral research and the internet*. New York: European Association of Methodology.

Das, M., Toepoel, V., & Van Soest, A. (2011b). Non-parametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods and Research, 40*, 32–56.

De Leeuw, E., Hox, J. J., & Dillman, D. (2008). *International handbook of survey methodology*. New York: European Association of Methodology.

Dillman, D. (2007). *Mail and internet surveys. The tailored design method*. Hoboken: Wiley.

Kent, R., & Brandal, H. (2003). Improving email response in a permission marketing context. *International Journal of Market Research, 45*, 489–506.

Lee, M. (2002). *Panel data econometrics*. Boston: Academic Press.

Lozar Manfreda, K., & Vehovar, V. (2008). Internet surveys. In: E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology*, (pp. 264–284). New York: European Association of Methodology.

Scherpenzeel, A., & Bethlehem, J. (2011). How representative are online panels? Problems of coverage and selection and possible solutions. In: M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet*. New York: European Association of Methodology.

Scherpenzeel, A., & Toepoel, V. (forthcoming). Recruiting a probability sample for an online panel: effects of contact mode, incentives and information. *Public Opinion Quarterly*

Sharot, T. (1991). Attrition and rotation in panel surveys. *The Statistician, 40*, 325–331.

Sikkel, D., & Hoogendoorn, A. (2008). Panel surveys. In E. D. De Leeuw, J. J. Hox, D. A. Dillman (Eds.), *International handbook of survey methodology*, (pp. 479–500). New York: European Association of Methodology.

Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. London: Sage.

Toepoel, V. (2011a). *Response quantity, quality, and costs of building an online panel via social contacts*. Presented at the Annual American Public Opinion Research Congress, Phoenix.

Toepoel, V. (2011b). *Panel recruitment via Facebook presented at the internet survey methods workshop*. The Netherlands: The Hage.

Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of design in web surveys. Comparing trained and fresh respondents. *Public Opinion Quarterly, 72*, 985–1007.

Trivellato, U. (1999). Issues in the design and analysis of panel studies: A cursory review. *Quality & Quantity, 33*, 339–352.

Zhang, C. (2011). Impact on data quality of making incentives salient in web survey invitations. Presented at the Annual American Public Opinion Research Congress, Phoenix.

Further Readings

Couper, M. P. (2008). *Designing effective web surveys*. New York: Cambridge University Press.

Couper, M. P. (2000). Web surveys. A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494.

Das, M., Ester, P., & Kaczmirek, L. (2011a). *Social and behavioral research and the internet*. New York: European Association of Methodology.

Dillman, D. (2007). *Mail and internet surveys. The tailored design method*. Hoboken: Wiley.

Poynter, R. (2010). *Handbook of online and social media research*. Esomar: Wiley.

Does Paying More Mean Getting a Better Product: Comparison of Modes of Survey Administration

21

Beau Shine and Brandon Dulisse

21.1 Review of Literature

The Total Design Method devised by Dillman (1978), created a framework for mail and telephone surveys with a primary focus on achieving high response rates. TDM emphasized the building of respondents trust and personalization, accompanied by the perception of increased rewards and reduced costs for respondent participation. The employment of these interconnected procedures resulted in high response rates, especially for mail surveys (Dillman 1978, 1991). This one-size-fits-all approach showed that general public populations could consistently produce high response rates, upwards of 70%. The reliance on one set of procedures Dillman (1999) claimed was also the biggest shortcoming of TDM. The proliferation of the Internet and the level of technology reached have made it impossible for most populations to be reached by a single mode survey.

The development of new technologies such as: increases of cell phone use and e-mail, decline in telephone response rates, and coverage problems all contributed to the emergence of mixed mode surveys (Dillman et al. 2009; Singer 2006). By adding a second or even third mode, researchers believed that they could increase response rates in hopes of reducing the potential for non-response error. In congruence with this argument, prior research has suggested mixed mode surveys can achieve higher response rates than single mode surveys. De Leeuw (2005) reported that using a second or third mode may improve response rates as well as coverage. Furthermore, some research suggests that mixed mode approaches may increase respondents' motivation because participants may appreciate being able to choose their response mode (Dillman 2000; Schaefer and Dillman 1998).

Mixed mode surveys are not without limitations as well as potential problems that have been pointed out by previous research. According to Dillman (1999), there are four main reasons differences between various modes: social desirability, acquiescence, question order effects, and primary/recency effects. These differences constitute response bias issues, which indicate problems in *how* one responds to survey questions (which it is important to note is different from non-response bias which is concerned with *who* responds). One major potential problem between modes derives from measurement differences leading to unique respondent answers depending on survey mode (De Leeuw 1992; Dillman 1999). More specifically, several decades of mode experiments have appeared in the survey

B. Shine (✉) · B. Dulisse
School of Criminal Justice, University of Cincinnati,
665 Dyer Hall, Clifton Avenue, P.O. Box 210389,
Cincinnati OH 45221-0389, USA
e-mail: shinebj@mail.uc.edu

B. Dulisse
e-mail: dulissbc@mail.uc.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_21, © Springer Science+Business Media New York 2012

literature suggesting that significant differences frequently occur in the answers that people give to aural and visual surveys (Dillman et al. 2009). Evidence from separate studies have shown that aural respondents tend to give more positive extreme answers than do mail respondents when answering opinion questions (Dillman and Mason 1984; Tarnai and Dillman 1992; Krysan et al. 1994). In addition, studies comparing telephone and web respondents found differences in responses for various kinds of scale questions with telephone respondents giving more positive answers (Christian et al. 2008). Ramifications for these differences in mixed mode surveys may result in different analytical conclusions and recommendations (Cobanoglu et al. 2001).

The observable advantages and limitations of survey method techniques illustrate the need for a cost/benefit analysis taking into account the exhaustive research on unimodal and mixed mode surveys. For example, implementation approaches that may be beneficial for mail surveys may not directly translate into Web-based survey benefits (Couper 2000). Issues such as these create saliency concerns that may result in the misinterpretation of measures of equivalency such as: response rates, response time, and costs. In an effort to reliably choose a survey strategy that most appropriately corresponds with the research being conducted, researchers must understand and demonstrate this equivalency, or relative strengths of alternative modes (Dillman 2000).

In recent years, the popular press has reported that response rates for all types of surveys are declining (Bickart and Schmittlein 1999). Research has also found that the US population is being over-surveyed through an exponential growth in the amount of survey research which has resulted in an increase in the number of requests for survey response (Groves et al. 1992). Whatever the reason may be, low response rates are a concern for researchers and have the possibility of creating a biased estimate of the characteristics of the population (Bean and Roszkowski 1995). These concerns make choosing the most effective survey mode even more invaluable, by weighing the costs and benefits of each and deciding on the most appropriate methodology.

21.2 Mail Surveys

Mail surveys are used more than any other mode of survey administration. Mail surveys have been used for decades, and according to Dillman's Total Design Method (Dillman 1978) Total Design Method (Dillman 1978), mail surveys allowed for a cookie-cutter approach to sampling various populations of interest. And while mail surveys do offer several advantages over other survey designs, research has since revealed clear limitations of the survey mode as well (Dillman 1999; Cobanoglu et al. 2001). This section will examine the advantages and disadvantages of mail surveys, paying close attention to how changes in technology have affected this mode of administration over time.

21.2.1 Strengths

Traditionally, mail surveys have been employed more frequently than any other type of survey. And while response rates varied in their predictability prior to Dillman's Total Design Method (1978), according to Donald Dillman, survey designers could achieve, on average, a 74% response rate using mail surveys that followed his method. Such a percentage is very respectable, especially in light of the relatively low cost of administering such a survey (see below).

Another strength of mail surveys is that they were traditionally able to reach larger geographic populations than face-to-face (FTF) interviewers or telephone interviews (Cobanoglu et al. 2001). This is because staffing expenses were next to nothing, as were long-distance expenses. For the cost of a preaddressed stamped envelope (which is a fixed price nationwide), survey administrators were able to tap into a larger sampling frame, and thus, generalize results to a larger proportion of the U.S. population. Also, along with reduced staffing expenses, much less training and requirements are

needed for those who are employed to help conduct the survey.

Finally, mail surveys do not have to be concerned with the interviewer and response biases as much as their telephone and FTF counterparts (Groves 1989). This is because survey respondents have privacy and anonymity when filling out mail surveys, thus minimizing the likelihood of response bias and eliminating the opportunity for interview bias altogether.

21.2.2 Limitations

While response rates for mail surveys were strong prior to the advent of the Internet, since then they have declined precipitously (Dillman 1999; Cobanoglu et al. 2001). Mail survey response rates have plunged from 71% in the late 1970s to 38% in 1992 (Parker 1992), all the way down to 26.27% in 2000 (Cobanoglu et al. 2001). In fact, Donald Dillman addressed the decrease in response rates between his 1978 Total Design Method in his revised work, "The Tailored Design Method" (1999), noting that with the advent of the fax machine and the world wide web, mixed modes of data collection are now superior to their unimodal design counterparts, yielding better response rates and more representative samples.

In addition to the decline of response rates over time, mail surveys have an additional limitation: since there is no interviewer, there is no opportunity for question flexibility, and/or interviewer assistance in explaining the questions (www.cdc.gov). This lack of flexibility and opportunity for an interviewer to clarify questions regarding the survey makes answering the survey more difficult for respondents, particularly those who are cognitively challenged.

Finally, since response rates on mail surveys have dropped so sharply over the years, there is now reason to believe that individuals who choose to respond to mail surveys may be different from those who do not, suggesting self-selection bias may be a concern. Self-selection bias is a form of non-random error that occurs when "respondents having the same observable demographic characteristics, but different unobservable characteristics, respond to a survey with a different likelihood" (Ethier et al. 1997).

For a full discussion on the advantages and limitations of mail surveys see Chap.17 by Henninger and Sung.

21.3 Telephone Surveys

Another commonly employed mode of survey administration is telephone surveying. Telephone surveying has been used for decades, as it has been estimated that 96% or more of all residences in the United States have a minimum of one telephone (American Statistical Association 1999). And in fact, following Dillman's Total Design Method (1978), response rates reached upwards of 91% through the late 1970s and early 1980s. However, over time and particularly since the proliferation of cellular phones and e-mail, telephone response rates have declined precipitously (Curtin et al. 2005). This section addresses the strengths and weaknesses of telephone surveys, paying close attention to the changes in response rates over time.

21.3.1 Methods

In general, there are two ways to reach out to prospective respondents: through telephone directories and random dialing. Each method will be discussed below, along with its advantages and disadvantages.

21.3.1.1 Telephone Directories

One way to reach potential survey respondents is to use telephone directories. While this would seem like a terrific resource, there are many limitations to using such a method (American Statistical Association 1999). First, not all residential lines are published in directories. Many residents choose not to appear in directories. This allows for sampling bias, a form of non-random measurement error. In fact, residences with unlisted phone numbers are likely to have lower incomes, more single-headed households, and be more concentrated in heavily populated

locations such as large cities and metropolitan areas.

Additionally, telephone directories exclude new phone numbers that are created since a given directory's publication (American Statistical Association 1999). In fact, between unlisted and unpublished phone numbers, it has been estimated that at least 30% of dwellings with telephones are not found in directories.

21.3.1.2 Random Digit Dialing

Random digit dialing (RDD) is another way to reach potential respondents. The theory underlying the method is that choosing a given area code and prefix (the first six numbers of any phone line), and choosing the last four numbers at random will prevent the sampling bias associated with telephone directory sampling. However, its strength also comes with quite an expense: it is very costly and time-consuming to call so many inactive phone numbers (American Statistical Association 1999). Additionally, when attempting to calculate total error estimates, response rate calculations may be inaccurate due to difficulty in determining whether certain phone numbers are valid (Massey et al. 1997; Sangster and Meekins 2003).

Variations in RDD techniques have been used to alleviate the concerns of RDD, including Mitofsky–Waksberg sampling and list-assisted telephone sampling; however, each method has its own set of limitations, and such methods are beyond the scope of this chapter.

21.3.2 Strengths

Historically, telephone surveying has been a popular survey mode of administration (Dillman 1978; American Statistical Association 1999). Given that estimates of households with at least one telephone have exceeded 96% (American Statistical Association 1999), sample representativeness has traditionally been viewed as one of the mode's advantages. Sample representativeness is an important consideration for surveyors, because omitting residences without phones from a sample may lead to nonrandom measurement error, thus biasing the results of the survey.

Another advantage of telephone surveys is that they are less expensive to conduct than FTF interviews (American Statistical Association 1999). Additionally, interviewer effects are reduced in phone surveying compared to FTF interviewing, reducing the potential for response bias. This is because phone surveys offer more anonymity than FTF interviews, which in turn triggers more honest responses from the respondents relative to their FTF counterparts.

While telephone surveys are more expensive to conduct than self-administered mail surveys, the data are collected quicker, and thus, available for analysis much sooner (American Statistical Association 1999; McNamara 2009). Thus, if a quick turnaround time is sought, telephone surveys have a distinct advantage over mail surveys. Additionally, telephone interviews offer the opportunity for additional qualitative data, as respondents have the opportunity to discuss issues with interviewers that could not otherwise be examined without such interpersonal interaction. Finally, quality control is improved with telephone surveys relative to mail surveys, as trained individuals are entering the data (McNamara 2009).

21.3.3 Limitations

What was once viewed as a major strength of telephone surveying has since become one of its greatest weaknesses: response rate. By the 1970s, less than 10% of U.S. households did not have telephones, making a large proportion of the population easily accessible (Massey et al. 1997). In fact, by the 1980s, response rates exceeded 70%. However, with the advent of the Internet and cellular phones, response rate has decreased significantly. Moreover, answering machines and caller IDs have also inhibited survey response. According to the American Statistical Association (1999), upwards of 55% of all households with landlines use answering machines to screen their calls. And to make matters worse for telephone surveying, the national Do Not Call registry has enabled households to block themselves from receiving unsolicited phone calls, further weakening

response rates and potentially affecting sample representativeness.

Another noteworthy limitation of telephone surveys is the lack of visual materials for a respondent (McNamara 1999). The lack of visuals means the respondents cannot read (and reread) the questions, inhibiting their understanding of the topic(s) being covered. In addition, respondent inattentiveness is a concern as well, as people often multi-task—or carry out other responsibilities—while on the phone. Also, telephone surveys using RDD are no longer representative of the United States' population, as home phones tend to be maintained by older respondents. Additionally, respondents are not usually willing to spend more than 15–20 min on a telephone survey. Finally, data have suggested that Internet surveys are preferred to their telephone counterparts, another potential factor that could harm response rates. For a more thorough discussion on phone and cell phone surveys, their advantages and disadvantages refer to Chap. 16 by Vehovar, Slavec and Berzelak.

21.4 Face-to-Face Surveys

Face-to-face surveys are surveys conducted by interviewers who are physically present and personally ask respondents the questions on a survey (De Leeuw and Van Der Zouwen 1988). This mode of survey administration is quite unique, and has distinct advantages and disadvantages over all other modes of survey administration. Each of these strengths and limitations are discussed below.

21.4.1 Strengths

One of the major strengths of FTF surveys is that they allow respondents to ask the interviewer questions or clarify items on the survey that they do not understand (Doyle, n.d.). This strength is particularly well suited when surveying respondents who are illiterate, have poor writing skills, or have physical or mental disabilities that would prohibit them from completing a survey administered through the mail, telephone, or online. Additionally, FTF interviewers can bring visual aids that can provide additional assistance to respondents, should they need it.

Another major advantage of FTF surveys is that they allow for more data to be gathered than do other modes of survey administration. While telephone respondents typically would not answer surveys longer than 15–20 min, most respondents will spend up to an hour answering questions with a FTF interviewer (Doyle, n.d.). Of equal importance is that the additional response time allows for longer, more in-depth questions that can provide additional data that would be left unexamined in shorter surveys.

An additional strength of FTF surveys is that well-trained interviewers have control over the quality of the data being collected (Doyle, n.d.). By engaging in an interpersonal exchange FTF, respondents are more likely to focus exclusively on the task at hand, thus lessening the opportunity for extraneous factors to bias their responses.

The greatest strength of FTF surveys, however, is the response rates they produce. According to Doyle (n.d.), FTF surveys yield the highest response rate of any type of survey. Potential respondents are much less likely to turn away someone at their doorstep than they are to decline a telephone or mail survey. In fact, some FTF surveys have obtained response rates in excess of 90%!

Finally, FTF surveys tend to produce a sample more representative of the general population than do mail surveys (Doyle, n.d.). This is because individuals who respond to mail surveys tend to be better educated than those who choose not to fill out mail surveys, for example.

21.4.2 Limitations

While there are distinct advantages to FTF surveys, they also have unique disadvantages. First and foremost is the tremendous financial cost of conducting an FTF survey. Such a survey design requires large amounts of money

for several reasons. First, the travel data collection expenses are enormous compared to other modes of administration. Hourly employees need to be paid, as do their expenses to and from the data collection points. In addition, interviewers must be carefully selected and trained, which is a large expense as well. In fact, research has shown that FTF surveys can be double the cost of telephone surveys, and can take upwards of three times longer to collect the data than its telephone survey counterpart (Groves 1989).

Tied into high costs is that FTF surveys are limited to a relatively small geographic region (Groves 1989). Due to budgetary constraints, conducting FTF interviews across the country would be next to impossible. Moreover, some types of respondents are not easily reached because they are not home often (such as college students), or access to their residences is restricted. Additionally, surveying some neighborhoods may place interviewers at an elevated risk of victimization.

In regard to the quality and reliability of the data, there is evidence suggesting personal questions are not as likely to be answered as honestly when a respondent is completing a FTF interview (Groves 1989). This should make sense—some people are less likely to be forthright with their responses when they fear being judged.

Finally, tied into data quality is the greatest limitation of FTF surveys: the potential for interviewer bias. Interviewer bias is a bias that occurs when interviewers—through verbal and/or nonverbal queues—influence how respondents answer questions. This concern, while slight in telephone interviews, is enormous in FTF surveys, and great amounts of interviewer training are needed to mitigate the potential for such nonrandom measurement error.

The advantages and disadvantages of FTF surveys were discussed earlier in this section by Neuman in Chap.14 and by Stoop and Harrison in Chap.15.

21.5 E-Mail Surveys

21.5.1 Strengths

The use of e-mail as a method of survey distribution has been around as early as 1986, when results from the first e-mail survey were published (Kiesler and Sproull 1986). Up until the late 1990s, e-mail survey distribution was the most popular form of Internet surveying (Solomon 2001). The research regarding e-mail survey techniques have suggested several advantages compared to postal surveys in terms of response speed and cost efficiency. Studies measuring delivery speed found that mail surveys took an average of 11.8 days to return while e-mail surveys were returned in 7.6 days (Sheehan and McMillan 1999). E-mail surveys are also more cost-efficient according to studies, showing e-mail costing only an estimated 5–20% of a paper survey (Sheehan and Hoy 1999; Weible and Wallace 1998). A majority of the cost savings originate from the reduction/elimination of paper and mailing costs.

In addition to being cost-efficient and possessing quicker rates of response, some e-mail surveys offer software that allows for the precise tracking of surveys. Having this knowledge can help researchers know the number of undeliverable e-mails as well as when the e-mail survey was opened, replied to, and deleted; all of which can improve sampling procedures. E-mail surveys can also elicit heightened response quality; as research has suggested respondents tend to provide longer open-ended responses than to other types of surveys (Paolo et al 2000). These open-ended responses allow for more self-disclosing comments, possibly due to the speed of typing over handwriting (Bachmann and Elfrink 1996), although no research has carefully investigated this difference.

E-mail surveys also serve a geographical advantage over many other types of survey techniques. Worldwide e-mail surveys have the

potential to reach nearly every part of the world, where mail, telephone, fax, or even other web-based surveys may not. One example, Swoboda et al. (1997) performed a worldwide e-mail survey and achieved only a 20% response rate, but received responses from all parts of the world (90% of them within 4 days). E-mail users in developing nations can access e-mail surveys at the same speed as those in developed nations. This suggests that if the target population has e-mail and many live in remote places, e-mail provides the most effective communication method for quick data collection (Yun and Trumbo 2000).

21.5.2 Weaknesses

While e-mail survey techniques offer advantages compared to more traditional survey modes, there are also limitations that must be noted. When e-mail addresses are changed there is usually no way to forward e-mails that were sent to an old address (Cobanoglu et al. 2001). Finding participants e-mail addresses may also create potential problems as contacting e-mail addresses for people without a prior established relationship with the survey sponsor is considered an unacceptable survey practice (Dillman et al. 2009). Unsolicited e-mail invades an individual's private space, and sending too many of these messages or surveys will bother some people. The abuse of the e-mail survey may actually damage the e-mail survey environment (Yun and Trumbo 2000). Research in the 1990s suggested this was the case by predicting that the increase in junk e-mail would result in a reduction of response rates, similar to mail and telephone rates as well (Dillman 2000; Parker 1992; Schaefer and Dillman 1998; Tse 1998).

Limitations also arise with e-mail surveys with trying to obtain a representative sample (Dillman 2000). Internet users changing their Internet Service Provider and their e-mail address create a 'churn effect', changing the accessibility and representativeness of samples. Also, the possession of multiple e-mail addresses by one individual can increase under-representation (Bradley 1999). These issues, according to Dillman (2000), strengthen the argument for using Web-based surveys as an alternative to e-mail, since anyone with Internet access can respond to the survey. Mesch, Chap.18, discussed e-mail surveys in more detail.

21.6 Web-Based Surveys

The exponential growth of the Internet has influenced practically every feature of society. Survey methodology and research certainly are no exception. The use of the Web as a medium for survey administration offers significant advantages over previous modes of survey distribution (i.e. mail, telephone, and fax); however, methodological limitations still apply. One of the biggest challenges stems from the lack of knowledge on the most effective ways to conduct surveys over the Internet and their relatability to other survey modes (Solomon 2001). While much research is yet to be done on the most effective uses of web-based survey techniques, a large amount has already been published on the strengths and limitations of this method.

21.6.1 Strengths

According to Wyatt (2000), web-based surveys possess a number of strengths over more conventional-type surveys (mail, paper, or FTF). One such strength is the overall net financial cost of using web-based surveys. After initial setup fees, Web-based surveys are cheap to carry out, allowing for the recruitment of large numbers of participants or to collect data repeatedly, on multiple occasions. This allows for significant increases in sample sizes with minimal additional costs compared to other survey modes. While many mail and fax surveys require the use of purchasing pre-stamped envelopes or a toll-free fax number, Web-based surveys require no variable cost to the respondent. Web-based surveys also eliminate the need for printing and mailing of survey instruments as well as allowing researchers to acquire survey data that is already in electronic format (Cobanoglu

et al. 2001). Not only does this advance help to eliminate costs of data acquisition, but it also significantly reduces the labor required, especially when compared to mail surveys.

In addition to cost savings, Web-based surveys provide improved convenience and interactive feedback for respondents. More recent online questionnaires are often designed to provide feedback in the form of summary statistics about an individual's responses, which may act as an incentive to participate (Dillman 2000; Schmidt 1997). Online feedback also assists researchers as Web-based technology now allows for data capture with rapid checking of responses and immediate validity checking of individual data items. This feedback allows Web-based surveys to rapidly update questionnaire content and question ordering according to user responses, helping to control for recency, and other question order effects (Wyatt 2000). Web-based surveys can also be convenient for participants, since completion is based on the respondent's leisure, which may increase the likelihood of participation (Sax et al. 2003).

Research has also suggested that Web-based surveys yield higher response rates for college populations. According to Carini et al. (2001) and Handwerk et al. (2000), variables that positively predict response to Web-based surveys in comparison to paper surveys include: being a traditional-aged college student, living on campus, majoring in math/science, declaring multiple majors, attending a selective institution, and attending an institution that provides substantial academic support for students.

21.6.2 Weaknesses

While Internet access in the U.S. has been increasing, as of 2007, only about 67% of American adults (18 and older) currently possess access from their homes (Horrigan and Smith 2007). Although Internet access has steadily increased in the last 20 years, coverage is still not sufficient for general public surveys (Dillman et al. 2009). This coverage bias creates significant validity concerns since results are only generalizable to those who are keyboard and Internet literate; which in 2000, constituted only a third of the US population (Wyatt 2000). Technical difficulties may also increase this coverage bias. When browsers do not have the same capacity to view websites, it can cause surveys to have a different appearance from one respondent to the next (Dillman 2000). This may decrease response rate or even prevent an individual from accessing the survey (Smith 1997).

In addition to a coverage bias and reduced generalizability, the prevalence of non-deliverable surveys is greater for Web-based surveys as people are more likely to change their e-mail addresses and Internet Service Providers than postal addresses (Zatz 2000). This creates a problem since the increase occurrence of e-mail change potentially limits the ability to carry out repeated assessments of the same individual for epidemiologic purposes (Wyatt 2000). Furthermore, some of the more "technologically advanced" populations such as college students have been responding at lower rates than previous decades which may have biasing effects on the data (Bradburn 1992; De Leeuw and Heer 2002; Dey 1997; Fraenkel and Wallen 1993; Schuman 2002; Smith 1995; Steeh 1981). These effects create a non-response bias worth noting that may limit the generalizability of Web-based survey data.

Research has shown that incentives such as cash and non-cash enticements can significantly increase response rates. (Shank et al. 1990; Hare et al. 1998; Dillman 1999; also see Toepoel Chap.13). However, limitations arise with Web-based surveys since tangible incentives cannot be included in Web-based media like they are for mail and FTF surveys. One potential solution to this limitation, according to Cobanoglu et al. (2001), is the inclusion of coupon attachments. Another significant difference between mail and Web-based surveys stems from concerns of potential survey participants of Internet security and the receipt of electronic "spam" or "junk mail" (Smith 1997; Sills and Song 2002).

Research has suggested both advantages and limitations in regard to respondent concerns with Internet security. A study by Bachmann et al. (2000) found that responses to e-mail surveys

tend to be more candid than mail or phone responses, potentially leading to increases in response quality. While Web-based surveys may still be more candid than other survey modes, concerns of guaranteed anonymity by respondents of Web-based surveys may prevent respondents from submitting personal information and/or sensitive issues over the Internet (Wyatt 2000). These security and data integrity differences reaffirm that some implementation approaches that benefit response rates in other modes may not translate directly to benefits in Web-based surveys. Potentially, as a result of these concerns unique to Web-based methodology, research has shown that response rates tend to be lower for Internet surveys than for other modes (Cook et al. 2000; Couper 2000).

21.6.3 Summary

Web-based surveys hold clear advantages when compared to other survey modes in that they have the potential to save researchers time and money, as well as stimulating respondents and gather other important respondent data. Yet, for all of these benefits, there are still significant limitations with coverage, anonymity, and confidentiality, as well as technical problems that arise with the use of Web-based surveys. The constant evolution and progress of Internet and computer technologies create an opportunity for more research on Web-based survey techniques since most findings are inconsistent (Sax et al. 2003). Until then, Web-based models can be most effective in surveys designed by overemphasizing the benefits and limiting the significance of the weaknesses.

Manzo and Burke earlier in this handbook, Chap.19, further discuss the multifaceted issues revolving around Web-based/Internet surveys.

21.7 Mixed Mode Surveys

According to Dillman et al. (2009), mixed mode survey techniques became popular based on developing research indicating preferences by respondents to certain modes over others. The advent and proliferation of new technologies such as the Internet and cell phones have supported the popularity of this trend. These new networks, combined with a decrease in coverage for RDD surveys and declining home telephone response rates, have forced researchers to move in the direction of additional survey administration techniques. A mixed mode approach was developed with the goal of optimizing data collection while reducing total survey error (TSE), owing to time and budgetary constraints (De Leeuw 2005).

Mixing modes creates an opportunity for survey designers to balance the weaknesses of each individual mode at an affordable cost (De Leeuw 2005). For example, written modes (like mail surveys) may contain low question flexibility coupled with low requirements of survey staff. Other modes like telephone surveys offer high question flexibility, but very high requirements of survey staff. By combining these modes, the researcher has the best of both worlds: more flexibility in question design and less demand on survey staff. If the more expensive mode is used second, researchers would only need to survey non-respondents from the first wave that used the first mode of mail surveys. This technique combines less costs and less error than in a unimodal approach, which optimizes benefits while achieving a higher response rate (De Leeuw 2005; Dillman et al. 2009).

The first advantage of mixed mode techniques is the reduction of non-response error. In official statistics as well as in academic research, response rates have been on the decline over the years (De Leeuw and Heer 2002; Balden 2004). To reduce non-response while maintaining low cost, Fowler et al. (2002) conducted telephone follow-ups to initial mail surveys. Results indicated that these follow-ups actually raised overall response rates and may even reduce non-response bias in mail surveys. Mixed mode survey techniques have also been shown to reduce selective nonresponse. A study by Beebe et al. (2005) included statements on the cover of their mail survey that were in several languages. The statement urged respondents interested in

Table 21.1 Strengths and limitations of the modes of survey administration

Issue	Mail	Telephone	Face-to-face	Web based
Cost	Medium	High	Very high	Low/medium
Data collection time	Medium	High	Very high	Low/medium
Demands on/requirements of staff	Low/medium	High	High	Low
Question flexibility	Low	High	High	High

Information from this chart was provided by www.cdc.gov

completing a telephone survey to call the center where bilingual interviewers were available. These techniques were shown to raise the overall response rates compared to unimodal counterparts (De Leeuw 2005).

While the mixing of different survey techniques has shown to increase response rates, different strategies such as pre-notice contact and phasing can also impact response rates. Research studying mixed mode techniques has demonstrated that surveys using increased number of contacts and pre-notice contacts appear to have the strongest impact on response rates (Dillman 2000; Dillman et al. 1995). In a mixed mode survey of college students, Kaplowitz et al. (2004), combined a pre-notice postcard with an e-mail survey in comparison with four other mixed methods (mail; postcard, e-mail survey, postcard; e-mail, postcard; and e-mail only). The mail survey mode yielded a 31.5% response rate at a cost of $10.97 per response, while the pre-notice followed up by an e-mail survey yielded a 29.7% response rate at a cost of only $1.31 per response. The other survey modes yielded lower response rates and higher costs than the pre-notice postcard with e-mail survey design.

These results support previous research, suggesting that pre-notice increases response rates while also finding increased cost-effectiveness when compared to other modes. It is also important to note the substantial cost difference between the mail survey and the advanced notice e-mail survey. This evidence suggests that in populations with full Internet access, Web-based survey designs can achieve comparable response rates to mail surveys if the Web-based design is preceded by advanced notice (Kaplowitz et al. 2004; also see Mesch, Chap.18, Manzo and Burke, Chap.19). More broadly, these findings illustrate the potential for more cost-effective mixed mode survey designs based on the recognition of the populations intended for research.

Pre-notice contact is not the only technique that can be used in mixed mode surveys to achieve an increased response rate. Combining various survey modes by administering "phases" can have a positive effect on response rates as well. Phasing involves the administration of one type of survey technique, then switching to a second type of mode once the first phase has been completed. The second phase only targets non-respondents from the first phase. For instance, Dillman et al. (2009) achieved response rates of 80–83% overall by combining mail and telephone surveys, regardless of order. These response rates were higher than single mode designs, which yielded a 44% response rate for telephone and 75% for mail. These findings are similar to other studies looking at phasing with other modes such as mail-based and Web-based combinations (Börkan 2006).

Table 21.1 presents a comparative summary of the main survey modes—mail, phone, FTF, and Web-based—on their characteristics. Readers are advised to use this table for consideration of mixing their data collection. Specifically, and as can be seen from Table 21.1, whenever it is possible to begin data collection using mail or e-mail surveys, researchers should attempt to use this mode, as it is cheaper and may be highly effective in academic surveys where scales are used. Similarly, FTF option, being the most expensive data collection mode, should be reserved for such cases where more flexibility is needed, and where there is an added value

to having a trained interviewer to convert non-response. It is quite obvious from Table 21.1 that those modes that lean on interviewer interaction are the most expensive ones. However, they are also the ones that enable more question flexibility and conversion of non-response. Accordingly, it is recommended that these methods will not be the preliminary default for all survey topics (also refer to Table 11.2 in Chap.11 by Albaum and Smith for a more comprehensive evaluation of the alternative survey methods of data collection and their overall cost and benefit).

21.8 Conclusion

Mixed mode survey techniques have clear advantages compared to unimodal designs with regard to increasing response rates while also being relatively affordable. As previously discussed in the literature review, while mixed modes offer many benefits, it is important to acknowledge the possibility of mode effects. These mode effects should dictate how and when a survey designer uses a mixed mode approach to creating an effective survey. Many of the benefits of combining survey techniques may be lost if mode effects are not taken into account by researchers (Dillman et al. 2009). For example, a survey designer who uses a seven-point "agree–disagree" scale for a telephone mode may also offer the scale in a mail or Internet version. The question wording effects could add to significant measurement error (De Leeuw 2005; also see Gideon, Chap.7). According to Dillman and Christian (2003), questions should be constructed to provide an equivalent stimulus for respondents across various modes.

References

Bachmann, D., & Elfrink, J. (1996). Tracking the progress of e-mail versus snail-mail. *Marketing Research, 8*(2), 31–35.

Bachmann, D. P., Elfrink, J., & Vazzana, G. (2000). Email and snail mail face off in rematch. *Marketing Research, 11*(4), 10–15.

Balden, W. (2004). *Highlights from the session on Web-based data collection*. Federal Committee on Statistical Methodology: Statistical policy seminar session on web based data collection, December, Washington, D.C.

Bean, A. G., & Roszkowski, M. J. (1995). The long and short of it. *Marketing Research, 7*(1), 20–26.

Beebe, T. J., Davern, M. E., McAlpine, D. D., Call, K. T., & Rockwood, T. H. (2005). Increasing response rates in a survey of medicaid enrollees: The effect of a prepaid monetary incentive and mixed-modes (mail and telephone). *Medical Care, 43*, 411–414.

Bickart, B., & Schmittlein, D. (1999). The distribution of survey contact and participation in the United States: Constructing a survey-based estimate. *Journal of Marketing Research, 36*, 286–294.

Börkan, B. (2006). Effectiveness of mixed-mode survey designs for teachers using mail and web-based surveys (Doctoral dissertation). Retrieved from http://etd.ohiolink.edu/view.cgi/Borkan%20Bengu.pdf?osu1158597296.

Bradburn, N. M. (1992). Presidential address: A response to the non-response problem. *Public Opinion Quarterly, 56*, 391–398.

Bradley, N. (1999). Sampling for Internet surveys: An examination of respondent selection for Internet research. *Journal of the Market Research Society, 41*(4), 387–395.

Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2001). *College student responses to web and paper surveys: Does mode matter?* Paper presented at the 41st Annual Forum of the Association for Institutional Research, LongBeach, CA, June 4, 2001.

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2008). The effects of mode and format on answers to scalar questions in telephone and web surveys. In J. Lepkowski, C. Tucker, M. Brick, E. De Leeuw, L. Japec, P. Lavrakas, M. Link, & R. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 250–275). New York: Wiley. (Chapter 12).

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research, 43*, 441–452.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or Internet-based surveys. *Educational and Psychological Measurement, 60*, 821–826.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69*(1), 87–98.

De Leeuw, E. D. (1992). *Data quality in mail, telephone, and face-to-face surveys*. Amsterdam: TT Publications.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233–255.

De Leeuw, E. D., & Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41–54). New York: Wiley.

De Leeuw, E.D., & Van Der Zouwen, J. (1988). Data quality in telephone and face-to-face surveys: A comparative analysis. In R.M. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nichols III, & J. Waksberg (Eds.), *Telephone survey methodology*. New York: Wiley.

Dey, E. L. (1997). Working with low survey response rates: The efficacy of weighing adjustments. *Research in Higher Education, 38*, 215–227.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.

Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225–249.

Dillman, D. A. (1999). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dillman, D. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dillman, D. A., & Christian, L. M. (2003). Survey mode as a source of instability in responses across surveys. *Field Methods, 15*(2), 1–22.

Dillman, D. A., Clark, J. A., & Sinclair, M. A. (1995). How prenotice letters, stamped return envelopes, and reminder postcards affect mailback response rates for census questionnaires? *Survey Methodology, 21*, 1–7.

Dillman, D. A., & Mason, R. G. (1984). *The influence of survey method on question response*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research. Delavan, WI.

Dillman, D. A., Phelps, G., Tortora, R. D., Swift, K., Kohrell, J., & Berck, J. (2009). Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response, and the Internet. *Social Science Research, 38*, 1–18.

Doyle, J. (n.d.). 9 February, 2012, http://www.wpi.edu/Images/CMS/SSPS/Doyle_-_Face-to-Face_Surveys.pdf.

Ethier, R. G., Poe, G.L., Schulze, W.D., & Clark, J. (2000). A comparison of hypothetical phone and mail contingent valuation responses for green-pricing electricity programs. *Land Economics 76*(1), 54–67.

Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (2nd ed.). New York: McGraw Hill.

Fowler, F. J, Jr, Gallagher, P. M., Stringfellow, V. L., Zalavsky, A. M., Thompson, J. W., & Cleary, P. D. (2002). Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Medical Care, 40*, 190–200.

Groves, R. (1989). *Survey errors and survey costs* (1st ed.). New York: Wiley-Interscience.

Groves, R. M., Cialdini, R. B., & Courier, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475–495.

Handwerk, P. G., Carson, C., & Blackwell, K. M. (2000). Online vs. paper-and-pencil surveying of students: A case study. Paper presented at the 40th Annual Forum of the Association for Institutional Research, Cincinnati, OH, May 21–24, 2000.

Hare, S., Price, J. H., Flynn, M. G., & King, K. A. (1998). Increasing return rates of a mail survey to exercise professional using a modest monetary incentive. *Perceptual and Motor Skills, 86*(1), 217–218.

Horrigan, J. B., & Smith, A. (2007). Home broadband adoption, 2007. Pew Internet & American Life Project. Accessed July 10, 2011. Available from: http://www.pewinternet.org/pdfs/PIP_Broadband%202007.pdf.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of Web and mail survey response rates. *Public Opinion Quarterly, 68*, 94–101.

Keisler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*(3), 402–413.

Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58*, 381–399.

Massey, J. T., O'Connor, D., & Krotki, K. (1997). Response rates in random digit dialing RDD telephone surveys. In: *1997 Proceedings of the section on survey research methods. American Statistical Association.* (pp. 707–712). Alexandria, VA.

McNamara, C. (2009). General guidelines for conducting interviews. Retrieved June 8, 2011, from http://managementhelp.org/evaluatn/intrview.htm.

Paolo, A. M., Bonaminio, G. A., Gibson, C., Patridge, T., & Kallail, K. (2000). Response rate comparisons of e-mail and mail distributed student evaluations. *Teaching and Learning in Medicine, 12*(2), 81–84.

Parker, L. (1992). Collecting data in the e-mail way. *Training and Development*, *46*, 52–54.

Sangster, R. L., & Meekins, B. J. (2003). Data concerns for hard to reach and reluctant respondents in telephone panel surveys, *Presented for the 14th International Workshop on Household Survey Nonresponse*, Leuven, Belgium.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44*, 409–432.

Schaefer, D. R., & Dillman, D. A. (1998). Development of a standard e-mail methodology. *Public Opinion Quarterly, 62*, 378–397.

Schmidt, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments and Computers, 29*, 274–279.

Schuman, H. (2002). Sense and nonsense about surveys. *Contexts, 1*, 40–47.

Shank, M. D., Darr, B. D., & Werner, T. C. (1990). Increasing mail survey response rates: Investigating

the perceived value of cash versus non-cash incentives. *Applied Marketing Research, 30*(3), 28–32.

Sheehan, K. B., & Hoy, M. G. (1999). Using e-mail to survey Internet users in the United States: Methodology and assessment. *Journal of Computer Mediated Communication, 4*(3). [Online]. Available: http://www.ascusc.org/jcmc/vol4/issue3/sheehan.html.

Sheehan, B. K., & McMillan, J. S. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising, 39*(4), 45–54.

Sills, S. J., & Song, C. (2002). Innovations in survey research: An application of web surveys. *Social Science Computer Review, 20*, 22–30.

Singer, E. (2006). Introduction: Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 637–645.

Smith, T. (1995). Trends in nonresponse rates. *International Journal of Public Opinion Research, 7*, 157–171.

Smith, C. B. (1997). Casting the net: Surveying an Internet population. *Journal of Communication Mediated by Computers, 4*(3). http://www.ascusc.org/jcmc/vol3/issue1/smith.html.

Solomon, D. J. (2001). Conducting Web-based surveys. *Practical Assessment, Research, and Evaluation*, *7*(19), 1–5.

Steeh, C. (1981). Trends in nonresponse rates. *Public Opinion Quarterly, 45*, 40–57.

Swoboda, S. J., Muehlberger, N., Weitkunat, R., & Schneeweiss, S. (1997). Internet surveys by direct mailing: An innovative way of collecting data. *Social Science Computer Review, 15*(3), 242–255.

Tarnai, J., & Dillman, D. A. (1992). Questionnaire context as a source of response differences in mail versus telephone surveys. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research*. New York: Springer.

Tse, A. (1998). Comparing the response rate, response speed and response quality of two methods of sending questionnaires: E-mail vs. mail. *Journal of the Market Research Society, 40*(4), 353–361.

Weible, R., & Wallace, J. (1998). The impact of the Internet on data collection. *Marketing Research, 10*(3), 19–23.

Wyatt, J. C. (2000). When to use web-based surveys. *Journal of the American Medical Informatics Association, 7*(4), 426–429.

Yun, G. W., & Trumbo C. W., (2000). Comparative response to a survey executed by post, e-mail, and web form. Available at http://jcmc.indiana.edu/vol6/issue1/yun.html.

Zatz, D. (2000). Create effective e-mail surveys. *HR Magazine, 45*(1), 97–103.

Part V

Sensitive and Difficult Survey Topics

22 Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items

Susan McNeeley

There is general agreement that certain survey topics pose problems for researchers due to the reluctance of respondents to discuss those topics. Social scientists have become increasingly interested in such topics due to the call for increased research in areas related to victimization, sexually transmitted diseases, and drug and alcohol use. Measurement error can occur when surveying on these sensitive topics since respondents may choose not to participate in the survey, not to answer particular questions, or not to give accurate answers to those items. Survey methodologists have identified several methods of reducing the threat associated with sensitive survey items; these methods include changes in mode of administration, question wording, and interviewer training. Such methods are essential in the attempt to reduce the total survey error (TSE), by minimizing response and nonresponse biases.

The first section of this chapter will identify and describe common sensitive topics, as well as populations that may view certain questions as especially threatening, in order to aid survey designers in determining whether sensitivity will be an issue for their survey. The second section will explain in more depth the problems posed by sensitive issues in surveys. The third section will offer methods to reduce the threat associated with sensitive issues: choosing the appropriate mode of administration, wording questions appropriately, training interviewers to ease participants' discomfort, and additional techniques that lessen the anxiety caused by sensitive issues.

22.1 Sensitive Topics

According to Tourangeau and Smith (1996, p. 276), "a question is sensitive if it raises concerns about disapproval or other consequences (such as legal sanctions) for reporting truthfully or if the question itself is seen as an invasion of privacy." More specifically, Barnett (1998) identified several ways of identifying topics that are sensitive. The first definition of "sensitivity" is based on the perceived costs of answering the question. If respondents are expected to worry about repercussions associated with answering, then the question is considered sensitive or threatening. A second way that researchers have determined if questions are sensitive is to examine whether anonymity makes a difference. If respondents whose identities are known refuse to answer the question or answer it inaccurately, then the item is considered sensitive. It is assumed that respondents would only lie about behaviors that are sensitive or threatening. Therefore, many topics are

S. McNeeley (✉)
School of Criminal Justice, University of Cincinnati,
665 Dyer Hall, PO Box 210389, Cincinnati,
OH 45221-0389, USA
e-mail: mcneelsm@mail.uc.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_22, © Springer Science+Business Media New York 2012

considered sensitive or threatening because studies have shown that respondents tend to answer inaccurately. Third, if respondents confirm that the subject makes them feel uneasy (or they feel that it would make others uncomfortable), then the subject is considered sensitive.

There has been little effort to empirically examine the sensitive nature of the subjects that researchers perceive to be threatening. Many sensitive topics are identified using "common sense;" certain subjects are considered threatening or sensitive due simply to the researcher's gut feeling regarding these topics (Barnett 1998). It is not unusual for researchers to determine that topics or questions are sensitive after the survey has been conducted (Sudman & Bradburn 1982).

However, some scholars have attempted to empirically determine the degree of sensitivity of a particular topic. Surveys asking respondents how sensitive they considered the questions to be have found that certain topics are generally considered sensitive. Blair et al. (1977) asked respondents whether they felt that certain topics would make people feel uneasy; a large percentage of participants rated topics related to sexual behavior and drug or alcohol use as topics that would make people feel very uneasy.

Research on sensitivity generally focuses on behavior rather than attitudes (Barnett 1998). This focus is likely due to the ability to validate some reports of behavior; attitudes cannot be verified against other types of data and are therefore more difficult to measure. This trend may be changing, as socially undesirable attitudes are beginning to be examined by scholars. Recent studies compare results from the Implicit Association Test (IAT) to answers to attitudinal questions to determine if respondents answer these questions dishonestly (Greenwald et al. 2009; Slabbinck & Van Kenhove 2010).

Box 22.1: Implicit association test

The Implicit Association Test, or IAT, was proposed by Greenwald & Banaji (1995) and introduced to social scientists by Greenwald et al. (1998). It is meant to link implicit and explicit memory to determine individuals' associations between concepts.

The IAT is a computer-based test. Users are asked to rapidly associate concepts; the time that the user takes to make these associations are measured to determine the user's ease of association. Generally, the test begins with two categories showing in the left and right areas of the computer screen; users are asked to associate words that appear in the middle with a certain category. The position of the categories and the associations are switched in order to measure the ease with which the user makes each association.

The most well-known IAT is used to measure the extent to which a person believes in stereotypes. There are a number of stereotype IATs, which test a wide range of ethnic, gender-based, and religious stereotypes. In addition to stereotype IATs, there are IATs that measure self-esteem and preferences for certain racial or ethnic groups.

There has been some criticism of the IAT; however, a great deal of research has been done to determine its validity. A recent meta-analysis determined that the IAT is an effective predictor of behaviors and attitudes. In fact, the IAT outperforms self-report measures in some studies (Greenwald et al. 2009).

For more information on the IAT, or to take the test, visit www.implicit.harvard.edu.

While there is not a clear standard definition of sensitive issues, there is general consensus that particular topics are sensitive. One of those topics is drug or alcohol use. Because of the illegal nature of some type of drug use, respondents could fear legal penalties if their answers became known. Furthermore, substance use, including drinking alcohol and smoking cigarettes, is considered deviant, and therefore is likely to cause embarrassment. Research has shown that certain drug or alcohol use is likely to be underreported due to the social undesirability associated with such behavior. Light or experimental users are less likely than habitual users to admit substance use (Mensch & Kandel 1988). For example, respondents tend to underreport drug use; this tendency is especially notable with use of "hard" drugs like cocaine, since other drugs are more socially acceptable (Sloan et al. 2004).

Scholars have become increasingly interested in conducting research on sexual behavior, especially risky sexual activity that is related to the transmission of STIs. The private nature of sexuality, as well as the empirically noted tendency of respondents to answer questions about sexual behavior inaccurately, has led to the inclusion of sexual behavior as a sensitive topic. It is expected that reports of sexual behavior would be equal across genders; however, these rates differ, suggesting that either men or women, or both, report inaccurately (Poulin 2010; Smith 1992). In face-to-face interviews, girls are unwilling to report sexual behavior, likely due to the embarrassing and socially undesirable nature of the topic (Mensch et al. 2003, 2008). On the other hand, boys often exaggerate sexual behavior in face-to-face interviews as a way to impress or shock the interviewer (Mensch et al. 2003).

Underreporting of homosexual behavior is an even more serious problem for researchers, especially when surveying juveniles. Juveniles report a lower rate of homosexual behavior than adults retroactively report engaging in as teenagers (Turner et al. 1996, 1998). Young males are less likely to report male–male sexual behavior than heterosexual behavior, suggesting that, at least for males, homosexual behavior is a more sensitive topic than is heterosexual behavior (Turner et al. 1998).

Private family issues are also considered sensitive topics, as it assumed that respondents are either unwilling or consider themselves unable to discuss these issues with interviewers. Studies show that relationship violence is underreported; both male and female partners tend to underreport marital violence (Szinovacz & Egley 1995). Furthermore, women in abusive relationships are less likely to participate in surveys, since such participation violates the isolation inherent in abusive relationships (Records & Rice 2006). Another family issue that is considered sensitive is loss or death in the family, as participants (and sometimes interviewers) can become emotionally distressed while discussing the topic (Kitson et al. 1996).

Victimology research relies heavily on survey data, as estimates obtained by official data are believed to be invalid due to the tendency of victims to opt out of notifying the police. Experiences of victimization are considered a highly sensitive topic, as these experiences are often traumatic and can be embarrassing (see Kleck and Roberts Chap. 24, and Yang and Hinkle Chap 25). Researchers have noted a tendency of survey participants to underreport experiences of victimization, especially sexual victimization (Langhinrichsen-Rohling et al. 2006; Krebs et al. 2011; Records & Rice 2006; Smith 1994). Misreporting of victimization experiences can be attributed to at least two factors—the sensitive nature of such victimizations and individuals' inability to define the incidents as criminal.

It has been suggested that survey respondents are unwilling to report criminal or delinquent activity (Tracy & Fox 1981). The sensitive nature of criminal behavior is clear, since participants would be admitting to behavior that could result in legal sanctions. Furthermore, since criminal behavior is considered morally wrong, responses to such questions could be damaging to one's reputation.

Questions regarding one's health are likely to be considered sensitive, as the respondent may view them as a violation of privacy. Therefore, researchers interested in medical conditions, suicide, or abortion should attempt to reduce the threat associated with such questions. Several medical conditions are underreported in surveys, while screening behaviors are overreported (Martin et al. 2000). Suicidal thoughts and attempts are considered sensitive topics, both because they are likely to be underreported and because discussing suicide may cause distress in respondents (Langhinrichsen-Rohling et al. 2006). Questions regarding experiences of abortion are likely to be considered sensitive for similar reasons (Fu et al. 1998). An experiment comparing survey responses to medical records demonstrated that 83.4% of women who had had abortions accurately responded to the question (Rasinski et al. 1999). However, surveys on medical topics that are conducted in clinical settings may not be considered as sensitive as those conducted in other settings, as respondents are used to discussing personal matters in that setting (Millstein 1987).

Questions on demographic information are considered sensitive because respondents may perceive these questions as identifying (Giles & Feild 1978). Individuals who complete surveys may worry that their unique responses to demographic questions could allow researchers to identify them, especially if they are part of a known sample, such as a survey conducted within one's workplace.

Some issues are considered sensitive because they are socially desirable. Individuals may be reluctant to admit that they do not engage in activities that are widely encouraged in society. One such behavior is voting. Voting is an especially popular topic when studying sensitive issues or social desirability due to the ability to use official voting records to verify self-reports (Silver et al. 1986; Tourangeau et al. 2010). Other issues that are sensitive due to social desirability are hygienic practices, diet, and exercise; many people are unwilling to report a lack of cleanliness or poor diet and exercise use due to the embarrassment that could result (Hebert et al. 1997; Motl et al. 2005; Visser et al. 1989).

22.2 Sensitive Populations

Cultural characteristics of the group being surveyed must be considered when attempting to conduct surveys on sensitive topics, as the sensitivity of the topic can vary depending upon the target population (Barnett 1998). Different cultural groups have different norms and preferences. These differences create a context in which a topic that is normally benign can be viewed as threatening. In contrast, certain groups can perceive a behavior that is usually socially undesirable as acceptable or even desirable.

As an example of the effect of cultural differences on surveys of sensitive behavior, researchers have found widely varying levels of sexual behavior across Sub-Saharan African countries. (Mensch et al. 2003). These differences are possibly due to cultural differences in willingness to report such behavior. For example, in Malawi, pre- or extra-martial sexual behavior has become morally reprehensible due to the AIDS epidemic (Poulin 2010). Therefore, surveys about sexual activity conducted there may be more vulnerable to misreporting due to the social undesirability of such behavior than surveys in other countries.

The cultural views of the target population, or segments of the population, can also affect the interview process. Certain groups, such as lower socioeconomic and minority groups, are more suspicious of the interview process (Mensch et al. 2008). More specifically, group membership and culture can affect whether the respondent believes the researcher's assurances of anonymity (Barnett 1998). Therefore, anonymity or confidentiality assurances will be a more effective method of dealing with sensitive topics for some groups than others. Furthermore, the culture of the population of interest should be considered when choosing a mode of administration, since the use of technological devices such as computers or tape recorders can make

some groups of respondents suspicious (Mensch et al. 2003; Poulin 2010).

Researchers should be especially careful when conducting surveys of children or adolescents, as sensitive topics pose additional ethical issues for this population. It is important to note that children or teenagers are often surveyed in the presence of others; surveys often take place in school in the presence of teachers and classmates or at home in the presence of family members (Phillips 1994). In these settings, juveniles may be very worried that their answers will be seen or overheard. Furthermore, they may feel that they have to participate in the study, as they may face ridicule if they refuse. Scholars attempting to survey these populations should take extra steps to ensure the privacy of respondents.

It has been suggested that teenagers and children are more emotionally affected by survey questions than are adults (Langhinrichsen-Rohling et al. 2006). Certain topics may be especially embarrassing for young participants (Harden et al. 2000). Children may also have increased difficulty in completing surveys due to lack of understanding of the questions (Jacobs et al. 2008). Also refer to Chap. 24 by Kleck and Roberts for further discussion on issues of privacy and sense of embarrassment.

22.3 Problems with Sensitive Topics

There are two main methodological issues associated with surveys covering sensitive topics. The first issue is *nonresponse*: potential respondents may refuse to participate in the survey when they see what topics are covered or may refuse to answer certain questions. The second issue is *misreporting*: a respondent who agrees to participate may not answer sensitive questions honestly, creating measurement error. In addition to these methodological issues, there are *ethical considerations* when asking threatening questions in surveys.

22.3.1 Nonresponse

A major issue associated with sensitive topics is refusal to participate. Since researchers are ethically obligated to explain the purpose of the survey and the topics that will be covered, respondents have the ability to opt out of surveys that cover issues that they consider threatening or invasive. This refusal to participate causes two main problems for the validity of the results. First, surveys that experience high refusal rates, especially surveys of sensitive topics, are unlikely to fully cover the population (Johnston & O'Malley 1985). It is possible that certain segments of the population, such as minority ethnic or religious groups, are underrepresented in surveys that contain questions on sensitive issues. Therefore, estimates obtained from such a survey may not be generalizable to the entire population of interest. Second, individuals whose information the researchers are trying to obtain may be the most likely to refuse to participate in the survey (O'Brien et al. 2006). For example, individuals who participate in risky AIDS-related behaviors may refuse to participate in a survey of such behavior. Therefore, estimates of the extent to which individuals participate in that behavior may be low, and important information that could be obtained through the survey may not be included.

Nonresponse is conditional on a number of factors, including an individual's general dislike for surveys; characteristics of the respondent, such as sociodemographic characteristics and household makeup; characteristics of the interviewer, such as sociodemographic characteristics, prior experience, and demeanor; and factors related to the survey design, such as mode of contact, mode of administration, and length of time to complete the survey (Groves 1989, 1992; O'Brien et al. 2006). In addition, survey participation may depend upon the topic(s) that the survey covers and the sensitive nature of those topics (Groves et al. 1992; Steeh 1981). Recent research indicates that surveys containing

embarrassing or socially undesirable topics are likely to have low response rates (O'Brien et al. 2006; Tourangeau et al. 2010).

Some individuals may refuse to participate in certain surveys because of their perception of the potential repercussions associated with answering. Research has identified several populations as less likely to participate due to this perception. First, arrestees may fear that their answers will be used against them (National Institute of Justice 2003). Second, studies examining alcohol and drug use among welfare recipients note the possibility that these participants may feel that their answers will get back to the government, especially if the study is funded by a grant (Pridemore et al. 2005). Finally, members of dysfunctional or abusive families may be unwilling to participate in surveys. For women in abusive relationships, either participating in a study or answering questions about their relationship may be considered dangerous, as their partners may retaliate with violence (Records & Rice 2006)

In addition to refusing to participate, respondents can refuse to answer specific questions. Item nonresponse can result in high levels of missing data for specific questions. Missing data affects the generalizability of estimates obtained from surveys, as it is unknown whether those individuals who chose not to answer are different from those who did answer. Furthermore, as explained above, statistics based on small sample sizes suffer more from sampling error. Therefore, survey designers must attempt to decrease item nonresponse.

Threatening survey questions are likely to suffer from nonresponse (Mangione et al. 1982; Mensch et al. 2008; Sakshaug et al. 2010). Nonresponse may be more problematic in reports of socially desirable behaviors than in reports of socially undesirable items (Sakshaug et al. 2010). Respondents may simply choose not to respond to non-threatening questions that they would prefer not to answer honestly.

22.3.2 Misreporting

Another major issue associated with surveys of sensitive topics is misreporting. Respondents who agree to participate in the survey may still feel unable or unwilling to answer honestly. This reluctance to give honest answers can stem from different sources. Respondents may be uncomfortable refusing to participate or respond, since such a refusal could be seen as an indication that their answers would be socially undesirable (Phillips 1994). Respondents may choose to answer dishonestly due to a desire to present themselves in the best light to the interviewer or to avoid potential repercussions (Pridemore et al. 2005; Szinovacz & Egley 1995). Respondents sometimes give inaccurate responses due to their unwillingness to think about embarrassing or traumatic events or inability to understand what has happened to them (Tourangeau & McNeeley 2000). Misreporting is problematic because it can result in large biases in estimates obtained from sensitive survey estimates. Rates of socially undesirable behavior can be underestimated by as much as 20% (Sakshaug et al. 2010)

Respondents purposely misreport for several reasons; to avoid embarrassment or stigmatization, to avoid potential repercussions, and to present themselves in a positive manner (Pridemore et al. 2005). There is a tendency to overreport socially desirable behavior, characteristics, or attitudes. For example, voting is considered highly desirable; therefore, individuals tend to report voting in elections even if they did not do so (Silver et al. 1986). In addition to overreporting socially desirable traits, respondents are likely to underreport socially undesirable behavior, characteristics, or attitudes. For example, drug use is likely to be underreported; the extent to which substance use is underreported varies by the level of social undesirability of the particular substance in question (Sloan et al. 2004).

It is generally assumed that misreporting will occur in a specific direction (as over- or underreporting, depending on whether the item is socially desirable or undesirable). However, it is possible that measurement error associated with reports of sensitive behavior occurs in both directions, regardless of the nature of the behavior in question. A meta-analysis found that errors were balanced on either side of the

average rather than occurring in the same direction, as would be expected if the respondent answered dishonestly due to the social desirability (or undesirability) of the behavior (Marquis 1984). Therefore, it is important that researchers do not assume that higher estimates of socially undesirable behavior are necessarily more accurate (Barnett 1998; Rand & Rennison 2005). Unfortunately, a great deal of research on improving accuracy in sensitive survey items makes that assumption.

22.3.3 Link Between Nonresponse and Misreporting

Tourangeau et al. (2010) found support for a relationship between nonresponse and misreporting, with sensitive surveys resulting in low response rates and inaccurate responses. This link between nonresponse and misreporting has not been replicated in other studies (Sakshaug et al. 2010). Therefore, more research must be conducted before it is clear whether such a relationship exists.

There is often a tradeoff between nonresponse and misreporting associated with choosing a mode of administration (Sakshaug et al. 2010). As will be discussed below, the various modes of administration have different effects on nonresponse and misreporting. Therefore, when choosing a mode of administration for a survey that includes sensitive items, one must take into account the specific nature of the topic in order to determine which source of error will be most problematic.

22.3.4 Ethical Considerations

Researchers must consider the emotional effect that their surveys will have on participants. Survey designers should attempt to phrase questions in ways that will ease respondents' discomfort. This consideration is especially important when conducting surveys in which respondents will be asked to recall and discuss traumatic past or ongoing events. Furthermore, surveys on certain subjects, such as family violence, can create dangerous situations for respondents if their family members become aware of their participation or responses (Lynn & Kaminska 2011; O'Brien et al. 2006; Records & Rice 2006). Therefore, every effort must be made to decrease these potential harms.

The emotional effects of survey questions are especially salient for children and teenagers (Langhinrichsen-Rohling et al. 2006; Phillips 1994). The results of a study conducted by Langhinrichsen-Rohling et al. (2006) suggest that sensitive survey items have a negative emotional effect on adolescents. A small percentage (between 2.5 and 7.6%) of participants reported feeling upset frequently while completing a survey on suicidal thoughts and attempts, illicit drug use, and experiences of victimization. Those who had experienced or engaged in the threatening behaviors or experiences were more likely to feel upset while participating in the survey. Demographic characteristics and mode of data collection were not significantly related to participants' emotional state.

22.4 Privacy, Anonymity, and Confidentiality

Survey techniques that guarantee anonymity are expected to result in more accurate reporting of sensitive issues. Respondents feel more comfortable admitting to embarrassing, stigmatizing, or illegal behavior when participating in more confidential interview processes, such as computer-based surveys (Mensch et al. 2003, 2008; Tourangeau & Smith 1996)

Box 22.2: Increasing privacy in face-to-face interviews

A useful way to improve accuracy in face-to-face interviews is to utilize the sealed booklet method (see Makkai & McAllister 1992). During an interview, respondents are handed a sealed booklet and told that the interviewer does not have a copy of the questions or answers. The numbers of the answers are also scrambled

to assure the respondent of the anonymous nature of the questions. The respondent simply calls out the number that corresponds to the answer, and the interviewer marks the answer. The use of the sealed booklet results in more accurate estimates of sensitive behavior; respondents feel more comfortable answering honestly due to the increase in privacy. It also allows the researcher to ask a greater number of sensitive questions.

A similar technique to the sealed booklet method is the use of a portable audiocassette tape player with headphones (see Horm et al. 1996). Devices such as these minimize the chance of a third party overhearing the respondent's answers by allowing the respondent to privately listen to the survey questions and answer categories. They also reduce the potential for embarrassment, since the respondent does not have to directly respond to the interviewer. This method is beneficial because it increases privacy without requiring a high degree of literacy and is inexpensive when compared to methods such as ACASI.

Research suggests that the presence of family members or others during interviews decreases truthful answering to sensitive questions (Rasinski et al. 1999; Smith 1995). However, while there seems to be a tendency to underreport embarrassing or undesirable behavior in the presence of others, it does not appear that such a tendency exists for questions regarding socially desirable behavior. For example, the presence of third parties does not affect misreporting of voting behavior (Silver et al. 1986).

While more anonymous or confidential interviews seem to be helpful in increasing accuracy of responses, it seems that assurances of anonymity or confidentiality are not as beneficial. Excessive assurances of confidentiality may give respondents the idea that the questions are more sensitive than they actually are, decreasing their willingness to participate or give honest answers (Rasinski et al. 1999). Some studies find anonymity assurances to be related to higher levels of nonresponse and misreporting; anonymity can lower accountability, making respondents feel that they do not need to respond to certain questions or that it does not matter if their responses are accurate (Fuller 1974; McDaniel & Rao 1981)

This inability of confidentiality assurances to improve accuracy is likely due to the fact that they typically do not address all of the concerns that participants may have about answering truthfully. An experiment conducted by Rasinski et al. (1999) found that respondents consider the chance of an unauthorized group gaining access to survey data to be low. However, they rate the chance of other negative consequences to be higher. Participants may worry that a third party (such as a family member) will overhear their answers during the interviewer, discover the respondent's secrets, and/or become upset with the respondent. Furthermore, participants worry that the interviewer will disapprove, causing the respondent to experience embarrassment. For survey conditions (including mode of interview and privacy) that respondents consider high risk, they consider the likelihood of truthful reporting to be low (Rasinski et al. 1999).

22.5 Mode of Administration

When surveying on sensitive topics, a researcher should consider the effect of the mode of administration on respondents' willingness to answer honestly. The choice of mode of administration is often dependent on other factors, such as the researcher's budget and time available in which to administer the survey. However, when possible, this decision should take into account the topic of the survey, as research suggests that the accuracy of estimates of sensitive behavior and attitudes varies by mode of administration. This section will discuss various modes of administration

and their effect on accurate reporting of sensitive behaviors.

22.5.1 Face-to-Face

Although one might expect lower rates of socially undesirable behavior in surveys conducted via face-to-face interviews, this is not always the case. Some studies obtain more accurate estimates when conducting surveys in person than when using other modes of administration. A study by Mangione et al. (1982) found that respondents reported higher rates of drinking behavior in face-to-face interviews than in telephone interviews or self-administered questionnaires. A study comparing abortion reports from personal interviews to those from self-reports obtained higher rates of abortion in self-administered questionnaires than in face-to-face interviews; however, some respondents who reported having abortions in the interviews did not report these incidents in the self-administered portion (Fu et al. 1998).

More specifically, it is possible to determine when face-to-face interviews are more appropriate for surveys of sensitive topics. It seems that in-person interviews are more suitable for extremely sensitive items. For example, a study examining the drug and alcohol use of welfare recipients obtained similar lifetime estimates of drug and alcohol use in telephone and face-to-face surveys. However, respondents were more likely to admit to recent drug and alcohol use in face-to-face interviews than in telephone surveys (Pridemore et al. 2005). Participants may have seen questions on more recent substance use as especially threatening due to the fear of negative consequences of answering those items truthfully.

The tendency of respondents to answer honestly during face-to-face interviews is not surprising considering the conversational nature of such interviews (Barnett 1998; Beck et al. 2002). Because surveys are social interactions between the interviewer and the participant, rapport with the interviewer can put respondents at ease and neutralize the threatening aspect of the topic (Poulin 2010; Smith 1994). In fact, in a flexible interview in which the participant feels a connection with the interviewer, he or she may joke or tell stories, even when discussing extremely sensitive topics (Poulin 2010). When respondents feel more comfortable with the interviewer, they are more likely to give honest answers (Poulin 2010). If the participant views the interviewer as an understanding listener, discussing sensitive information can actually be relieving or therapeutic (Rasinski et al. 1999). The flexible nature of interviews also grants the interviewer the ability to probe or ask about contradictory answers (Poulin 2010).

22.5.2 Telephone

Surveys administered over the telephone have become quite popular due to several procedural advantages over face-to-face interviews. Telephone interviews are less expensive than face-to-face interviews, easier to administer, and more convenient for both the data collectors and the respondent (Mangione et al. 1982; Pridemore et al. 2005). Furthermore, this mode of administration has the added benefit of quality control through monitoring and recording of interviewers (Pridemore et al. 2005). Respondents find surveys conducted over the telephone to be as enjoyable as in-person interviews (Mangione et al. 1982).

In many cases, telephone surveys result in similar estimates as those obtained from face-to-face interviews, although the latter may be more reliable for extremely sensitive items (Mangione et al. 1982; Pridemore et al. 2005). However, in some circumstances, telephone surveys are superior; the potential for embarrassment is lower since the interviewer and the respondent do not physically meet.

Box 22.3: T-ACASI

Telephone surveys can be further modified to increase anonymity. Modifications of telephone surveys can be used to mimic the beneficial aspects of audio computer-assisted self-interviewing (ACASI) while

maintaining the convenience and budgetary benefits associated with telephone surveys. These modifications are commonly referred to as telephone ACASI, or T-ACASI. Using touch-tone data entry, a recorded voice will read the questions and answers to the respondent, who will enter the answers using the phone's keypad. Similarly, interactive voice response (IVR) can be used; in this technique, the computer is programed to comprehend the participants' verbal responses. These surveys have been demonstrated to be useful in collective sensitive information (Blumberg et al. 2003). As with other computer-based methods, this mode of administration can suffer from its impersonality; respondents may be unwilling to participate in such surveys.

The increase in cell phones has resulted in higher quality data obtained from telephone surveys. At first glance, it may seem that those using cell phones would be more likely to take surveys in public in the presence of others who could overhear their answers. However, research suggests that respondents are less likely to give socially desirable answers when using mobile phones than when using fixed phone lines (Lynn & Kaminska 2011). Participants using cell phones have greater control over the presence of others within earshot; if someone is nearby, the participant has less restricted movement to avoid allowing that person to eavesdrop. Furthermore, cell phone survey respondents do not have to worry that someone in their household is listening on another line, thereby reducing the possibility that family members will discover their answers. Therefore, the use of cell phones results in greater privacy, encouraging participation in the survey and honest answers to threatening or embarrassing questions.

22.5.3 Self-Administered

Because the problems posed by sensitive survey items are considered to be related to the respondent's discomfort or embarrassment, it is commonly assumed that more reliable estimates can be obtained through survey methods that decrease the respondent's contact with others. Therefore, it is commonly believed that the use of self-administered surveys will suffer less from measurement error.

When using self-administered surveys, participants report higher levels of socially stigmatizing behaviors that are commonly believed to be underreported in face-to-face interviews (Aquilino 1994; Beach et al. 2010; Beck et al. 2002; Fu et al. 1998; Mensch et al. 2003; Turner et al. 1996). Since there is no interviewer to impress, respondents are also less likely to overestimate socially desirable behaviors and other shocking behaviors, such as sexual exploits, when using self-administered survey methods (Mensch et al. 2003). The usefulness of self-administered surveys has been confirmed by studies conducted outside of North America—Europe (Beck et al. 2002; Gmel 2000), South America (Mensch et al. 2008), and Africa (Mensch et al. 2003; Poulin 2010). Self-administered surveys can be conducted in a number of ways: written self-administered questionnaires (SAQ), which can be conducted through the mail or in the presence of the data collector; web-based surveys; and computer-based surveys taken in the presence of the data collector. However, other studies have found that face-to-face or telephone interviews result in more accurate estimates, possibly because self-administered surveys are less personal (Pridemore et al. 2005).

Surveys administered by mail are considered beneficial because they allow respondents to answer the questions privately and at their convenience. However, mail surveys have major disadvantages, depending upon the subject of the questions. Surveys regarding very sensitive topics should not be conducted through the mail because of the possibility that others in the household or the community may intercept and read the questionnaire. This possibility is especially problematic when conducting surveys on family or relationship issues, such as domestic abuse.

Records and Rice (2006) propose using a modified version of Dillman's method for surveys on sensitive topics; specifically spousal and domestic abuse. Dillman (1978) suggests four mailings: the original survey distribution, a postcard reminder, and two more distributions of the survey. Each of the three survey mailings should include a different cover letter that emphasizes the importance of the project and acknowledges the effort put forth by the respondent. The major modification suggested by Records & Rice (2006) for sensitive topics was the use of a letter for the second mailing instead of a postcard, because a postcard could be more easily intercepted and read by others in the household. Other modifications were made, including the use of a standard cover letter for each mailing, to decrease the cost and effort put forth by the agencies assisting in the data collection. The modified version of the Dillman method for sensitive issues increased the recruitment rate from 13 to 35% (Records & Rice 2006). However, it could not be determined whether this method increased the number of abuse victims in the sample.

As an alternative to mail surveys, researchers may consider a drop-off/pick-up method in which the interviewer delivers the survey and sets a time to pick it up. This face-to-face contact allows the respondent to establish a rapport with an interviewer but maintain privacy while completing the survey (Mangione et al. 1982). When using this method, the interviewer can assist those that are not capable of taking the survey on their own. While this method of conducting interviews is convenient, it does not increase accuracy of estimates over telephone or face-to-face interviews, and respondents may find it less enjoyable (Mangione et al. 1982). This method is also more costly than mail surveys in terms of time commitment.

Computer-based survey methods are becoming popular due to problems associated with other modes of administration. One such computerized method is computer-assisted self-interviewing, or CASI. CASI is similar to an SAQ in that participants are able to read the questions on the screen and respond. Another recent trend in social science research is the audio computer-assisted self-interview (ACASI) system. ACASI is a computer-based interview system that includes audible instructions, questions, and answers. The questions and instructions are usually displayed on the screen and read by the computer at the same time; respondents use the keyboard or mouse to progress through the survey (NIMH 2008).

There are several advantages associated with the use of computer-based surveys. First, these surveys may be more enjoyable for participants than other modes of administration (Millstein 1987; NIMH 2008). This is especially true for children; young respondents tend to prefer computer-based methods because they are "fun" or "interesting" (Jacobs et al. 2008). Second, a slight majority of participants report having a preference for computer-based surveys (Millstein 1987; NIMH 2008). Third, some participants who have little computer experience report feeling empowered by the opportunity to use a computer (NIMH 2008). Fourth and more specifically, ACASI is especially beneficial since it does not require any degree of literacy to be used. Studies have found that individuals with less education report sensitive behavior more accurately using ACASI than paper SAQ (Turner et al. 1998).

Computer-assisted self-interviewing methods (CASI and ACASI) produce more accurate reports of embarrassing information (sexual behavior and drug use) than do personal interviews (Gribble et al. 1999; Tourangeau & Smith 1996; Turner et al. 1996). ACASI also results in more reliable reporting of sensitive behaviors over time (Mensch et al. 2008). The use of ACASI also improves estimates of sensitive behavior over written self-administered questionnaires (Beach et al. 2010; Gribble et al. 1999; Mensch et al. 2003; Rhodes et al. 2002; Tourangeau & Smith 1996; Turner et al. 1996, 1998). Surveys on sexual behavior (including male homosexual behavior and contraceptive use), intimate partner violence, and drug and alcohol use have acquired higher estimates using ACASI than those using paper SAQs. Nonresponse is also less of a problem when using

ACASI than when using paper SAQ; respondents being surveyed via ACASI are more likely to use the "Don't know" or "Refuse to answer" options than are those filling out standard questionnaires (Turner et al. 1998).

While self-interviewing may result in less misreporting due to the increased privacy for respondents, there are also disadvantages associated with the use of self-administered survey methods. A major disadvantage of self-administered surveys is the absence of an interviewer. In other modes of administration, respondents who are confused by a question can ask for clarification. Self-administered modes of survey administration are limited in their ability to aid respondents in understanding questions. Therefore, results obtained from self-administered surveys may reflect measurement error due to participants' confusion. Furthermore, in telephone and face-to-face surveys, interviewers can identify answers that are inconsistent and ask for clarification (Mensch et al. 2008). Self-administered surveys suffer from an inability to check and/or correct for inconsistency in participants' answers. In addition, interviewers establish a rapport with respondents that eases their discomfort about answering honestly; they can also encourage hesitant participants to answer questions by reminding them of the confidential nature of the survey (Mensch et al. 2008). Without this encouragement, self-administered surveys may suffer more from missing responses (Turner et al. 1998).

When considering the use of paper-and-pencil SAQs, there are issues that a researcher needs to keep in mind. First, respondents may still be suspicious about the anonymity of their information, since their identification number and answers are located on the same form (Turner et al. 1998). This suspicion may discourage respondents from answering honestly.

A second disadvantage of written surveys is that they require the respondent to be able to read and write. Because of this issue, the use of paper-and-pencil surveys could bias estimates, as individuals who do not read or write well may be unwilling or unable to participate in the survey. An interviewer may assist individuals who are not literate; however, these interviews would not benefit from increased privacy or anonymity. In addition to a simple requirement of literacy, SAQs can be difficult for some literate respondents. Complicated surveys, such as those with skip patterns, are confusing even for respondents who can read and write (Turner et al. 1998). Therefore, written SAQs are not ideal for use with young children or other populations that are not well educated.

Computer-based survey methods have additional weaknesses as well. First, an understanding of computers and confidence in one's ability to use a computer is an important factor in self-reporting using computer-based surveys. The problem of computer illiteracy is especially salient when surveying certain groups (such as the elderly) or conducting surveys in areas (such as developing countries) in which computer use is less common (Beach et al. 2010; Mensch et al. 2003).

A second disadvantage is that the use of computers may increase suspicion about the survey process. Again, this disadvantage of computerized surveys is especially important in developing countries or low-income areas. The community may be apprehensive about the idea of computers being used to collect information about members of the community (Mensch et al. 2003). This apprehension may reduce response rates as well as discourage those who do participate in the survey from answering accurately.

Third, computer-based surveys (as well as other self-report modes of administration) of children and other at-risk populations may produce biased results since they have difficulty in understanding the questions asked (Jacobs et al. 2008). Because of this weakness, the use of ACASI under some circumstances can result in biased estimates of behavior or incidents. A study in which child endangerment reports generated from surveys using ACASI were investigated found that most of the reports were false (Jacobs et al. 2008). Sixteen percent of the sample reported incidents but only 3.6% reported incidents that were confirmed.

Fourth, computer-based survey methods can be expensive, as costly equipment must be

purchased and maintained. In addition to the initial cost of obtaining and programing equipment for use in computer-based survey methods, there is additional cost associated with the maintenance and replacement of computers if they malfunction or are stolen while surveys are being administered (NIMH 2008).

22.5.4 Survey Design

The wording of the survey description can affect the respondents' willingness to participate and answer honestly. First, researchers should be sure to stress the purpose of the survey and the way that the results will be used; respondents appear to be more inclined to answer honestly when the benefits of the survey results to society are emphasized (Rasinski et al. 1999). Second, whenever possible, survey designers should avoid the use of lengthy standard introductions (O'Brien et al. 2006). These introductions impede conversational flow, which is important for establishing rapport between the interviewer and respondent. As discussed earlier, this rapport makes the participant feel more comfortable divulging sensitive information to the interviewer. Also, a standard script makes it difficult for the interviewer to discover and address reasons for the respondent's unwillingness to participate in the survey.

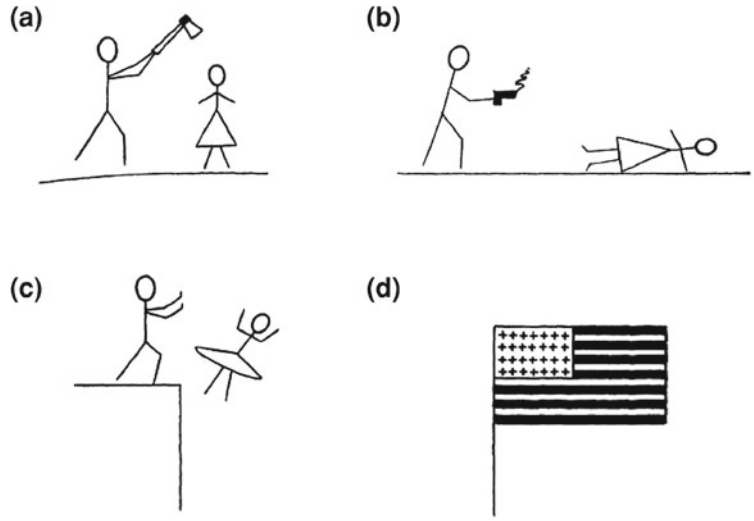
The wording of specific questions can also make an impact on misreporting; effective wording can make the question seem less threatening and put the respondent at ease. A common method of reducing the threat associated with particular topics is to load the question so that the behavior seems normal or common. Loading the question can be done in several ways. First, assume that respondents engage in the behavior of interest and ask about frequency or other more specific details ("How many alcoholic beverages do you consume per day?"). Second, the question can begin with a statement asserting that the behavior is common or understandable ("Even the calmest parents get angry with their children sometimes. Have your children done anything in the past 7 days to make you angry?") This can be combined with the third technique, the use of authority, in which experts are cited to prove to the respondent that the behavior is common or acceptable. Fourth, the question can be asked casually to make the behavior seem unimportant. See Fig. 22.1 for a popular example of these techniques of reducing question sensitivity: the "Did you kill your wife?" example by Barton (1958).

Another method of reducing the threat associated with a survey item is to use clear questioning that reduces confusion and aids recall. Questions that are easy to understand and answer will be less embarrassing for the respondent. The length and specificity of cues have been found to assist the respondent in understanding questions and easing discomfort. Such questions result in higher estimates of sensitive behavior (Blair et al. 1977; Mooney & Gramling 1991). Furthermore, using words that are familiar to the respondent and avoiding the use of jargon or technical terms results in higher estimates of sensitive behavior (Blair et al. 1977).

Indirect questioning is often used as a method of alleviating the threat associated with direct questions regarding sensitive behaviors or attitudes. In one form of indirect questioning, instead of asking respondents for reports of their own behavior or attitudes, the survey asks them to consider a third person (often, the "average" person). The wording of these questions should be as close as possible to that used in direct questioning; the only difference should be the use of third person rather than first person. Because the respondent uses his or her own behavior or attitude as a reference point, these items are meant to proxy the respondent's actual behavior or attitude. This type of indirect question reduces social desirability bias (Fisher 1993).

Indirect questions are believed to be less threatening and remove the tendency of respondents to answer in the way they perceive to be the most socially desirable. However, there are doubts regarding the validity of these items, since the respondent is not giving a true account of his or her behavior or mindset (Slabbinck & Van Kenhove 2010). Furthermore, it is possible

1. **The Casual Approach: "Do you happen to have murdered your wife?"**
2. **The Numbered Card: "Would you please read off the number on this card which corresponds to what became of your wife?" (HAND CARD TO RESPONDENT)**
   a. Natural death
   b. I killed her
   c. Other (What?)
   d. GET CARD BACK FROM RESPONDENT BEFORE PROCEEDING
3. The Everybody Approach: "As you know, many people have been killing their wives these days. Do you happen to have killed yours?"
4. The "Other People" Approach:
   a. "Do you know any people who have killed their wives?"
   b. "How about yourself?"
5. The Sealed Ballot Technique
   a. In this version you explain that the survey respects people's right to anonymity in respect to their marital relations, and that they themselves are to fill out the answer to the question, seal it in an envelope, and drop it in the box conspicuously labelled "Sealed Ballot Box" carried by the interviewer.



6. The Protective Technique: "What thoughts come to mind as you look at the pictures?" (Note the relevant responses will be evinced by picture D).
7. The Kinsey Technique: Stare firmly into the respondent's eyes and ask in simple, clearcut language such as that to which the respondent is accustomed, and with an air of assuming that everyone has done everything, "Did you ever kill your wife?"
8. Putting the question at the end of the interview.

Source: Barton, A.J. (1958).

Fig. 22.1 Loading the question, "Did you kill your wife?"

that respondents are more likely to perceive a socially undesirable behavior or trait as uncommon; this abnormality is the reason that scholars assume the topic is sensitive. Indeed, research shows that results from indirect questions are not free of social desirability bias (Slabbinck & Van Kenhove 2010).

Another type of indirect question is the Randomized Response Technique (RRT) (Warner 1965). Respondents are randomly assigned to answer one of two questions; the interviewer is unaware of which question they have been assigned to. The two questions can be opposite views of a controversial issue (for example, A) I

am in favor of legalizing gay marriage; B) I am against legalizing gay marriage) or one of the questions can be an unrelated question with a known probability for each answer. If the respondent answers, "yes," his or her answer to the item of interest is still unknown to the interviewer. This method is believed to improve reporting rates because respondents should feel less pressure to misreport if the interviewer does not know which question they are answering.

In another variation of RRT, sometimes known as the forced alternative method, respondents answer the same question, but are randomly assigned to either answer correctly or give a predetermined answer. In this case, respondents are expected to answer honestly because the interviewer will not know whether their answers are genuine.

A meta-analysis of studies using randomized response found that utilizing the technique resulted in significantly lower rates of socially desirable answers (Lensvelt-Mulders et al. 2005). However, a review of studies verifying RRT responses against actual incidences of behavior found that the responses still suffer from misreporting (Umesh & Peterson 1991).

It has been suggested that participants do not always understand the instructions or the way that the technique will protect their privacy, making them less likely to follow the directions and answer as instructed (Krosnick & Presser 2009; Holbrook & Krosnick 2010). Further, the unwillingness to report socially undesirable behavior may cause some respondents to lie, even if they have been randomly assigned to the group that is told to answer a certain way rather than answer honestly. A cheating-detection modification can be employed to account for this issue (Clark & Desharnais 1998; Moshagen et al. 2010).

A third type of indirect question is the Unmatched Count Technique (UCT), also called item count. Respondents are given a list of items and asked to give the number of items that apply to them. Respondents are randomly assigned to two groups; one of the groups receives the item of interest as part of the list and the other does not. Results from the two groups can be compared in order to determine the proportion of the sample that answered yes to the additional item. This technique is believed to increase accuracy because respondents are not asked which of the statements are true; therefore, they are less likely to answer dishonestly out of embarrassment or fear of sanctions. The use of UCT results in higher reporting rates of socially undesirable behaviors than direct questioning (Tourangeau & Yan 2007). However, item count does not significantly increase accuracy of reporting of victimization experiences (Krebs et al. 2011). Krebs et al. (2011) noted that estimates using UCT are unreliable since they are based on half the sample; they corrected for this by giving respondents two sets of items. Respondents in this study were randomly assigned to answer either item counts 1 and 4 or item counts 2 and 3 (Fig. 22.2).

Open-ended questions may be better for some threatening topics, as they allow participants to explain their answer rather than choose from a few narrow responses. This technique is recommended when surveying on such topics as victimization and criminal behavior (Smith 1994). For example, respondents may feel unwilling to answer "yes" to a question regarding spousal abuse, as they may prefer to explain the incident rather than vilify their spouse. Open-ended questions are also beneficial in increasing rapport between interviewers and respondents, which encourages accurate reporting (Smith 1994). However, for some questions, such as those asking for demographic information, categorical questions may be better, as they decrease the possibility of identifying the respondent (Giles & Feild 1978).

Open-ended questions are especially useful for asking questions regarding the frequency of socially undesirable behavior (Sudman & Bradburn 1982). The ordinal categories chosen by researchers may cause additional embarrassment to respondents who engage in behavior more often, as people tend to avoid choosing the extreme answers on lists. For example, a question asking respondents how often they drink

Fig. 22.2 Unmatched count technique

Item Count 1. How many of the following things have happened to you since you entered college?

- You have woken up with a hangover because you drank too much the night before
- Someone has pushed, slapped, or punched you
- You have received a failing grade on an exam or a paper
- You have been involved in a car accident
- Someone has stolen something from you

Item Count 2. How many of the following things have happened to you since you entered college?

- You have woken up with a hangover because you drank too much the night before
- Someone has pushed, slapped, or punched you
- You have received a failing grade on an exam or a paper
- You have been involved in a car accident
- Someone has stolen something from you
- Someone has had sexual contact with you by using physical force or threatening to physically harm you

Item Count 3. How many of the following things have happened to you since you entered college?

- You have had an alcoholic beverage before you turned 21
- You have used marijuana
- Someone has offered to pay you to write a paper or take a test for them
- You have lost your credit card or ATM card
- You have had mono (mononucleosis)

Item Count 4. How many of the following things have happened to you since you entered college?

- You have had an alcoholic beverage before you turned 21
- You have used marijuana
- Someone has offered to pay you to write a paper or take a test for them
- You have lost your credit card or ATM card
- You have had mono (mononucleosis)
- Someone has had sexual contact with you by using physical force or threatening to physically harm you

Source: Krebs et al. (2011)

alcohol could simply allow participants to supply a number of alcoholic beverages consumed in a specific period of time.

Sudman and Bradburn (1982) suggest adding questions to the survey to determine whether respondents considered the topics to be threatening or embarrassing. Because the definition of "sensitivity" is unclear, asking participants about their perception of the topics can be an ideal way to determine if the items included in the survey are sensitive. This ability to determine sensitivity is especially important when considering the context in which the question is asked, which can determine the sensitive nature of the subject in question (Barnett 1998). Furthermore, such questions can allow for further research that identifies particular topics as threatening, which can benefit future survey designers. Sections like these are also useful in determining if the respondents believe the researcher's assurances of anonymity (Lautenschlager & Flaherty 1990).

Survey designers should also consider the format of the survey, as the placement of sensitive items can be an important factor in easing the threat associated with particular questions. Several scholars suggest placing sensitive questions that are likely to make respondents feel uncomfortable at the end of a questionnaire (Andrews 1984; Blair et al. 1977; Giles & Field 1978; Krosnick & Presser 2010). According to Andrews (1984), questions do not perform as well at the very beginning of a survey as they do in other parts of the questionnaire. Therefore, more accurate responses can be expected if sensitive items are not included.

Fig. 22.3 Embedding sensitive items

Did you ever, even once, do any of the following?		
Commit armed robbery:	Yes	No
Break into a home, store, or building?	Yes	No
Take a car for a ride without the owner's knowledge?	Yes	No
Take something from a store without paying for it?	Yes	No

Source: Sudman and Bradburn (1982)

Also, respondents are more likely to end the interview if discomfort due to the topic of the survey occurs early.

The formatting of certain questions can influence the answers to those questions. Sensitive items may suffer less from nonresponse if options such as "don't know" or "not applicable" are not available (Krosnick & Presser 2010). Sudman & Bradburn (1982) suggest embedding sensitive items of interest within lists of items that are even more threatening. This type of question can encourage respondents to admit to mildly sensitive behavior or attitudes. For example, rather than simply asking whether the respondent has shoplifted, respondents could answer a series of questions such as that given in Fig. 22.3. Note that the items are in order from most to least severe, encouraging the respondent to view committing shoplifting as relatively harmless and therefore less embarrassing to admit.

22.5.5 Interviewer Training

Researchers conducting telephone surveys should have their interviewers undergo refusal aversion training, in which interviewers are taught how to detect reasons for refusals and encourage respondents to complete surveys (see O'Brien et al. 2006, for a more detailed description of refusal aversion training). This type of training is especially beneficial for surveys with sensitive topics because a participant who decides to complete the survey based on rapport with the interviewer is also more likely to accurately respond to survey questions. Interviewers should be taught to notice and address subtle cues of reluctance to participate or respond. These cues can include the respondent's tone of voice and pace of speech, as well as a tendency for the respondent to respond in a format unlike that which is instructed. When attempting to recruit and interview participants for sensitive telephone surveys, interviewers should be alert to cues that will inform them about the respondent's emotional state, privacy, and current safety.

The comfort level of the interviewer when surveying about sensitive topics may help put the participant at ease, affecting the participant's willingness to answer such questions (Rasinski et al. 1999). Therefore, interviewers should be selected based on their ability to remain professional and calm regardless of the topic being discussed, and training should further build upon this ability. The emotional state of the interviewer is important in avoiding nonresponse; interviewers who are nervous or uncomfortable about the topic are likely to obtain more missing responses than other interviewers (Groves et al. 1992). Respondents are concerned with how the interviewer will react to their answers; this concern affects the truthfulness of reporting sensitive behavior (Rasinski et al. 1999). Because of this concern, it is important that interviewers be trained to remain objective and refrain from visible or audible signs of judgment.

22.6 Conclusion

Survey designers must be mindful of the impact of sensitive questions on the results of their surveys. First, survey items of a threatening or sensitive nature must be identified so that they can be worded and/or formatted appropriately. Key suggestions for the wording and formatting

of sensitive items include: (1) Use open-ended questions when possible, especially for items regarding frequency, (2) Load questions so that the behavior seems common and normal, (3) Use indirect questions when appropriate, and (4) Place sensitive items at the end of the questionnaire or interview. Second, consider the effect of the survey topic when choosing a mode of administration. Self-administered methods may be more effective if the topic is somewhat sensitive due to the increased privacy and anonymity. However, the social exchange associated with interviews and the rapport established between the interviewer and the respondent may make face-to-face interviews more appropriate for extremely sensitive topics. Third, when conducting interviews, be sure that interviewers are trained well so that respondents are comfortable discussing sensitive information with them.

References

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly, 48*, 409–442.

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly, 58*, 210–240.

Barnett, J. (1998). Sensitive questions and response effects: An evaluation. *Journal of Managerial Psychology, 13*, 63–76.

Barton, A. J. (1958). Asking the embarrassing question. *Public Opinion Quarterly, 22*, 67–68.

Beach, S. R., Schulz, R., Degenholtz, H. B., Castle, N. G., Rosen, J., Fox, A. R., & Morycz, R. K. (2010). Using audio computer-assisted self-interviewing and interactive voice response to measure elder mistreatment in older adults: Feasibility and effects on prevalence estimates. *Journal of Official Statistics, 26*(3), 507–533.

Beck, F., Peretti-Watel, P., & Andreyev, Z. (2002). The impact of data collection methodology on the reporting of illicit drug use by adolescents. *Population, 57*(3), 571–592.

Blair, E., Sudman, S., Bradburn, N. M., & Stocking, C. (1977). How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *Journal of Marketing Research, 14*, 316–321.

Blumberg, S. J., Cynamon, M. L., Osborn, L., & Olson, L. (2003). The impact of touch-tone data entry on reports of HIV and STD risk behaviors in telephone interviews. *The Journal of Sex Research, 40*, 121–128.

Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods, 3*, 160–168.

Dillman, D. A. (1978). *Mail and telephone surveys*. New York: Wiley.

Fisher, R. J. (1993). Social Desirability and the validity of indirect questioning. *Journal of Consumer Research, 20*, 303–315.

Fu, H., Darroch, J. E., Henshaw, S. K., & Kolb, E. (1998). Measuring the extent of abortion underreporting in the 1995 National Survey of Family Growth. *Family Planning Perspectives, 30*, 128–133 +138.

Fuller, C. (1974). Effect of anonymity on return rate and response bias in a mail survey. *Journal of Applied Psychology, 59*, 292–296.

Giles, W. F., & Feild, H. S. (1978). Effects of amount, format, and location of demographic information on questionnaire return rate and response bias of sensitive and nonsensitive items. *Personnel Psychology, 31*, 549–559.

Gmel, G. (2000). The effect of mode of data collection and of non-response on reported alcohol consumption: a split-sample study in Switzerland. *Addiction, 95*, 123–134.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41.

Gribble, J. N., Miller, H. G., Rogers, S. M., & Turner, C. F. (1999). Interview mode and measurement of sexual behaviors: Methodological issues. *The Journal of Sex Research, 36*, 16–24.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*(4), 475–495.

Harden, J., Scott, S., Backett-Milburn, K., & Jackson, S. (2000). Can't talk, won't talk? Methodological issues in researching children. *Sociological Research Online, 5,* http://www.socresonline.org.uk/5/2/harden.html.

Hebert, J. R., Ma, Y., Clemow, L., Ockene, I. S., Saperia, G., Stanek, E. J., et al. (1997). Gender differences in social desirability and social approval bias in dietary self-report. *American Journal of Epidemiology, 146*, 1046–1055.

Holbrook, A. L., & Krosnick, J. A. (2010). Measuring voter turnout by using the randomized response technique: Evidence calling into question the method's validity. *Public Opinion Quarterly, 74*, 328–343.

Horm, J, Cynamon, M., & Thornberry, O. (1996). The influence of parental presence on the reporting of sensitive behaviors by youth. In Warnecke, R. (Ed.), *Health Survey Research Methods Conference Proceedings* (pp.141-145), DHHS Pub. No. (PHS) 96-1013.

Jacobs, M. A., Bruhn, C., & Graf, I. (2008). Methodological and validity issues involved in the collection of sensitive information from children in foster care. *Journal of Social Service Research, 34*(4), 71–83.

Johnston, L. D., & O'Malley, P. M. (1985). Issues of validity and population coverage in student surveys of drug use. In Rouse, B. A., Kozel, N. J., & Richards, L. G. (Eds.), Self-report methods of estimating drug use: Meeting current challenges to validity, *NIDA Research Monograph 57* (pp. 31–54).

Kitson, G. C., Clark, R. D., Rushforth, N. B., Brinich, P. M., Sudak, H. S., & Zyzanski, S. J. (1996). Research on difficult family topics: Helping new and experienced researchers cope with research on loss. *Family Relations, 45*, 183–188.

Krebs, C. P., Linquist, C. H., Warner, T. D., Fisher, B. S., Martin, S. L., & Childers, J. M. (2011). Comparing sexual assault prevalence estimates obtained with direct and indirect questioning techniques. *Violence Against Women, 17*, 219–235.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *The Handbook of Survey Research* (pp. 263–314). Bingley, UK: Emerald.

Langhinrichsen-Rohling, J., Arata, C., O'Brien, N., Bowers, D., & Klibert, J. (2006). Sensitive research with adolescents: Just how upsetting are self-report surveys anyway? *Violence and Victims, 21*(4), 425–444.

Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75*, 310–314.

Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research, 33*, 319–348.

Lynn, P., & Kaminska, O. (2011). The impact of mobile phones on survey measurement error. ISER Working Paper Series No. 2011-07.

Makkai, T., & McAllister, I. (1992). Measuring social indicators in opinion surveys: A method to improve accuracy on sensitive questions. *Social Indicators Research, 27*, 169–186.

Mangione, T. W., Hingson, R., & Barrett, J. (1982). Collecting sensitive data: A comparison of three survey strategies. *Sociological Methods & Research, 10*, 337–346.

Marquis, K. M. (1984). Record checks for sample surveys. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Washington, DC: National Academy Press.

Martin, L. M., Leff, M., Calonge, N., Garrett, C., & Nelson, D. E. (2000). Validation of self-reported chronic conditions and health services in a managed care population. *American Journal of Preventive Medicine, 18*, 215–218.

McDaniel, S. W., & Rao, C. P. (1981). An investigation of respondent anonymity's effect on mailed questionnaire response rate and quality. *Journal of the Market Research Society, 23*, 150–160.

Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal cohort: Individual and interviewer effects. *Public Opinion Quarterly, 52*, 100–124.

Mensch, B. S., Hewett, P. C., & Erulkar, A. S. (2003). The reporting of sensitive behavior by adolescents: A methodological experiment in Kenya. *Demography, 40*, 247–268.

Mensch, B. S., Hewett, P. C., Jones, H. E., Luppi, C. G., Lippman, S. A., Pinho, A. A., et al. (2008). Consistency in women's reports of sensitive behavior in an interview mode experiment, São Paolo, Brazil. *International Family Planning Perspectives, 34*(4), 169–176.

Millstein, S. G. (1987). Acceptability and reliability of sensitive information collected via computer interview. *Educational and Psychological Measurement, 47*, 523–533.

Mooney, L. A., & Gramling, R. (1991). Asking threatening questions and situational framing: The effects of decomposing survey items. *The Sociological Quarterly, 32*, 289–300.

Moshagen, M., Musch, J., Ostapczuk, M., & Zhao, Z. (2010). Reducing socially desirable responses in epidemiologic surveys: An extension of the Randomized-Response Technique. *Epidemiology, 21*, 379–382.

Motl, R. W., McAuley, E., & DiStefano, C. (2005). Is social desirability associated with self-reported physical activity? *Preventive Medicine, 40*, 735–739.

National Institute of Justice. (2003). *2000 arrestee drug abuse monitoring: Annual Report*. Washington, DC: United States Department of Justice.

NIMH. (2008). Designing an Audio Computer-Assisted Self-Interview (ACASI) system in a multisite trial: A brief report. *Journal of Acquired Immune Deficiency Syndromes, 49*, S52–S58.

O'Brien, E. M., Black, M. C., Carley-Baxter, L. R., & Simon, T. R. (2006). Sensitive topics, survey nonresponse, and considerations for interviewer training. *American Journal of Preventive Medicine, 31*, 419–426.

Phillips, S. R. (1994). Asking the sensitive question: The ethics of survey research and teen sex. *IRB: Ethics and Human Research, 16*(6), 1–7.

Poulin, M. (2010). Reporting on first sexual experience: The importance of interviewer-respondent interaction. *Demographic Research, 22*(11), 237–288.

Pridemore, W. A., Damphousse, K. R., & Moore, R. K. (2005). Obtaining sensitive information from a wary population: A comparison of telephone and face-to-face surveys of welfare recipients in the United States. *Social Science and Medicine, 61*, 976–984.

Rand, M., & Rennison, C. M. (2005). Bigger is not necessarily better: An analysis of violence against women estimates from the National Crime Victimization Survey and the National Violence Against Women Survey. *Journal of Quantitative Criminology, 21*, 267–291.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology, 13*, 465–484.

Records, K., & Rice, M. (2006). Enhancing participant recruitment in studies of sensitive topics. *Journal of the American Psychiatric Nursing Association, 12*, 28–36.

Rhodes, K. V., Lauderdale, D. S., He, T., Howes, D. S., & Levinson, W. (2002). Between me and the computer: Increased detection of intimate partner violence using a computer questionnaire. *Annals of Emergency Medicine, 40*, 476–484.

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly, 74*(5), 907–933.

Silver, B. D., Abramson, P. R., & Anderson, B. A. (1986). The presence of others and overreporting of voting in American national elections. *Public Opinion Quarterly, 50*, 228–239.

Slabbinck, H., & Van Kenhove, P. (2010). Social desirability and indirect questioning: New insights from the Implicit Association Test and the Balanced Inventory of Desirable Responding. Paper presented at the 2009 Association for Consumer Research North American Conference.

Sloan, J. J., Bodapati, M. R., & Tucker, T. A. (2004). Respondent misreporting of drug use in self-reports: Social desirability and other correlates. *Journal of Drug Issues, 34*, 269–292.

Smith, T. W. (1992). Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Social Biology, 39*, 203–211.

Smith, M. D. (1994). Enhancing the quality of survey data on violence against women: A feminist approach. *Gender and Society, 8*, 109–127.

Smith, T. W. (1995). The impact of the presence of others on a respondent's answers to questions. *International Journal of Public Opinion Research, 9*, 33–47.

Steeh, C. (1981). Trends in nonresponse rates. *Public Opinion Quarterly, 45*, 40–57.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass.

Szinovacz, M. E., & Egley, L. C. (1995). Comparing one-partner and couple data on sensitive marital behaviors: The case of marital violence. *Journal of Marriage and Family, 57*, 995–1010.

Tourangeau, T., & McNeeley, M. (2000). "Measuring Crime and Victimization: Methodological Issues" unpublished manuscript. Survey Research Center, University of Michigan. (2002). In Pepper and Petrie (editors). *Measurement Problems in Criminal Justice Research: Workshop Summary*. Washington D.C.: The National Academies Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly, 74*, 413–432.

Tracy, P. E., & Fox, J. A. (1981). The validity of randomized response for sensitive measurements. *American Sociological Review, 46*, 187–200.

Turner, C., Ku, L., Sonenstein, F.L., & Pleck, J.H. (1996). Impact of ACASI on reporting of male–male sexual contacts: Preliminary results from the 1995 National Survey of Adolescent Males. In Warnecke, R. (Ed.), *Health Survey Research Methods Conference Proceedings* (pp.171–176), DHHS Pub. No. (PHS) 96-1013.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867–873.

Umesh, U. N., & Peterson, R. A. (1991). A critical evaluation of the randomized response method: Applications, validation, and research agenda. *Sociological Methods and Research, 20*, 104–138.

Visser, A. P., Eijkman, M. A., & Wiegman, H. (1989). Social desirability and questioning children in dental research. *Community Dentistry and Oral Epidemiology, 17*(3), 127–130.

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*, 63–69.

Researching Difficult Populations: Interviewing Techniques and Methodological Issues in Face-to-Face Interviews in the Study of Organized Crime

23

Jana Arsovska

23.1 Researching Organized Crime

There are many reasons for conducting criminological research; the primary one being the need to provide convincing answers to complex questions about human nature. One decade ago, I applied for a PhD scholarship at University of Leuven in Belgium. The topic of my dissertation was: "The role of cultural codes in the evolution of Albanian organized crime." The decision regarding my future funding was in the hands of a panel composed of twelve professors, all of whom were from different departments and disciplines. I was invited for an interview, and the first question the jury asked me, was: "Explain how do you plan to investigate the Albanian mafia? How does one actually study "the mafia"?" None of the people on my panel was a criminologist and this field of research seemed foreign to them, yet I managed to convince them that research on organized crime is feasible.

Conducting an objective study on topic such as organized crime is without a doubt challenging. The problem of doing reliable empirical research on an "entity" concealed in the criminal underworld is of main concern. Due to its obscure nature, organized crime is less accessible research field. Only rarely law enforcement officers and researchers can observe the activities of a criminal group and gain direct access to the internal dynamics of a criminal organization.[1] Since accounts of offenders and informers can be self-glorifying fantasies, one should also remain skeptical of their reliability.

Social scientists are forced to rely on accounts given by police, judicial authorities, and the media, and on reports of witnesses. Getting access to police files and interviewing victims or witnesses is also not a simple task. The police are not eager to give up their "information position" and the victims are often reluctant to say a word.

Experiences show that victims of organized crime also fear retribution and are afraid to come forward and speak to police officers. Trafficking victims in particular tend to stay "underground" due to their illegal status. Many are also traumatized as a consequence of the physical and psychological abuse they endure. Systematic isolation makes these victims dependent on the traffickers. In some cases, they do not consider themselves victims and refuse to provide reliable information to outsiders. Moreover, as collaborators of justice, police informants are often entitled to police protection, living allowances, and a reduction in their prison terms, which also may affect the reliability of their testimonies.

J. Arsovska (✉)
Department of Sociology, John Jay College of Criminal Justice, 899 Tenth Avenue, New York, NY 10019, USA
e-mail: jarsovska@jjay.cuny.edu

[1] There has been a limited number of qualitative studies where the researchers gathered data by developing a sustained relationship with criminals while they go about their normal activities (Chambliss 1988; Bourgois 1995a, 1995b; Blok 1981; Wright & Decker 1994; Ianni 1974).

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_23, © Springer Science+Business Media New York 2012

Thus, reliance on police sources and the media is not unconditional.

Furthermore, collecting scientific data on organized crime across countries is difficult because of definitional problems and language barriers. As Paoli (2003) wrote, the legal view tends to privilege those aspects of the phenomena that assume a greater relevance under the judicial-formal profile. There is also no guarantee that official agencies concentrate on the most serious crime problems in their territory.

Although organized crime researchers face numerous methodological challenges, the topic offers a unique opportunity to utilize an interdisciplinary research methodology. It allows researchers to combine elements of various distinctive research techniques into a unique data-collection strategy. This chapter evaluates employed methodologies in the study of organized crime. It provides a personal reflexive account of the fieldwork and survey experience, and it elaborates on the main challenges and lessons learned from attempting to conduct study in this field. Then, it addresses issues critical to the proper design and successful implementation of various quantitative and qualitative methods in the study of organized crime. It discusses the method of (multiple) triangulation—used to obtain confirmation of findings through convergence of different perspectives. Non-probability and probability sampling techniques including purposive, snowball, and cluster sampling are discussed, as well as the proper use of distinctive research methods such as face-to-face interviews and focus groups. This chapter elaborates on methodological challenges in face-to-face interviews with experts, difficult populations, and common people.

23.2 Researching the Literature and Developing a Research Question

A criminological research begins with a clear statement of the research problem, which affects all parts of the subsequent study. Before developing research methods, researchers must have narrow research question, know the literature on their topic, and be familiar with how research in the field of organized crime has been conducted in the past.

However, what is a clearly defined research question? Research has shown that no single factor explains why some individuals engage in illegal activities, or why crime is more prevalent in some communities than in others. Violence and crime are the result of the complex interplay of individual, social, cultural, and situational factors. They should be seen as products of multiple levels of influence on behavior: biological and personal historical factors that individuals bring to their behavior; proximal social relationships; the community contexts in which social relationships are embedded; and factors that create an acceptable climate for crime (Krug et al. 2002, pp. 12–13). This does not mean that the research questions should be vague and complex.

Studying Albanian and West African organized crime networks the main focus has been on the larger societal factors in Albania and West Africa that influence peoples' perceptions of violence and crime. For example, studying the importance of the Albanian customary code, the Kanun of Lek Dukagjini—considered the foundation of the Albanian culture—in order to understand the behavior and structure of contemporary ethnic Albanian organized crime groups (see Arsovska & Verduyn 2008; Arsovska & Craig 2006; Arsovska & Lantsman 2010; Arsovska 2007). Issues such as honor, revenge killings, violence, codes of silence (besa), trust, clan mentality and attitudes toward women have been analyzed in the context of organized crime.

Similarly, my research looks at the relevance of "voodoo" rituals, "black" magic, obia/obeah, juju, and other cultural practices in West African immigrant communities, and the extent to which these practices are used by West African organized crime groups, particularly human trafficking networks, to gain power, profit, and/or control over their victims' lives (Arsovska & Lantsman 2010).

These are some examples of research questions developed for the purpose of the above-mentioned studies: "What role did the Kanun Laws play in the evolution of Albanian organized crime groups in Europe?"; and "How, and if, West African human trafficking networks use cultural rituals/practices to control their trafficking victims?" These questions were selected because there is often bias against immigrant communities worldwide as well as lack of understanding of culture-specific behaviors. Terms like "cultural defense" have become commonly misused in colloquial conversation. This has allowed criminals to use "culture" for justification of their criminal behavior. Thus, my main research interest has been to gain understanding of the interplay between culture-specific behaviors and organized crime.

Once the researchers have developed and operationalized their main research questions, they need to familiarize themselves with proper research methods in their field of studies. Reviewing the literature and knowing the limits of existing research techniques is of uppermost importance. The available literature provides researchers with a rich overview of research methods, such as studying archive data, police and judicial reports, criminal biographies, life histories and transcripts of trials, interviewing informants, offenders and victims, analyzing documents, and conducting ethnographic studies. A field research design that enables researchers to get hold of primary source data has been highly desirable; however, as mentioned above, ethnographic studies on organized crime that involve participant observation are rare.[2] The question thus remains: What are some good research methods to be applied in the study of organized crime?

23.3 Multiple Triangulation

Researchers need to decide which individuals, families, groups, organizations communities, or events they would like to study in order to answer their research question(s), and what sample they will choose for this purpose. They also need to make a plan on how to carry the study and to decide on the research design and methods.

Since organized crime is a less accessible research field, scholars have tried to combine various distinctive research methodologies to explain the phenomenon. The method of multiple triangulation can be an appropriate method for researching organized crime figures (or so-called mafia bosses), prostitutes, pimps, punters, and other hidden populations. In the social sciences, triangulation is often used to indicate that more than two methods are applied in a study of the same phenomenon to double or triple-check the results (Baulch & Scott 2006; Booth et al. 1998; Jakob 2001). This is a type of "cross examination."

Triangulation is often considered an alternative to criteria like reliability and validity. *Reliability* is the relative absence of errors of measurement in a measuring instrument.[3] The general procedure to increase reliability is to minimize the error variance. Poor wording, unclear recording procedures, lack of clear standardized instructions, and other errors may lead to low levels of reliability of the instrument. The *validity* of an instrument is the extent to which it measures what is supposed to measure. Making use of multiple research methods, i.e., triangulation, improves the reliability and the validity of a study. The idea is that one can be more confident with a result if different methods lead to the same result.

[2] American FBI Agent Joe Pistone infiltrated for 6 years (1976–1982) in the Bonnanno family in New York (1988). Donnie Brasco: my undercover life in the Mafia; Chambliss (Chambliss 1988). On the take: from petty crooks to presidents. Chambliss did a detailed analysis of the symbiotic relationship between organized criminals, the police, politicians and businessmen in his study of Seattle (USA). See also: Anton Blok (1981). The Mafia of a Sicilian Village: a study of violent peasant entrepreneurs 1860–1960; Ianni (1974). Black mafia: ethnic succession in organized crime; Rubinstein & Reuter (1977). Numbers: The Routine Racket; Fact, fancy and organized crime (1978). See also Polsky (1969); Ianni (1972); Taylor (1984); Adler (1985); Williams (1990); Hobbs (1995); Bourgois (1995a, 1995b); Zaitch (2002).

[3] Instruments such as test–retest method, the multiple or alternate forms method, the split-half method and internal consistency measure exist to establish the reliability of an instrument (see Raymond 1996).

If a researcher uses only one method, the temptation is strong to believe in the findings. If two methods are used, the results may contradict each other. By using three or more methods, the researcher hopes that two of the three sources will produce similar results. If three contradictory answers are produced, then the researcher knows that the question needs to be reframed and the methods reconsidered. Therefore, by combining multiple observers, theories, methods, and empirical materials, criminologists can hope to overcome the intrinsic biases linked to single method studies. The use of different qualitative and quantitative approaches, and exploiting the strengths of each, is necessary in the study of organized crime since there is limited literature on the topic and the research subjects are less accessible.

In the following, several methods—both quantitative and qualitative—for researching difficult populations are discussed as well as sampling strategies, interviewing techniques, and methodological issues in face-to-face interviews.

23.4 Non-Probability Sampling and Expert Panels

Sample is a portion of a population selected for study. Researchers often come across terms like random sampling—a procedure for drawing a sample from a population so that every element in the population has an equal chance of being selected. Although positivistic social scientists prefer random sampling methods, they are limited by constrains, such as budget, unwilling participants, and lack of information about population. A number of sampling procedures have been developed to cope with these problems (Raymond 1996).

For instance, non-probability sampling does not involve random selection, and, therefore, it does not depend upon the rationale of probability theory. With non-probability samples, the researcher may not necessarily represent the population well. Nevertheless, in social research, there are circumstances where it is not feasible to do random sampling. If one studies organized crime, for example, probability sampling is not efficient sampling method, as there is no available sampling frame—a list of members who are affiliated with the illegal group and activities that are examined.

Purposive sampling is one type of non-probability sample, often preferred by qualitative or grounded theory researchers[4] (Raymond 1996), p. 216. Purposive sampling can be useful for situations where one needs to reach a targeted sample quickly, and where sampling for proportionality is not of concern. In this method, study participants are handpicked by the researcher to serve the purpose of the study (Raymond 1996), p. 113. Researchers might sample for specific types of people as in modal instance, expert, or quota sampling. They might also take advantage of informal social networks to identify respondents who are hard to locate otherwise, as in snowball sampling.

Although most qualitative researchers limit their focus to just one or a few sites, so that they can focus all their attention on the social dynamics of those settings, this does not mean that sampling is unimportant. Qualitative researchers use theoretical sampling when they focus their investigation on particular processes that seem to be important and select instances to allow comparisons or checks with which they can test these perceptions (Fig. 23.1).

Purposive sampling is useful for organized crime research, because one way to obtain information is to interview law enforcement officials who have investigated organized crime cases; another option is to talk to other experts in the field. The most important aspect is identifying the "expert" population and deciding on the type of interviews to be conducted. The process of expert sampling involves the assembling of a sample of persons with demonstrable expertise in a given area. For example, when conducting research on organized crime, the researcher might convene an expert panel consisting of persons with insight into this field and ask them to examine the researchers' hypotheses, concepts and definitions,

[4] Grounded theory is a method for studying complex social phenomena that was developed and first presented in 1967 by two sociologists, Glaser & Strauss. Rather than starting with a theory, generating hypotheses from that theory, and then testing them, in this approach, theory is generated from data. Grounded theorists usually study areas in which there is little or no theory to explain human behavior (Raymond 1996, p. 215).

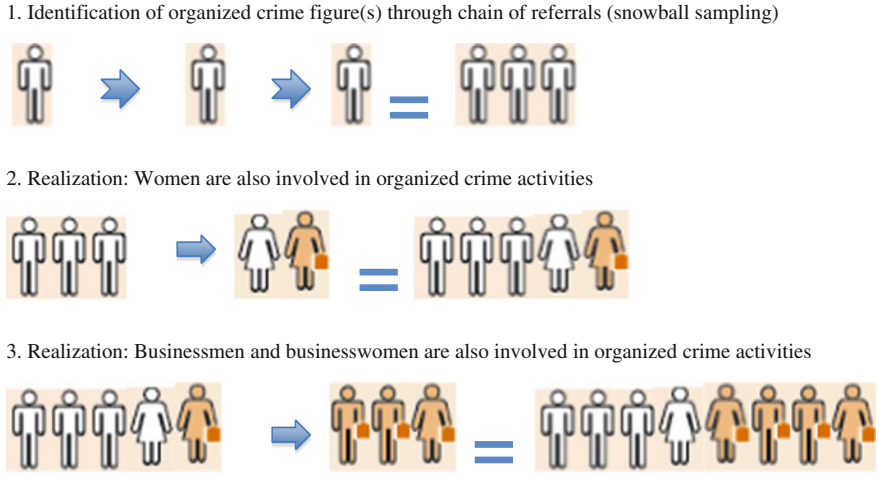


Fig. 23.1 Identifying members of organized crime groups by using theoretical and snowball sampling techniques

and comment on their appropriateness and validity. Expert interviews are particularly useful during the first stages of the research.

Using theoretical sampling, it is the task of the researcher to identify the experts in his or her field and to establish contact with them. Researchers should try to select interviewees who are not only very knowledgeable about the subject of the interview but who are also open to talking and who represent a range of perspectives. The researcher might want to consider interviewing the experts individually (face-to-face depth interviews), or to run focus groups. Both intensive (depth) interviewing and focus groups are considered qualitative methods that involve open-ended, relatively unstructured, questioning in which the interviewer seeks in-depth information on the interviewer's feeling, experiences, and perceptions (Lofland & Lofland 1984).

During the research on Albanian organized crime, the goal was to gain better understanding of the phenomenon before developing research questions. Intensive face-to-face interviews with law enforcement officials seemed like logical thing to do. Locating specialized police units that deal with Albanian criminal groups, and searching for the most knowledgeable officers in this area was the first step. In order to get in contact with the right police officers, researchers often relay on chains of referrals made by other experts in their respective fields. Identifying experts from reputable newspaper articles that make references to names of detectives or prosecutors working on cases of interest is another way to identify an "expert population."

Once the interviewees have been identified, the interviewer needs to set an appointment for an interview. The interviews usually take about 1 h. Interviewers should be open to suggestions but should never go unprepared to an interview. They must plan carefully their main questions around an outline of the interview topic. The questions should generally be short and to the point. Semi-structured interviews with primarily open-end items are effective in interviews with experts. Having a pilot testing of the draft questionnaire is always beneficial.

For my project two semi-structured questionnaires with both open-end and closed-end items were developed. Questionnaire N°1 aimed at assessing the opinions of organized crime experts familiar with ethnic Albanian organized crime groups. It required that the respondent has a comprehensive understanding of the organized crime situation in his/her country, as well as a solid knowledge of Albanian-speaking criminal groups

operating therein. This part targeted experts on the topic, but not necessarily law enforcement officials.

Questionnaire N°2 targeted law enforcement officials (prosecutors, judges and police officers) from different European countries that have been actively involved in arrests, prosecutions, and investigations of ethnic Albanian criminals in their own countries. The questions were structured around issues such as methods of operation, group structure and group dynamics, codes of conduct, and attitudes. The questionnaire was used during the interviews as a guide, see Box 23.1. After this type of interviews, the researcher might consider adjusting the research questions or changing the direction of the research project. One disadvantage of this is that experts are often wrong, thus the researchers should always cross-examine and double-check their answers by comparing them to other data sources. In fact the researchers might use expert sampling to provide evidence for the validity of another sampling approach they have chosen.

Box 23.1: Example Questions for Law Enforcement Officials (Closed- and Open-End)

Please provide a short summary/background of the case you have selected:

_____

_____

_____

_____

_____

_____

_____

1. English name, aliases of group; original language name, or acronym:

_____

_____

☐ NOT KNOWN ☐ N/A

2. Is the group defined as **organised crime** group in your country?

☐ YES ☐ NO ☐ N/A

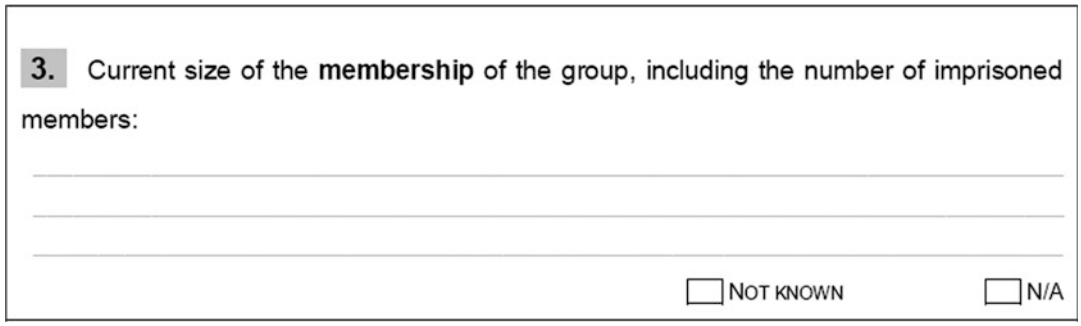
► *If yes/no please provide explanation why:*

_____

_____

_____

_____


3. Current size of the **membership** of the group, including the number of imprisoned members:

NOT KNOWN N/A

Focus groups, if planned well, are very beneficial for researchers working on understudied topics. Focus groups are groups of unrelated individuals that are formed by a researcher and then led in group discussion of a topic for about 2 h. The aim is to identify 8–12 experts and to invite them to be a part of an "expert panel." The researcher asks specific questions and guides the discussion to ensure that group members address these questions, but the resulting information is qualitative and relatively unstructured. Focus group methods can be an indispensable aid for developing hypotheses and survey questions, for investigating the meaning of survey results, and for quickly assessing the range of opinion about an issue. This method of gathering information is less costly and less time-consuming.

It is good to start a focus group discussion with a vignette (crime story) and to structure the discussion around it, see Box 23.2 and Box 23.3. The following vignettes were used in a study on the role of voodoo rituals in West African trafficking networks (Arsovska & Lantsman 2010).

Box 23.2: Albinism and Witchcraft: Importance of Cultural Practices

In 2007, light was shed on the murders of 53 children and adults suffering of albinism in Africa. These killings were done for voodoo purposes. Potions and charms were made from the skin, hair and body parts of the people suffering from albinism in order to "ensure wealth, good health and good fortune" to the people who posses them. Ever since then the numbers of murders have been on the rise. When the trend of using albino for witchcraft first started, the perpetrators were robbing bodies from graveyards and selling them to witch doctors. Once the grave robbers ran out of bodies to steal they started terrorizing and killing people.

Open-ended questions (part of semi-structured questionnaire):

1. Are you familiar with some cases of murders of African people suffering from albinism?
2. Why and how are albino people used in witchcraft?
3. How widespread is this phenomenon in your country?

Box 23.3: Control Mechanism and West African Trafficking Networks

Mrs. Akouavi Afolabi and her ex-husband Lassissi Afolabi of Togo were sentenced to 27 years for human smuggling of West African girls as young as 10 to work in hair braiding salons in New Jersey. Between October 2002 and September 2007 at least 20 girls and women were taken from Togo using fraudulent visas. Mrs. Afolabi made her victims undergo a voodoo ritual to frighten them into believing they would go insane if they escaped. Before leaving for America, the girls were given kola nuts to eat by a man in Togo who told them they were taking a vow of loyalty to the woman.

Open-ended questions (part of semi-structured questionnaire):

1. What is the significance of rituals for members in the American West African diasporas?
2. What is the prevalence of rituals and voodoo practices in the West African diasporas?
3. What is known about the methods of control employed by West African trafficking networks?

4. Are you aware of some ways in which West African organized crime groups in your country use voodoo rituals?

Although vignettes are useful during interviews, the researcher might decide to have open-ended questions structured around different themes not necessarily around vignettes. The questions should serve as a guide, but the discussion on the various topics can be open. Open-ended items usually provide a frame of reference while putting a minimum restrain on the respondent's answer. There are some things the researcher should take into consideration when preparing for the interview and developing open-end questions: (1) Is there a purpose for every item on the questionnaire? (2) Is the wording of the instrument appropriate to the level of the respondent? (3) Are closed- or open-ended questions most appropriate for the information we want? (4) Does the question assume knowledge that the respondent might not have? (5) Does the instrument use filter questions? (6) Have you avoided double barrel questions? (7) Have you avoided leading questions? (8) Is the order of the items appropriate? (9) Does the appearance of the instrument make it easy for the respondent? (10) Have you pretested the instrument? Paying closer attention to these points could improve the reliability and validity of the measuring instrument, and as a result the responses received, thus minimizing response biases.

Furthermore, tape recorders are commonly used to record interviews; however, from personal experiences it is more comfortable for the interviewee if tape records are not used. Written notes are effective means of recording interviews in the field of organized crime (Emerson et al. 1995). It is almost always a mistake to try to take comprehensive notes while engaged in the discussion with the research subjects—the process of writing is disruptive. The usual procedure is to jot down brief notes about highlights of the conversations. These brief notes (called jottings) can then serve as memory joggers when writing the actual notes. Usually writing up notes takes much longer—at least three times longer—than the observing did. Direct quotes should be distinguished from paraphrased quotes. Selection of new interviewees should continue, if possible, until the saturation point is reached, the point when new interviews seem to yield little additional information.

Comparing the law enforcement perspectives with those of NGO representatives, representatives from victim shelters, journalists working in the field, and criminal lawyers is important for triangulating the collected data and for more reliable conclusions. "Mixed" focus groups often yield good results because of the extensive—and often constructive—discussions between experts with somewhat different opinions. Finally, establishing a good relation with law enforcement officials and NGO representatives is crucial because it can lead to gaining access to police and court files or to having the possibility to interview victims and offenders.

23.5 Difficult Populations, Qualitative Research, and Snowball Sampling

Participant observation is another qualitative method for gathering data on organized crime that involves developing a sustained relationship with people while they go about their normal activities. It is based on the need to study intensively people, places, or phenomena of interest. The first concern of all participant observers is deciding what balance to strike between observing and participating. In complete observation, researchers try to see things as they happen, without participating in these events. The researcher must take into account the ways in which his or her presence as an observer alters the social situation being observed. Some field researchers adopt a complete participation role, in which they operate as a fully functioning member of the setting. This, however, raises a lot of ethical and security questions.

Most field researchers adopt a role that involves some active participation in the setting. This is known as mixed participant observation/participation. In this case, usually researchers inform at least some group members of their

research interests, but then they participate in some group activities to gain a direct sense of what the group members experience.

Participant observation can provide a rich view of the social world but it remains only a view. The researchers need to investigate what the subjects think as well. For this purpose the researcher should conduct intensive interviews as those conducted with the experts. No doubt that if one studies organized crime, culture, values, and perceptions, interviewing offenders, despite some disadvantages, is a valuable asset. In order to understand the criminal actions, it is useful to consider their self-created narratives that help to give form and significance to those actions. These inner narratives can be a goldmine for researchers.

In the Albanian study, several unstructured and semi-structured interviews were conducted with less accessible research subjects—the actual criminals. Mixed participation was the preferred research method. While reading some police files on Albanian-speaking groups, an important criminal character became available—a Kosovo-Albanian offender labeled by Western European police agencies as an Albanian mafia boss. When arrested in the late 1990s, the offender was given a prison sentence on the basis of multiple charges, such as, membership in a criminal organization, prostitution, women trafficking, kidnapping, racketeering, extortions, possession of weapons, and fights. He was suspected of arms trafficking, financing of the Kosovo Liberation Army (KLA), controlling his gang from prison, and even murders.

In 2005, the offender was released from prison on the condition that he returns to his place of birth, Kosovo. Conducting a face-to-face interview with him seemed like a great idea at that time; using contacts from the Kosovo police department an attempt was made to get in touch with the offender. This approach did not work out. After 2 months of waiting for a reply, the Kosovo police made the decision not to get involved.

During one of my stays in Kosovo, I met a person who indirectly knew the offender and offered his help in providing contacts. Through my contact, I met two of the offender's close friends, and we had extensive discussions on several occasions. They wanted to make sure that I am not linked to the police before they put me in contact with "their friend." It is natural to ask, why on Earth should the offenders talk to you? You need to overcome their initial suspicions: Who are you? Why are you here? Who is paying for this so-called research you say you are doing? Are you an undercover cop? Are you looking for illegal immigrants? One of the points emphasized by those who have done this sort of research is; do not ask for names of the people you interview. Make it clear that you are not interested in names and not interested in any information about particular crimes. That will help to establish that you are not working for law enforcement. Patience, flexibility, and trustworthiness are important characteristics for an organized crime researcher.

After several meetings, an interview with the offender was finally arranged. We met in a very informal setting—a coffee bar in Prishtina. His friends were present during our meeting, as well as my initial contact. The starting point in this type of research is often difficult.

During my first interview, I had several pages filled with questions. The questions were more general, and their aim was to assess the offender's perceptions and opinions on various issues such crime, honor, gender, politics, institutionalized justice, revenge, values, ambitions, and so on, see Box 23.4. To my surprise, the offender responded to my questions with great care. During the interviews no recording device was used, only notes were taken.

Box 23.4: Unstructured Interview (Guiding Questions—Examples)

A. General

1. Tell me little bit about your life. When and where were you born? Tell me more about your childhood.
2. What did you experience from the regime of Milosevic? Were you in favor of the non-violent philosophy of LDK and Ibrahim Rugova?
3. When did you move to Belgium? Why? With whom? For which purpose?

4. Did you go to school there? Did you have some relatives there to help you out? Was it easy for you to get a job? What kind of job did you get?

B. Court

5. Can you tell me something about the court decision and your case? For which offences you were charged? How many people were involved, arrested, prosecuted? Please give me a summary of the case?
6. Did you find the court judgment just? Was the trial fair? Were you the only person charged/prosecuted?

C. Organized Crime

7. What do you think of the statement "the Albanian mafia is the new Sicilian Mafia of Europe"? Is there strong Albanian mafia nowadays in Europe?
8. What do you think, which are the core activities of the Albanian organized crime groups nowadays?

D. Values

9. What about the role of the women in Albanian society? What is your view on this?
10. What is your perception of prostitution? Should it be legalized or prohibited? Why?

When comparing the offender's statements with those of law enforcement officials, there are many differences in opinions. The "facts" were no longer "facts". However, the question is how far the researchers can trust what people are saying to them. The offenders may be looking for an opportunity to brag about how important they are; they may exaggerate their achievements, or they may talk because they want to show how unhappy they are in a life of crime. There are all sorts of motives at work; nevertheless, the offender's perspective is very important because it allows the researcher to see the offender's world through his or her eyes. As one law enforcement official pointed out, "it is interesting to note, that despite the fact that criminals tend to make things 'slightly' more beautiful than reality, in general they never completely lie". This must be why, at our level, we are inclined to corroborate information from different sources and try to look at things from different perspectives to get a clearer global image. Matching criminal stories with court cases, police files, victim statements, and journalistic accounts is what brings us closer to the actual "truth".

After the first interview, the offender agreed to meet me again. The second interview was to take place in the same coffee place as the first one. This time I met the offender alone, but I instructed a local friend familiar with the situation to give me occasional phone calls in order to check if all is going well. Thirty minutes into the interview the offender's phone rang. He suggested that we continue the conversation in his house because he needs to bring some bricks and construction materials back home since his uncle was working on the house. Trust and flexibility are some key issues raised in research with difficult populations. Becker (1965, p. 602) argued that, "No matter how carefully one plans in advance, the research is designed in the course of its execution. The finished monograph is the result of hundreds of decisions, large and small, made whilst the research is underway." I was in a dilemma and was not sure if it was safe to get into a car with a well-known offender. At that moment I decided to go, and the offender did what he promised to do. We drove to his house after picking up bricks from a local store. He showed me documentaries of his trials and photos of his family, friends, and associates.

We spent about two additional hours in the house, having discussions on variety of topics. He remained relatively open and patient throughout the whole conversation. Then we went to dinner in a famous restaurant in Prishtina, where he introduced me to several other people—including the owner of the restaurant and some ex-prisoners, with whom I had a brief discussion. The second round of interview was extremely beneficial for my research. I believe that I did gain the offender's trust because I showed him my courage and honest intentions. He was comfortable with me, which ensured better quality answers to my questions. This

time the interview was unstructured and, again, no recording device was used. I was, however, allowed to take a series of photos from him and his home.

Sharpe (2009) conducted a street corner research on prostitutes and pimps and noted that her fieldwork too did not get off to a promising start because of the lack of trust among prostitutes who feared their names would be published or that they would be recognized by their description. One month into her research Sharpe (2009, p. 367) wrote in her journal:

> The women are practically ignoring me as if I do not exist. Some of them look straight through me, some of them look me up and down, some of them leave me in no doubt that if looks could kill…I have the feeling that any other intelligent person doing this kind of research would know exactly what to do and would say all the right things and everything would be fine, but I seem to be having to make this up as I go along and adapt to whatever situation I find myself in. I have been sworn at, spat at, laughed at, ridiculed and threatened. This is much harder in reality than I ever imagined it would be.

Months into her research Sharpe finally gained the trust of the women. She notes that standing her ground in the district and absorbing the abuse and threats eventually earned her a sort of respect from the prostitutes because she has showed to them that she has courage. There is a problem, however, with the subject becoming too comfortable with the interviewer. The researchers may find themselves having to do things for the offenders in order to establish their trustworthiness. Zaitch (2002), for example, found himself helping with legal advice and language issues, translating immigration documents, etc. But, as he argues, it is vital not to get compromised.[5]

Going back to my fieldwork experience, over a period of several months the offender put me in contact with two other important ethnic Albanian criminals who operated internationally. This chain of referrals is known as snowball sampling method, a special non-probability method used when the desired sample characteristics are rare. Snowball sampling relies on referrals from initial subjects to generate additional subjects (Vogt 1999). This strategy can be viewed as a response to overcome the problems associated with sampling concealed populations such as the criminals (Faugier & Sargeant 1997). Snowball sampling can be placed within a wider set of link-tracing methodologies which seek to take advantage of the social networks of identified respondents to provide a researcher with an ever-expanding set of potential contacts (Spreen 1992; Thomson 1997). This process is based on the assumption that a "bond" or "link" exists between the initial sample and others in the same target population, allowing a series of referrals to be made within a circle of acquaintance (Berg 1988).

Snowball sampling can be applied as an "informal" method to reach previously hidden populations (Hendricks & Blanken 1992). Techniques of "chain referral" may imbue the researcher with characteristics associated with being an insider or group member and this can aid entry to settings where conventional approaches find it difficult to succeed. Snowball sampling has been found to be economical, efficient, and effective in various studies (van Meter 1990).

In an organized crime research, the researcher ideally would also aim to interview victims or vulnerable populations. The purpose of an interview is to establish the facts of the case, to use the facts to corroborate the victim's story and establish his/her credibility, and to continually assess the risk to the victim and those associated with the victim. Effective interview techniques with vulnerable populations are vital in this respect. The safety and well-being of the victim should be a priority at all times. For example, trafficked victims and prostitutes working on the streets are often vulnerable or traumatized, and they should be interviewed in a sensitive manner, ideally by a trained interviewer.

Before conducting the interview, the researcher should find out about the age of the victim, the language, and the country of origin, including culture and state of development. Risk assessments must ensure the safety of the victim, and if there is some danger to the victim's life if

[5] See also Arsovska (2008).

she discusses the case with the researcher then the interview should not take place. The gender of the interviewer matters, too. Translators, if there is a need to use them, also have an important role. They must be reliable and unbiased. Translators are often more capable of winning the trust of the victim, if they have the same ethnic and cultural background. There have been some incidents, however, of translators being used by organized crime networks to intimidate victims and witnesses and to provide false information to police. This makes some victims highly distrustful of translators. Moreover, interviews with victims are sometimes conducted in an uncontrolled environment, and in such cases the researcher has to gain the trust of the victims, or research subjects (see Sharpe 2009).

23.6 Survey Methods, Quantitative Approaches, and Public Perception

23.6.1 Survey Research

Survey research is important area of measurement in applied social research. The broad area of survey research encompasses any measurement procedures that involve asking questions to respondents. Surveys represent one of the most common types of quantitative, social science research, often associated with structured questionnaires. They are usually used to study broad research questions that involve public opinion; differences among large groups of people; or differences based on fixed characteristics, such as gender, race, education, and geographic location (Raymond 1996, p. 182).

Surveys are roughly divided into questionnaires and interviews. Questionnaires are typically paper-and-pencil instruments that the respondent completes; a set of questions or items in a written form that is self-administered. Interviews are completed by the interviewer based on the respondent's answers. Interview is a face-to-face situation in which an interviewer asks questions. Sometimes it is hard to distinguish between a questionnaire and an interview. Some people think that questionnaires always ask short closed-ended questions while interviews broad open-ended ones. There are, however, questionnaires with open-ended questions and there are closed-ended questions asked in an interview. Overall, a variety of types of items is used in questionnaires and interviews, the most common being fixed-alternative (closed item), scales, and open-ended items.

Interviews are a far more personal form of research than questionnaires. In the personal interview, the interviewer works directly with the respondent. Unlike with mail surveys, the interviewer has the opportunity to ask follow-up questions. And, interviews are generally easier for the respondent, especially if what is sought is opinions or impressions. However, interviews can be resource intensive as well as time-consuming. The interviewer is considered a part of the measurement instrument and in order to improve the reliability of the instrument the interviewers have to be trained.

Survey research has changed dramatically in the last 10 years. Nowadays, there are automated telephone surveys that use random dialing methods. There are computerized kiosks in public places that allow people to ask for input. However, not all of these methods are appropriate for organized crime research. Once the proper survey method has been selected, the researcher needs to construct the survey itself. A number of issues should be address here: the types of questions to be asked; decisions about question content; decisions about question wording; decisions about response format; and, questions regarding placement in the instrument.

23.6.2 Large-Scale Public Survey and the Person Perception Method

When conducting research on organized crime, the researcher might be interested in broad research questions that involve public opinions, such as the links between societal values and organized crime, or gender and organized crime. In 2006, I applied quantitative research methods

and conducted a cross-national survey, with 864 ethnic Albanian respondents from three countries with large ethnic Albanian population (Albania, Macedonia and Kosovo).[6] My study examined the larger societal factors that influence perceptions of violence and crime in Albania. A questionnaire was prepared for the purpose of the survey (structured interviews including vignettes[7]). It aimed at assessing the importance of the customary Kanun laws (foundation of the traditional Albanian culture) among ethnic Albanians, as well as their genuine knowledge about the laws. It also tried to study the alleged link between the respondents' actual knowledge of the ancient code and their perceptions of crime, honor, and violence.

As noted, survey research is not appropriate for studying the actual criminals or difficult populations; nevertheless, it is useful, for example, when studying the relevance of cultural codes among different societal groups. Since criminals often use "culture" as a justification for committing criminal acts, it is important to have an understanding of the degree to which so-called culture-specific behaviors are indeed acceptable and relevant in a given society. The draft questionnaire was prepared with the advice of experts in the field of anthropology, social psychology, social statistics, and criminology, and when finalized, it was distributed in Albania, Kosovo, and Macedonia. The questionnaire made use mainly of fixed-alternative (closed item) and scales, although some open-end items were included as well.

In fixed-alternative or closed item the respondent is presented with two or more alternatives provided by the interviewer or the questionnaire, see Box 23.5.

Box 23.5: Fixed-Alternative Item

Please indicate your gender:

☐ M ☐ F

A scale item is a special type of fixed-alternative item. A scale is a set of questionnaire or interview items to each of which an individual responds by expressing degrees of agreement or disagreement or some other type of response (Raymond 1996), see Box 23.6.

Box 23.6: Scale Item (Fix-Alternative)

The economic situation in my country has been very bad during 2000 - 2005

STRONGLY DISAGREE			UNDECIDED			STRONGLY AGREE
1	2	3	4	5	6	7

☐N/A

I don't see any future possibilities for personal economic growth in my country

STRONGLY DISAGREE			UNDECIDED			STRONGLY AGREE
1	2	3	4	5	6	7

☐N/A

Importantly, the questionnaire was developed in such a way that it elicits, from respondents' attitudes about specific criminal situations, the underlying causes of violent, and criminal acts. It also explores how ethnic Albanians perceive various offenders. The person perception method,[8] which captures the immediate impressions of people and behaviors, when used as a complementary technique alongside other data-collection methods, is proven to be highly effective (Hughes 1998; Hastrof et al. 1970; Finch, 1987). Previous research has documented its validity and utility (Collins & Brief 1995; Hughes 1998; Hastrof et al. 1970).

In my study, participants were asked to read a vignette and rate the behavior of the offender on 7-point scales, in terms of how justifiable versus unjustifiable or honorable versus dishonorable their actions were. Vignettes are used to elicit cultural norms, derived from the respondents'

[6] In Albania a total of 412 Albanians were interviewed. They are representative of the broader sample according to subculture (Geg/Tosk), area (rural/urban), age, gender and education. In Kosovo 206 ethnic Albanians were interviewed and in Macedonia—264 (Arsovska & Verduyn 2008).

[7] Short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond.

[8] In the typical person perception study, participants read a vignette. Then, participants' immediate impressions are assessed by having them rate people (in our case mainly the offenders) in the story on bipolar adjective scales.

attitudes and beliefs about a specific situation, as well as to explore the participants' ethical frameworks and moral codes (Finch 1987). Vignettes are especially useful for sensitive areas of inquiry (Neale 1999; Wade 1999), see Box 23.7.

Box 23.7: Vignettes and the Person Perception Method

Revenge Killing

A man was killed in broad daylight in Shkoder. The victim's brothers went immediately to search for the killer (E.H.). Fearing that he might be discovered, the killer E.H. knocked at the first door he found, asking for *besa (protection)*. The head of the house, who was in fact the father of the victim (but did not recognise the killer of his son), welcomed the visitor. When the victim's brothers returned, they recognised the killer E.H. The oldest brother (V.K.) immediately shot their guest—killer. After this incident V.K. was arrested.

The second killer (V.K.) explained that he was just revenging the murder of his younger brother according to the customary laws of his country. None of the parties had previous criminal record.

In your opinion, how justifiable the offender's (V.K.) behaviour is? CIRCLE the appropriate response:

TOTALLY UNJUSTIFIABLE			UNDECIDED			TOTALLY JUSTIFIABLE
1	2	3	4	5	6	7

In your opinion, is the offender honourable or dishonourable person? CIRCLE the appropriate response:

VERY BAD			UNDECIDED			VERY GOOD
1	2	3	4	5	6	7

Simple random sampling and stratified sampling are the best ways to ensure representative samples. But they are not always possible. The most common alternative that is used frequently for large-scale surveys is cluster sampling. In this method the population is divided into clusters or units. Cluster sampling is a technique used when "natural" groupings are evident in a statistical population. The total population is divided into clusters and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. Subsample of elements may also be selected within each of these groups. A common motivation for cluster sampling is to reduce the average cost per interview. Therefore, in my study I made a use of cluster sampling, structured questionnaires, as well as face-to-face interviews.

The sample was not representative of the whole ethnic Albanian population, since the aim was not to generalize the findings to the whole population, but to make comparisons between various subgroups. The key demographic variables for the selection of the sample were: country, subculture (Gheg/Tosk), area (rural/urban), age, gender, socio-economic status, and education.

Basing my assumptions on prior research,[9] I postulated that cultural norms promoted by the Kanun of Lek Dukagjini would play a significant role in shaping ethnic Albanians' perceptions of violence and crime. Since research shows that the Kanun is associated much more with the Albanian Gheg—a subculture populating North Albania, Kosovo, and parts of Macedonia—significant differences in perceptions between Albanian Ghegs and Tosks was anticipated. In general, the study hypothesized that older Ghegs from rural areas would know the Kanun better and would support more offenders who have committed "honor" crimes than other types of crime; whereas, younger, less educated Ghegs and Tosks, who only know the laws superficially, would justify various crimes, using the Kanun laws as a shield. It also postulated that the Tosks would be more supportive of profit-oriented crimes than honor crimes because of the

[9] A survey on the Kanun estimated that over 50% of teenagers polled respected the Kanun laws and were willing to take revenge (International Crisis Group 2000). Ismet Elezi conducted a survey which shows that today few people under 35 know the exact wording of the Kanun—yet many invoke it as an excuse to kill (Mortimer & Toader 2005).

greater Western influence in south Albania (in its proximity to Greece and Italy). Finally, the study hypothesized that there would be significant differences in perceptions between the respondents from the three countries due to different socio-political environments.

The questionnaire was available in Albanian and English and it was slightly adjusted to each country and subethnic community. This was a very comprehensive fifteen-page questionnaire; hence, the average length of an interview was approximately 40 min. Local experts were also given the opportunity to comment on the structure and type of questions. Before proceeding with the main research ($N = 726$), a pilot testing of the questionnaire was conducted ($N = 138$). The aim was to test the questionnaire prepared for this research in Albania, Kosovo, and Macedonia.

The results of the pilot were almost identical with those of the main study, attesting to the validity and reliability of the research. Before finalizing the questionnaire used during the main study, an expert panel was consulted regarding methodological issues, clarity of questions, sequence of questions, and types of items to be used. For the final research we followed a similar sampling logic. Individual respondents from various households were taken as sampling units. These are some of the steps taken in each country for selection of the sample:

Step 1: Selection of cities and villages
Step 2: Selection of neighborhoods/areas (according to economic standard)
Step 3: Selection of streets
Step 4: Selection of households
Step 5: Identifying individuals in the households according to specific profiles (primarily gender, age, and education).

As noted above, according to previous research done on the topic, we anticipated to find main differences in perceptions and beliefs between the Gegs from the north and the Tosks from the south of Albania. Although today in Albania there is not any clear dividing line between the two subcultures, the demographic makeup of the country shows that the Gegs live mainly in the north of the country and the Tosks live in the south of the country. Therefore, the sampling frame in Albania was divided into two main regions, north and south and two subcultures associated with the north and south, the Gegs and the Tosks. We also included two central districts populated by both Gegs and Tosks. To ensure diversity of the sample we decided to include districts from Central, Mountain, and Coastal Areas.

The researchers identified the neighborhoods in each rural and urban area. For example, in Kukes district/urban area the interviewer—following fixed criteria—identified two wealthier and two poorer neighborhoods. Around 13 interviews were conducted in poor area of this city and 13 in wealthier neighborhoods. Once the neighborhoods were identified, two streets were randomly selected in each neighborhood. The researchers randomly selected the first street as a starting point and every third house on that street was visited (according to the numbers of the houses, 1-3-6-9, etc.). Approximately four interviews were conducted on each street. No one was interviewed between street 1 and 3, or 3 and 6. The 13 people selected from the poor neighborhoods as well as the 13 people selected from the wealthier neighborhoods had equal number of men and women and were from various age groups (up to 25; 25–35; 35–45 and above 45). They had different educational backgrounds as well (primary school or less, secondary school, university or higher).

During a training session the interviewers were instructed regarding the profiles of the people they need for the study. If there was no person with such profile in the household, the interviewers were instructed to look at the second profile on their list until they met the quota (the interviewers had a prepared list with profiles. Example: male below 35 with higher education, male above 35 with higher education, female bellow 35 with higher education, etc.).

In case a person from a household rejected the interview, the interviewers moved to the next house identified on their list. When the interviewers visited a building they always took apartment one as the starting point. If no one in the apartment was willing to do the interview,

they went to apartment 3 then 6, 9, etc. until someone with the right profile was willing to answer the questionnaire.

23.6.3 Interview Techniques in Survey Research

As argued throughout this Chapter, interviews are among the most rewarding forms of measurement, no matter whether conducted in the context of participant observation or survey research. The interviewers play a very important part because the reliability of the measuring instrument depends largely on their performance. Their roles are multifaceted. One of the most important aspects of any interview is the training of the interviewers themselves since the quality of the study is in their hands.

Interviewers need to know more than simply how to conduct an interview. They should learn about the study and its social relevance. The interviewer has to be convinced of the importance of the research. This is important because the interviewer needs to motive the respondents to provide quality answers. Often, the interviewer has to work at the least desirable times (after dinner or on weekends) because the respondents are most available at these time periods.

The interviewers need to know enough about the survey method as well. Explaining the sampling logic, the interview process and the structure of the questionnaire is crucial. The interviewers should also understand why sampling is so important and should be aware of the ways that they can bias the results. In my research, the interviewers had to speak the dialect of the people they interviewed, so they were carefully selected. Being culture and gender sensitive in a research is very important. The principle investigator should also make sure that the interviewers know how to identify households and respondents, and how to avoid biasing the research results. The interviewer is the one who has to find the respondent, thus he or she must be able to locate addresses. Training sessions should be organized by the principle investigator in order to familiarize the interviewers with the study, the methodology, and the interviewing process. Due to time and budget constrains, the principle investigator cannot always supervise personally all of the interviews conducted. However, in order to assure the quality of the responses, the principle investigator should observe a subsample of interviews, or conduct follow-up assessments of interviews with the respondents.

It is important to note that each interview is unique, although there are a few common elements: an opening, where the interviewer gains entry; a middle game, the heart of the process; and an endgame, where the interviewer and respondent establish a sense of closure. One of the first things the interviewer must do is gain entry. Some factors, such as initial appearance, can enhance the prospects. The interviewer needs to dress professionally. The way the interviewer appears to the respondent has to communicate that the interviewer is trustworthy. Brief introduction, such as stating ones name and the name of the organization one represents is very important. If the interviewer has an identification badge this should be visible to the potential respondent. The interviewer must appear as legitimate as possible.

If the potential respondent is willing to take the interview to the next level, the interviewer should explain briefly the aims and the objectives of the study. The respondent usually does not need to know all the details of the study. There should be two-sentence description of the study memorized. The interviewer should spend more time on assuring the respondent that they are interviewed confidentially, and that their participation is voluntary. If the respondent is comfortable then the interviewer can begin with the interview.

In my quantitative survey study, most of the questions were structured in such a way that the respondent could answer them without the help of the interviewer (except if they were illiterate or needed an assistance). The task of the interviewer was to clarify unclear and confusing points in order to ensure the quality of the answers. Respondents often raise concerns that are not anticipated and the interviewer must be able to respond candidly and informatively. The

reason for developing this type of structured questionnaire was to reduce the bias and the influences of the different interviewers.

Similar to interviews in the context of focus groups or participant observation, the use of recorders is not very desirable during survey interviews. Respondents are often uncomfortable when they know their answers will be recorded word-for-word. They may strain to only say things in a socially acceptable way. Taking brief notes is always safer. Finally, when the interviewer has finished with the interview, he or she needs to bring it to closure, thank the respondent, and tell them about the future of the study. Immediately after leaving the interviewer should write down any notes about how the interview went. In my study, the interviewers were instructed to describe the interior of the house, following certain criteria, in order to assess the relative wealth of the family.

23.7 Unobtrusive Measures, Secondary Sources, and Document Analysis

Unobtrusive measures are measures that do not require the researcher to intrude in the research context. They are contrasted with interviews and questionnaires, in that they try to find indirect ways to obtain data. For example, participant observations require that the researcher be physically present. This can lead the respondents to alter their behavior in order to look good in the eyes of the researcher. A questionnaire is an interruption in the natural stream of behavior. Respondents can get tired of filling out a survey or aggrieved by the questions asked.

Unobtrusive measurement presumably reduces the biases that result from the intrusion of the researcher. However, unobtrusive measures reduce the degree the researcher has control over the type of data collected. Available materials also known as secondary data, for example, refer to data that have been collected by others. The data available for secondary analysis include: census data, registration data, newspapers and periodicals, institutional records, personal documents and archives, crime records, standardized testing data, economic data, and consumer data. In our modern world there is mass of data that is routinely collected by various organizations. Many research projects store their raw data in electronic form in computer archives so that others can also analyze it.

In my research extensive use of existing datasets like the World Values Survey (WVS), conducted by the World Values Survey Association was made.[10] This is considered the world's most comprehensive investigation of political and socio-cultural change. Moreover, the findings of the public surveys conducted by the Internal Security Reform Sector in Kosovo (2006)[11] and the International Commission in the Balkans (2005)[12] were also used in the "Albanian criminality" research as a point of comparison. Murder numbers from the World Health Organisation, social statistics from the United Nations Development Programme, statistics on transnational crime from UNODC and INTERPOL, as well as reports from organizations working on the Balkan region, such as SEESAC and OSCE were utilized too. Often these sources are overlooked by researcher but can be very useful if combined with other data sources.

[10] More information on the World Values Surveys (WVS) can be found on www.worldvaluessurvey.org

[11] ISSR studied perception of threat in Kosovo. ISSR team engaged in a number of activities including: Public opinion surveys completed by approximately 800 homes; Comment boxes located in every municipality building, or in cultural centers in areas inhabited by minority groups; Recorded commentaries from citizens; Emails sent directly to ISSR address; Phone calls made to ISSR call line; Questionnaires completed by individual citizens. Public debates among Kosovo's key security figures regarding issues outlined in the ISSR report; Public meetings facilitated by OSCE across all 30 of Kosovo's municipalities and in 2 of Kosovo's "Pilot" municipalities.

[12] The Survey was conducted in November–December 2004 in Macedonia, Albania, Serbia and Montenegro (including Kosovo) and Bosnia and Herzegovina. The data was analyzed by BBSS Gallup International.

23.8 Concluding Remarks

This chapter elaborated on a variety of research methodologies, sampling strategies, interviewing techniques, and methodological challenges in the study of organized crime. Organized crime is a difficult topic for researchers; safety, uncertainty, and accessibility are some of the main reasons for this ambiguity. Prior experiences indicate that truly interdisciplinary and multi-dimensional research methodology must be developed in this area to ensure the quality of future studies. This chapter explains why the method of multiple triangulation can be an appropriate method for researching organized crime. This method is used to indicate that more than two methods are applied in a study of the same phenomenon to double or triple-check the results.

Moreover, some advantages and disadvantages of the purposive and snowball sampling techniques were identified in this chapter. The personal reflexive account of the fieldwork experience provided here highlighted good and bad practices in conducting ethnographic research on organized crime. The chapter also argued that researchers must be familiar with proper interviewing techniques to establish the facts of the case as accurately as possible, remain objective, and avoid becoming sensational. Professionalism, confidentiality and transparency are very important too not only for gaining the interviewee's trust, but also for ensuring the safety of the researcher. Finally, researchers working in this field should make extensive use of secondary data in order to compare and contrast their empirical findings.

References

Adler, P. (1985). *Wheeling and dealing: An ethnography of an upper level drug dealing and smuggling community*. New York: Colombia University Press.

Arsovska, J. (2007). Code of conduct: Understanding Albanian organised crime. *Jane's Intelligence Review, 19*(8), 46–49.

Arsovska, J. (2008). Interviewing serious offenders: Ms Egghead meets Mr Gumshoe. *Trends in Organised Crime, 11*(1), 42–58.

Arsovska, J., & Craig, M. (2006). 'Honourable' behaviour and the conceptualisation of violence in ethnic-based organised crime groups: An examination of the Albanian Kanun and the code of the Chinese triads. *Global Crime, 7*(2), 214–246.

Arsovska, J., & Lantsman, L. (2010). Magic touch: West African trafficking networks. *Jane's Intelligence Review, 22*(10), 38–41.

Arsovska, P., & Verduyn, P. (2008). Globalisation, conduct norms and "culture conflict": Perceptions of violence and crime in an ethnic Albanian context. *British Journal of Criminology, 48*(2), 226–246.

Baulch, B., & Scott, L. (Eds.). (2006). *Report on CPRC workshop on panel surveys and life history methods*. London: Workshop held at the Overseas Development Institute.

Becker, H. S. (1965). *Sociological work*. Chicago: Aldine.

Berg, S. (1988). Snowball sampling. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (vol. 8). New York: Wiley.

Blok, A. (1981). *The Mafia of a Sicilian village: A study of violent peasant entrepreneurs 1860–1960*. Long Grove: Waveland Press.

Booth D., Holland J., Hentschel J., Lanjouw P., & Hebert A. (1998). Participation and combined methods in African poverty assessment: reviewing the Agenda. DFID, Social Development Division, African Division. www.dfid.gov.uk/Pubs/files/sddafpov.pdf.

Bourgois, P. (1995a). *In search of respect: Selling crack in El Barrio*. Cambridge: Cambridge University Press.

Bourgois, P. I. (1995b). *In search of respect: Selling crack in El Barrio*. New York: Cambridge University Press.

Chambliss, W. (1988). *On the take: From petty crooks to president*. Bloomington: Indiana University Press.

Collins, B. E., & Brief, D. E. (1995). Using person-perception vignette methodologies to uncover the symbolic meanings of teacher behaviors in the Milgram paradigm. *Journal of Social Issues, 51*, 89–106.

Emerson, R., Fretz, R., & Shaw, L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing, 26*, 790–797.

Finch, J. (1987). The vignette technique in survey research. *Sociology, 21*, 105–114. Finch (1987).

Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory. Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Hastrof, A. A., Schneider, D. J., & Polefka, J. (1970). *Person perception*. Boston: Addison-Wesley.

Hendricks, V. M., & Blanken, P. (1992). Snowball sampling: Theoretical and practical considerations. *In Snowball Sampling: A Pilot Network, 2*(1), 1–18.

Hobbs, D. (1995). *Bad business: Professional crime in Modern Britain*. Oxford: Oxford University Press.

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health & Illness, 20*(3), 381–400.

Ianni, F. (1972). *A family business: Kinship and social control in organized crime*. New York: Russell Sage.

Ianni, F. A. J. (1974). *Black Mafia: Ethnic succession in organized crime*. New York: Simon and Schuster.

International Crisis Group. (2000). *Albania State of the Nation*, 1 March. www.asyl.net/Magazin/Docs/Docs09/I5806alb.htm.

Jakob, A. (2001). Reflections on a typology of conceptualising 'uncertainty' in the context of employment biographies. *Forum for Qualitative Social Research 2*(1).

Krug, E. G., Dahlberg, L., Mercy, J. A., Zwi, A., & Lozano, R. (Eds.). (2002). *World report on violence and health*. Geneva: World Health Organization.

Lofland, J., & Lofland, L. H. (1984). *Analyzing social settings*. Belmont: Wadsworth.

Mortimer, M., & Toader, A. (2005). Blood feuds blight Albanian lives. *BBC World Service*, 23 September.

Neale, B. (1999). Post divorce childhoods. Available online at http://www.leeds.ac.uk/family.

Paoli, L. (2003). *Mafia brotherhoods. Organized crime: Italian style*. New York: Oxford University Press.

Pistone, J. D., & Woodley, R. (1999). *Donnie Brasco: My undercover life in the Mafia*, London: Hodder & Stoughton.

Polsky, N. (1969). *Hustlers, beats and others*. New York: Anchor Books.

Raymond, M. (1996). *A research made simple: A handbook for social workers*. London: Sage Publications.

Rubinstein J., & Reuter, P. (1977). *Numbers: The routine racket.* New York: Policy Sciences Ctr, Inc United States Publication.

Reuter, P., & Rubinstein J. (1978). Facts, fancy and organized crime. *The Public Interest*, *53*, 45–67. [Reprinted in *The Public Interest on Crime and Justice*].

Sharpe, K. (2009). Sad, bad and (sometimes) dangerous to know. In King D. Roy & Emma (Eds.), *Doing research on crime and justice.* London: Oxford University Press.

Spreen, M. (1992). Rare populations, hidden populations and link-tracing designs: what and why? *Bulletin Methodologie Sociologique, 36*, 34–58.

Taylor, L. (1984). *In the underworld*. London: Counterpoint.

Thomson, S. (1997). Adaptive sampling in behavioural surveys. *NIDA Research Monograph, 167*, 296–319.

Van Meter, K. M. (1990). Methodological and design issues: techniques for assessing the representatives of snowball samples. *National Institute on Drug Abuse: Research Monograph Series Mono, 98*, 31–43.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. London: Sage.

Wade, A. (1999). New childhoods? children and coparenting after divorce. Available online at http://www.leeds.ac.uk/family.

Williams, T. (1990). *The cocaine kids: The inside story of a teenage drug ring*. London: Bloomsbury.

Wright, R. T., & Decker, S. (1994). *Burglars on the job: Streetlife and residential break-ins*. Boston: Northeastern University Press.

Zaitch, D. (2002). *Trafficking cocaine: Colombian drug entrepreneurs in the Netherlands*. The Hague: Kluwer International.

What Survey Modes are Most Effective in Eliciting Self-Reports of Criminal or Delinquent Behavior?

24

Gary Kleck and Kelly Roberts

The most important research strategy for testing criminological theories, by far, is the self-report survey, in which respondents (Rs) are invited to report criminal or delinquent acts they have committed (Thornberry and Krohn 2000; Kleck et al. 2006). These tests, however, are meaningful only to the extent that the surveys yield valid measures of those behaviors, which require effective methods for minimizing errors in response to the questions about criminal behaviors.

The most important threat to the validity of self-report measurements is probably social desirability bias, the tendency of Rs to present themselves in a socially desirable light. This causes Rs to understate socially disapproved behaviors like criminal or delinquent acts (and to overstate socially approved behaviors like voting, church attendance, and the like). While false positive responses to questions about disapproved behaviors (i.e., admitting crimes that the R had not in fact committed) do occur, they are far less common than false negatives (denying crimes the R had committed). For example, validity tests of drug use based on urine tests consistently indicate that false positives are rare and greatly outnumbered by false negatives (Harrison 1995). Likewise, in a unique study of self-reports of a wide array of criminal acts, Clark and Tifft (1966) found that false negatives outnumbered false positives by a margin of three to one.

Further, the main source of inaccurate responses to self-report questions is deliberate misreporting rather than recall failure or misunderstanding of questions. In their review of the literature, Tourangeau and Yan (2007, pp. 859, 876) concluded that: "The survey evidence also indicates that misreporting on sensitive topics is a more or less motivated process in which respondents edit the information they report to avoid embarrassing themselves in the presence of an interviewer or to avoid repercussions from third parties." Thus, if false negative responses are overwhelmingly the dominant kind of error in self-report surveys, better methods should yield higher rates of admission of illegal behaviors.

24.1 Survey Modes

There are many features of surveys that influence Rs' willingness to report such behaviors, including the wording of questions or the presence of third parties during interviews, but the focus of this chapter is on the survey modes used, i.e., the modes of communication by which

G. Kleck (✉) · K. Roberts
College of Criminology and Criminal Justice,
Florida State University, Tallahassee,
FL 32306-1127, USA
e-mail: gkleck@fsu.edu

K. Roberts
e-mail: kkr03c@fsu.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_24, © Springer Science+Business Media New York 2012

questions are conveyed to Rs and by which Rs convey their responses back to the researchers.

The body of evidence concerning the effects of different modes on response error is constantly evolving, as technological changes alter the available and feasible modes of communication. In particular, the increasing availability, smaller size, and lower cost of computers has increased their use in survey work, changing even the ways in which traditional survey modes were applied. Interviewers asking questions over the telephone or in person can now be assisted in their work by computers, reading questions off a computer screen and recording answers on the computer using a keyboard and mouse rather than pencil and paper answer form. Some surveys can now be entirely or partially conducted by computers over the telephone. The development of the Internet has produced a growing body of Web surveys, but little methodological research on the effects of this new mode on response errors and other survey errors has been done.

Some survey modes involve human interviewers asking questions, while others are self-administered by the R. Among those that are self-administered, some involve questions being presented to Rs by a computer, while others employ a paper questionnaire. Among those employing a human interviewer, some involve interviewers conveying their questions in a face-to-face contact, often in the R's home, while others rely on communication via telephone. In modes in which computers pose the questions, they can appear on a computer screen, or prerecorded questions may be played to the R, usually through headphones, thereby eliminating problems due to the R's literacy or vision limitations. The computer that is used may be the R's own home computer (as in most Web surveys) or may be a laptop or similar computer brought to the R by an interviewer. Rs may speak their responses, record them on paper, or record them using a computer's keyboard or mouse. Further, these elements may be combined in various ways, as in surveys in which the interviewer reads the questions but the R records his or her responses on a paper answer form or on a computer.

Because survey modes evolve as rapidly as communication technologies, methodological research has yet to catch up with some of the more recent developments. Thus, we have fewer experimental studies assessing response errors in Web surveys or Interactive Voice Response surveys, to name two of the newer survey modes. Thus, conclusions must often rely on a handful of studies that bear on a specific pair of randomly assigned modes, and are correspondingly tentative. Indeed, many pairs of survey modes have never, to our knowledge, been experimentally compared.

This review covers the following modes:

- *Face-to-face interviewing* (aka personal interviewing) (FTF). A human interviewer in the R's presence asks questions orally, reading from a printed questionnaire, and R answers orally.
- *Face-to-face interviewing with paper answer form* (F-P). A human interviewer in the R's presence asks questions orally, reading from a printed questionnaire, and R privately records answers on a paper form.
- *Telephone interviewing*, typically computer-assisted (Tel). A human interviewer asks questions over the telephone, and the R answers orally.
- *Computer-assisted self-interviewing* (CASI). Research staff provides a computer to the R and give instructions on its use, the R reads questions off a computer screen, and records answers using the computer's keyboard or mouse.
- *Audio computer-assisted self-interviewing* (ACASI). Research staff provides a computer to the R and give instructions on its use, the R listens to recorded questions delivered through earphones, and records answers using the computer's keyboard or mouse.
- *Telephone audio computer-assisted self-interviewing* (TACASI). A computer delivers questions over the telephone (usually after a human interviewer has established rapport and asked nonsensitive questions), and the R provides answers by pressing telephone keypad numbers and/or speaking responses.

- *Paper-and-pencil questionnaire* (aka self-administered questionnaire) (PAP). Paper questionnaires are distributed to Rs by research staff, often in group settings, the Rs read questions, and record answers on paper answer forms or the questionnaire itself.
- *Mail-out/mail-in surveying* (Mail). Paper questionnaires are mailed to Rs, who record answers on paper answer forms or the questionnaire itself, which are mailed back.
- *Web* (aka Internet). Rs are directed to a website with a computerized version of the questionnaire, read questions off a computer monitor (or hear them via headphones or computer speakers), and provide answers using the computer's keyboard or mouse.

This is an extensive but not exhaustive list of survey modes. It is not possible to make meaningful assessments of unique modes, used in only a single experimental study, so comparisons involving unique, mixed or "hybrid" modes were not included in this review.

Furthermore, all of the above mentioned modes are described and discussed in more details through the different chapters of this handbook, while also being compared for their advantages, disadvantages and related costs (see for example Chap. 11 by Albaum and Smith, as well as Chap. 21 by Shine and Dulisse).

24.2 Some Aspects of Modes that May Affect Willingness of Rs to Admit Criminal Acts

Why should variations in survey modes affect whether Rs report criminal or delinquent acts? A number of differences underlying mode variations could be hypothesized to influence responses to sensitive questions.

Sense of privacy. Rs may be more willing to respond honestly to questions about criminal acts if they can be confident that their answers are being provided in a private context in which the answers cannot be linked to them as individuals. For example, providing responses in a group context, where one R's answers will be submitted along with many other Rs' answers, gives Rs a greater sense of privacy because they know that their responses cannot be individually linked with them. This is an advantage enjoyed by PAP surveys in which questionnaires are distributed in classroom and other group settings. In contrast, Rs interviewed face-to-face in their homes know that their identities are known to the researchers, while those interviewed via telephone may suspect the same—a reasonable suspicion given the possibility of deriving names and addresses from telephone numbers using the reverse record checks available via online services like the White Pages. Likewise, Rs may feel greater privacy when they respond to a computer rather than a human, even though they presumably are aware that their computer-recorded responses will later be accessed by human researchers. Conversely, Rs who are interviewed when there are third parties present (e.g. parents present when adolescents are interviewed in their homes) probably feel less privacy than Rs who provide their answers while alone.

Embarrassment of directly speaking responses to interviewers. Directly speaking admissions of criminal acts to a human interviewer, either face-to-face or over the telephone, may cause Rs to feel embarrassed and may thereby discourage admissions. In contrast, recording admissions on a paper form or on a computer avoids the embarrassment of admitting a crime directly to another human being, even if a human asked the question in the first place. This may play a large role in the superiority of self-administered survey modes over interviewer-administered modes in eliciting admissions of socially undesirable behaviors.

Trust in surveyors. Personal contact between the R and an interviewer can help establish the R's trust in the researchers. Rs may be less suspicious of persons willing to "show themselves" to Rs than of faceless strangers who contact them via telephone or mail. On this dimension, face-to-face personal interviews may have an advantage over telephone interviews, mail surveys, or Web surveys, even though they also commonly suffer from the disadvantage of requesting Rs to speak answers to interviewers.

24.3 Scope of the Review

Our review covered all published English-language studies that used random assignment of survey modes and measured reporting of criminal or delinquent behavior (including illegal drug use), published since 1985. Random assignment of modes insures that the effects of mode are not likely to be confounded with other influences on responses, since any other factors that might affect the reporting of criminal behavior (e.g. attributes of the R, of the interviewer, or of the context in which Rs were interviewed) will be uncorrelated with the mode. Within each study, we identified every individual pairing of survey modes that was compared, and each type of criminal behavior to which the comparison was applied. Each paired comparison of modes with regard to a specific type of criminal behavior was treated as an individual finding. There were 326 such pairings of randomly assigned modes for which the frequency of some type of self-reported criminal behavior was noted.

Qualifying studies had to compare some fairly widely used survey modes, since there was no need to summarize the evidence concerning rarely used modes that had been assessed in a single study. Thus, while dozens of mixed or hybrid modes have been used in surveys asking questions about criminal behavior, there is no significant body of evidence to be summarized concerning any one of them.

In order to qualify for the review, studies had to examine behaviors that are criminal when they are committed by adults. Works that examined only status offenses (acts unlawful only for juveniles), attitudes related to deviant behavior, alcohol use, tobacco use, victimization, consequences of illicit drug use (as opposed to drug use itself), and experiences where it was ambiguous as to whether victimization or offending behaviors were being asked about were excluded. Studies that created indices in which legal and illegal behaviors were combined were likewise excluded. Also, studies that examined the use of certain illicit substances were not included in this review because they were not comparable with other studies examining illicit drug use. For instance, a study by Li et al. (2003) compared the use of betel quid by survey mode. Due to the nature of the drug use, this study did not meet this criterion for inclusion.

This review does not cover studies of values, attitudes, or personality traits purportedly related to illegal behavior, legal but disapproved behaviors such as alcohol or cigarette use, or experiences of criminal victimization. It was felt that some of the forces influencing Rs to conceal illegal behaviors are distinct, and do not apply to behaviors or attitudes that are merely socially disapproved. Admitting to behaviors such as excessive drinking may be embarrassing, but confessing to crimes may, at least in the minds of Rs, entail risks of legal punishment, perhaps even imprisonment.

Qualifying studies also had to either report the results of pairwise comparisons or provide prevalence estimates for each illegal behavior by mode such that we could perform difference-of-proportions tests ourselves. Some studies did not report pairwise comparisons but instead conducted only an analysis of variance across modes, which made it impossible to determine the rank order of the modes. Additionally, studies that reported effects for interactions (e.g., mode by privacy effects) were not included unless prevalence estimates were reported by mode alone.

24.4 Study Acquisition Methods

The articles included in this review were located in a search of the following databases: Web of Knowledge, ArticleFirst, Criminal Justice Periodicals Index (CJPI), Cambridge Scientific Abstracts (Subject Area = Social Sciences), National Criminal Justice Reference Service, and WorldCat. These databases covered a wide array of disciplines. In each database, the terms "survey," "methods," "modes," "administration," "interview," "self-report," "questionnaire," "comparison," "sensitive behaviors OR questions OR issues," "illegal OR illicit OR criminal behaviors OR acts," "drug OR substance use," and "violence" were used in the searches. The fields searched, when available in

a given database, were "abstract," "title," "keyword," "subject," "topic," and "first page." Reference lists of retrieved articles were also used to locate other studies which met the review criteria.

24.5 Findings

Our main findings appear in two tables. First, the Appendix table lists, in chronological order of publication, every study we reviewed, and displays each individual finding of those studies. An individual finding is a pairwise comparison of the rate of admission of each particular type of criminal behavior, between two survey modes. Thus, a study that compared a single pair of modes regarding three kinds of criminal behavior would have three findings, while a study that compared two pairs of modes regarding four kinds of criminal behavior would have eight findings. The abbreviations used to denote each survey mode are explained in the endnotes to the table. In total 27 studies reviewed containing a total of 326 distinct findings.

Each entry in the Appendix table provides the study's authors and date of publication, describes the sample of persons surveyed, and states the sample size. It is also noted whether this sample was selected using probability sampling techniques (P) or was a nonprobability sample (NP). The sample size is significant in that in many cases it helps explain why differences between modes were not statistically significant. Each line within the study's entry describes an individual finding, indicating the pair of survey modes that were compared, the illegal behavior that researchers asked about, and summarizes the finding. The last column displays which of the two modes had a higher rate of admissions of the illegal behavior asked about, and the statistical significance of the difference between the rates. The last column displays only an empty pair of parentheses, it indicates that the authors merely reported there was no significant difference between the modes, but did not report which mode had the higher (albeit nonsignificant) admission rate or what significance level was used.

Many of these findings pertain to the reporting of illegal drug use. We separately counted findings pertaining to (1) current or recent use (designated R) and findings pertaining to (2) lifetime use (designated L) as distinct findings, based on the belief that questions pertaining to recent or current use are considerably more sensitive and likely to lead to false denials than questions about lifetime use that might have occurred many years in the past.

This mass of data obviously does not lend itself to any simple interpretation, so Table 24.1 was created to summarize the individual findings shown in the Appendix table so as to make them more comprehensible. The findings were first sorted by the pair of modes compared, and then these pairs were sequenced in accordance with our initial judgments as to the effectiveness of the modes being compared. Thus, Table 24.1 is divided into panels, each of which summarizes the findings concerning a specific pair of survey modes. The panels appear in rough order of the effectiveness of the modes in eliciting admissions of criminal behavior, starting with the less effective modes in the first part of the table and progressing to the generally more effective modes in the later panels.

The first column of Table 24.1 reports the modes compared, presenting first the mode predicted to be more effective, i.e., to yield higher rates of admission of criminal acts. Thus, this column states a prediction concerning which was the putatively better mode, i.e., the mode expected to yield higher admission rates. This was based on our tentative conclusions from an initial reading of the literature, which yielded a provisional ranking of modes, subject to revision if the more systematic counting of findings did not support some of the ranks. This provisional ranking held up fairly well, in that few studies yielded more than one or two significant differences favoring the mode predicted to be worse in eliciting admissions.

Table 24.1 Summary of findings, sorted by mode pairs compared (sequenced from worst to best modes)

Modes compared	Study	Findings	Differences in predicted direction	Tentative conclusion
FTF > Tel	Johnson et al. (1989)	7/0/1 (1)	8 of 8	
	Aquilino (1994)	0/0/7 (5)	5 of 7	
Total		7/0/8 (6)	13 of 15	FTF is better
F-P > FTF	Schober et al. (1992)	5/0/1 (1)	6 of 6	
	Turner et al. (1992)	1/0/1 (5)	6 of 6	
	Aquilino (1994)	1/0/6 (5)	6 of 7	
	Tourangeau et al. (1997)	0/0/10 (?)	?	
	Rogers et al. (1998)	4/0/2 (?)	4+ of 6	
Total		11/0/24 (11+)	22+ of 25	F-P is better
F-P > Tel	Aquilino (1994)	4/0/3 (1)	5 of 7	F-P is better
PAP > F-P	Tourangeau et al. (1997)	0/0/10 (?)	?	No difference
PAP > FTF	Tourangeau et al. (1997)	0/0/10 (?)	?	No difference
CASI > PAP	O'Reilly et al. (1994)	2/0/7 (5)	7 of 9	
	Tourangeau et al. (1997)	0/0/10 (?)	?	
	Beebe et al. (1998)	0/2/2 (1)	1 of 4	
	Wright et al. (1998)	0/0/1 (1)	1 of 1	
	Brenek et al. (2006)	0/0/8 (7)	7 of 8	
	Hamby et al. (2006)	1/1/1 (1)	2 of 3	
Total		3/3/29 (15+)	18+ of 25	CASI may be slightly better
ACASI > PAP	O'Reilly et al. (1994)	0/0/9 (7)	7 of 9	
	Turner et al. (1998)	0/0/5 (5)	5 of 5	
Total		0/0/14 (12)	12 of 14	ACASI is better
CASI > F-P	Tourangeau et al. (1997)	0/0/10 (?)	?	
	Supple et al. (1999)	3/0/2 (2)	5 of 5	
Total		3/0/12	5 of 5	CASI is better
CASI > FTF	Tourangeau and Smith (1996)	1/0/5 (3)	4 of 6	
	Tourangeau et al. (1997)	0/0/10 (?)	?	
Total		1/5/15 (3+)	4 of 6	CASI is slightly better
ACASI > FTF	Tourangeau and Smith (1996)	3/0/3 (3)	6 of 6	
	Newman et al. (2002)	1/0/1 (1)	2 of 2	
Total		4/0/4 (4)	8 of 8	ACASI is better

(continued)

Table 24.1 (continued)

Modes compared	Study	Findings	Differences in predicted direction	Tentative conclusion
ACASI > CASI	O'Reilly et al. (1994)	0/0/9 (2)	2 of 9	
	Tourangeau and Smith (1996)	2/0/4 (4)	6 of 6	
	Couper et al. (2003)	0/0/2 (0)	0 of 2	
Total		2/0/14	8 of 17	ACASI is slightly better
Mail > Tel	Bason (2000)	0/0/4 (1)	1 of 4	Too little information
Web > Tel	Bason (2000)	0/1/3 (1)	1 of 4	Too little information
TACASI > Tel	Bason (2000)	0/1/3 (1)	1 of 4	
	Gribble et al. (2000)	4/0/7 (3)	7 of 11	
	Corkrey and Parkinson (2002)	2/0/7 (6)	8 of 9	
	Turner et al. (2005)	11/0/0 (0)	11 of 11	
Total		17/1/17 (10)	27 of 35	TACASI is better
Web > PAP	Wang et al. (2005)	1/0/4 (4)	5 of 5	
	Lucia et al. (2007)	0/0/6 (2)	2 of 6	
	Van de Looij-Jansen (2008)	1/0/1 (?)	?	
	Eaton et al. (2010)	4/0/0/(0)	4 of 4	
Total		6/0/11 (6)	12 of 17	Web is better
Mail > TACASI	Bason (2000)	0/0/4 (2)	2 of 4	
	Knapp and Kirk (2003)	0/0/1 (1)	1 of 1	
Total		0/0/5 (3)	3 of 5	No difference
Web > Mail	Bason (2000)	0/0/4 (3)	3 of 4	
	McCabe et al. (2002)	0/1/0 (0)	0 of 1	
	Knapp and Kirk (2003)	0/0/1 (1)	1 of 1	
	McCabe (2004)	2/2/12 (4)	6 of 16	
Total		2/3/14 (8)	10 of 22	No difference
Web > TACASI	Bason (2000)	0/0/4 (2)	2 of 4	
	Knapp and Kirk (2003)	0/0/1 (1)	1 of 1	
Total		0/0/5 (3)	3 of 5	No difference

To illustrate how to interpret the table, consider the first panel. The face-to-face mode (FTF) was predicted to be more effective than the telephone mode (Tel). The second column indicates which study is being described. The third column summarizes the number of findings of each possible type: (1) significant and in the predicted direction (i.e. the mode predicted to be better did obtain a higher admission rate), (2) significant and in the opposite direction from that predicted, and (3) not significantly different. The number in parentheses states the number of findings in the predicted direction, among the *non*significant findings. Thus, for the Johnson et al. (1989) study, there were seven findings indicating that the FTF mode was significantly better, as predicted, zero

Telephone/ CATI	< Face-to-Face Interviewing (FTF)	< Face-to-Face Interviewing with paper answer forms (F-P)	< Paper-and-Pencil (PAP)	< Computer-assisted Self-interviewing (CASI)	< Audio computer-assisted self-interviewing (ACASI)	< Telephone audio computer-assisted self-interviewing; (TACASI) Web; Mail[a]

Fig. 24.1 A tentative ranking of survey modes as to their ability to elicit admissions of criminal or delinquent behavior (reading from *left* to *right*, modes go from worst to best)

findings indicating the telephone mode was significantly better, and one finding of no significant difference. The "1" in parenthesis indicates that this one nonsignificant finding was in the predicted direction. It was felt that when cumulating many findings over multiple studies, a consistent pattern of findings in the predicted direction constituted meaningful support for the prediction, even if some of the findings, taken one at a time, were not statistically significant (often due to small sample sizes). Based on this reasoning, the fourth column reports the share of findings that were in the predicted direction, regardless of statistical significance, among those where a direction was reported by the original authors or could otherwise be determined. Some authors reported only that differences were not significant, without reporting the direction of difference or the admission rates. In these cases, a question mark was entered in the table where the number of findings in the predicted direction would be noted.

When summarizing the total number of findings in the predicted direction, across all studies comparing a particular pair of modes, the number of findings in the predicted direction sometimes could not be determined precisely because some authors did not report the direction of nonsignificant differences. Thus, if there were 11 findings with direction reported that were in the predicted direction, plus some others without direction reported, this means there were 11 or more findings in the predicted direction. This would be designated 11+, meaning "11 or more."

The last column reports our tentative conclusion as to which of the two compared survey modes appeared to be more effective. When fewer than 10 comparisons of a given pair of modes had been made in the set of reviewed studies, we concluded that there was too little information to draw even a tentative conclusion. When the findings were evenly or nearly evenly divided as to which mode was more effective, we tentatively concluded that there was no difference.

Finally, Fig. 24.1 visually summarizes the information from Table 24.1, displaying the rank ordering of survey modes as to their effectiveness in eliciting self-reports of criminal or delinquent behavior. As one reads from left to right, the modes shown are increasingly effective. It should be evident from Table 24.1 that many possible pairwise comparisons of modes have never been directly studied, so there often was no direct empirical foundation for saying mode A is better than C. We could, however, apply the transitive principle to those pairwise comparisons that have been studied, to draw conclusions about pairs not directly compared. That is, if some studies showed mode A to be better than mode B, and other studies showed mode B to be better than mode C, we felt justified in concluding that mode A was better than mode C, even though no study had ever directly compared modes A and C.

The body of experimental evidence available to date suggests that conventional telephone interviewing and face-to-face interviewing are the least effective survey modes in getting Rs to admit to unlawful behaviors, while the most effective modes are Web, mail, and TACASI. The latter three modes appear to be about equally effective, or at least, so far, there is too little evidence to convincingly separate them.

24.6 Conclusions

Valid measurement of delinquent and criminal behaviors is certainly crucial to survey research on these topics, but it is not the only consideration in doing good surveys. Good researchers try to reduce total survey error, which includes

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
Johnson et al. (1989)	P	780	FTF/Tel	Marijuana use, L	FTF, >.05
	U of Kentucky students, 1987		FTF/Tel	Cocaine use, L	FTF, <.01
			FTF/Tel	Tranquilizer use, L	FTF, <.001
			FTF/Tel	Amphetamine use, L	FTF, <.05
			FTF/Tel	Downer use, L	FTF, <.001
			FTF/Tel	Other pill use, L	FTF, <.001
			FTF/Tel	Hallucinogen use, L	FTF, <.01
			FTF/Tel	Heroin use, L	FTF, <.01
Schober et al. (1992)	P	9,308	F-P/FTF	Cocaine use, R	F-P, <.001
	NLSY, 1988		F-P/FTF	Cocaine use, past year	F-P, <.001
			F-P/FTF	Cocaine use, L	F-P, <.001
			F-P/FTF	Marijuana use, R	F-P, <.001
			F-P/FTF	Marijuana use, past year	F-P, <.001
			F-P/FTF	Marijuana use, L	F-P, > .10
Turner et al. (1992)	P	3,284	F-P/FTF	Marijuana use, R	F-P, .039
	NHSDA, 1990(HH residents in 33 US metro areas)		F-P/FTF	Marijuana use, past year	F-P, .18
			F-P/FTF	Marijuana use, L	F-P, >.20
			F-P/FTF	Cocaine use, R	F-P, .081
			F-P/FTF	Cocaine use, past year	F-P, .185
			F-P/FTF	Cocaine use, L	F-P, >.20
Aquilino (1994)	P	1,508	F-P/FTF	Marijuana use, L	F-P, >.10
	HH residents in 37 Largest US SMSAs, 1991		F-P/FTF	Marijuana use, past year	F-P, <.05
			F-P/FTF	Marijuana use, R	F-P, >.10
			F-P/FTF	Cocaine use,L	F-P, <.10
			F-P/FTF	Cocaine use, past year	F-P, >.10
			F-P/FTF	Cocaine use, R	()
			F-P/FTF	Crack use, L	F-P, >.10
		1,499	F-P/Tel	Marijuana use, L	F-P, >.10
			F-P/Tel	Marijuana use, past year	F-P, <.001
			F-P/Tel	Marijuana use, R	F-P, <.05
			F-P/Tel	Cocaine use, L	F-P, <.01

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			F-P/Tel	Cocaine use, past year	()
			F-P/Tel	Cocaine use, R	()
			F-P/Tel	Crack use, L	F-P, <.05
		1,489	FTF/Tel	Marijuana use, L	FTF, >.10
			FTF/Tel	Marijuana use, past year	FTF, <.10
			FTF/Tel	Marijuana use, R	FTF, >.10
			FTF/Tel	Cocaine use, L	FTF, <.10
			FTF/Tel	Cocaine use, past year	Tel, >.10
			FTF/Tel	Cocaine use, R	()
			FTF/Tel	Crack use, L	FTF, >.10
O'Reilly et al. (1994)	NP	26	ACASI/CASI	Marijuana use, R	ACASI, .58
	Volunteers		ACASI/CASI	Marijuana use, past year	CASI, .24
			ACASI/CASI	Marijuana use, L	CASI, .26
			ACASI/CASI	Cocaine use, R	()
			ACASI/CASI	Cocaine use, past year	CASI, .72
			ACASI/CASI	Cocaine use, L	CASI, .56
			ACASI/CASI	Crack use, R	()
			ACASI/CASI	Crack use, past year	CASI, .44
			ACASI/CASI	Crack use, L	ACASI, .58
		27	ACASI/PAP	Marijuana use, R	ACASI, .12
			ACASI/PAP	Marijuana use, past year	ACASI, .19
			ACASI/PAP	Marijuana use, L	ACASI, .29
			ACASI/PAP	Cocaine use, R	()
			ACASI/PAP	Cocaine use, past year	ACASI, .52
			ACASI/PAP	Cocaine use, L	ACASI, .06
			ACASI/PAP	Crack use, R	()
			ACASI/PAP	Crack use, past year	ACASI, .52
			ACASI/PAP	Crack use, L	ACASI, .12
		25	CASI/PAP	Marijuana use, R	CASI, .22
			CASI/PAP	Marijuana use, past year	CASI, .03
			CASI/PAP	Marijuana use, L	CASI, .06

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			CASI/PAP	Cocaine use, R	()
			CASI/PAP	Cocaine use, past year	CASI, .48
			CASI/PAP	Cocaine use, L	CASI, .04
			CASI/PAP	Crack use, R	()
			CASI/PAP	Crack use, past year	CASI, .22
			CASI/PAP	Crack use, L	CASI, .22
Tourangeau and Smith (1996)	P	197	ACASI/FTF	Marijuana use, L	ACASI, <.01
	HH residents in Cook County, Illinois, 1994		ACASI/FTF	Marijuana use, past year	ACASI, <.05
			ACASI/FTF	Marijuana use, R	ACASI, <.10
			ACASI/FTF	Cocaine use,L	ACASI, <.05
			ACASI/FTF	Cocaine use, past year	ACASI, >.10
			ACASI/FTF	Cocaine use, R	ACASI, >.10
		205	ACASI/CASI	Marijuana use, L	ACASI, >.10
			ACASI/CASI	Marijuana use, past year	ACASI, <.05
			ACASI/CASI	Marijuana use, R	ACASI, >.10
			ACASI/CASI	Cocaine use,L	ACASI, <.05
			ACASI/CASI	Cocaine use, past year	ACASI, >.10
			ACASI/CASI	Cocaine use, R	ACASI, >.10
		219	CASI/FTF	Marijuana use, L	CASI, <.05
			CASI/FTF	Marijuana use, past year	FTF, >.10
			CASI/FTF	Marijuana use, R	CASI, >.10
			CASI/FTF	Cocaine use, L	CASI, >.10
			CASI/FTF	Cocaine use, past year	CASI, >.10
			CASI/FTF	Cocaine use, R	FTF, >.10
Tourangeau et al. (1997)	P & NP	1,100	F-P/FTF	Any illegal drug use, L	?, >.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
	Chicago & sample of women who had an abortion, 1993		F-P/FTF	Marijuana use, L	?, >.05
			F-P/FTF	Amphetamine use, L	?, >.05
			F-P/FTF	Barbiturates use, L	?, >.05
			F-P/FTF	Tranquilizer use, L	?, >.05
			F-P/FTF	Psychedelics use, L	?, >.05
			F-P/FTF	Cocaine use, L	?, >.05
			F-P/FTF	Crack use, L	?, >.05
			F-P/FTF	Heroin use, L	?, >.05
			F-P/FTF	Injectable drugs use, L	?, >.05
			PAP/F-P	Any illegal drug use, L	?, >.05
			PAP/F-P	Marijuana use, L	?, >.05
			PAP/F-P	Amphetamine use, L	?, >.05
			PAP/F-P	Barbiturates use, L	?, >.05
			PAP/F-P	Tranquilizer use, L	?, >.05
			PAP/F-P	Psychedelics use, L	?, >.05
			PAP/F-P	Cocaine use, L	?, >.05
			PAP/F-P	Crack use, L	?, >.05
			PAP/F-P	Heroin use, L	?, >.05
			PAP/F-P	Injectable drugs use, L	?, >.05
			CASI/F-P	Any illegal drug use, L	?, >.05
			CASI/F-P	Marijuana use, L	?, >.05
			CASI/F-P	Amphetamine use, L	?, >.05
			CASI/F-P	Barbiturates use, L	?, >.05
			CASI/F-P	Tranquilizer use, L	?, >.05
			CASI/F-P	Psychedelics use, L	?, >.05
			CASI/F-P	Cocaine use,L	?, >.05
			CASI/F-P	Crack use, L	?, >.05
			CASI/F-P	Heroin use, L	?, >.05
			CASI/F-P	Injectable drugs use, L	?, >.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			PAP/FTF	Any illegal drug use, L	?, >.05
			PAP/FTF	Marijuana use, L	?, >.05
			PAP/FTF	Amphetamine use, L	?, >.05
			PAP/FTF	Barbiturates use, L	?, >.05
			PAP/FTF	Tranquilizer use, L	?, >.05
			PAP/FTF	Psychedelics use, L	?, >.05
			PAP/FTF	Cocaine use, L	?, >.05
			PAP/FTF	Crack use, L	?, >.05
			PAP/FTF	Heroin use, L	?, >.05
			PAP/FTF	Injectable drugs use, L	?, >.05
			CASI/FTF	Any illegal drug use, L	?, >.05
			CASI/FTF	Marijuana use, L	?, >.05
			CASI/FTF	Amphetamine use, L	?, >.05
			CASI/FTF	Barbiturates use, L	?, >.05
			CASI/FTF	Tranquilizer use, L	?, >.05
			CASI/FTF	Psychedelics use, L	?, >.05
			CASI/FTF	Crack use, L	?, >.05
			CASI/FTF	Cocaine use, L	?, >.05
			CASI/FTF	Heroin use, L	?, >.05
			CASI/FTF	Injectable drugs use, L	?, >.05
			CASI/PAP	Any illegal drug use, L	?, >.05
			CASI/PAP	Marijuana use, L	?, >.05
			CASI/PAP	Amphetamine use, L	?, >.05
			CASI/PAP	Barbiturates use, L	?, >.05
			CASI/PAP	Tranquilizer use, L	?, >.05
			CASI/PAP	Psychedelics use, L	?, >.05
			CASI/PAP	Cocaine use, L	?, >.05
			CASI/PAP	Crack use, L	?, >.05
			CASI/PAP	Heroin use, L	?, >.05
			CASI/PAP	Injectable drugs use, L	?, >.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
Beebe et al. (1998)	NP	368	CASI/PAP	Marijuana use, past year	PAP, <.05
	US students, 1996		CASI/PAP	LSD/psychedelic use, past year	PAP, >.05
			CASI/PAP	Amphetamine use, past year	PAP, <.05
			CASI/PAP	Cocaine use, past year	CASI, >.05
			CASI/PAP	Damaged property, past year	PAP, <.001
			CASI/PAP	Beat person up, past year	PAP, <.05
			CASI/PAP	Stolen something, past year	PAP, >.05
Rogers et al. (1998)	P	1,877	F-P/FTF	Cocaine use, R	F-P, .01
	NHSDA, 1990		F-P/FTF	Cocaine use, past year	F-P, .025
			F-P/FTF	Cocaine use, L	F-P, .33
			F-P/FTF	Marijuana use, R	F-P, .005
			F-P/FTF	Marijuana use, past year	F-P, .04
			F-P/FTF	Marijuana use, L	F-P, .37
Turner et al. (1998)	P	1,672	ACASI/PAP	Sex with prostitute	ACASI, <.01
	Male residents of US HHs, 1995		ACASI/PAP	Paid for sex	ACASI, <.10
			ACASI/PAP	Street drugs w/ needle	ACASI, <10
			ACASI/PAP	Injected drugs, past year	ACASI, .13
			ACASI/PAP	Daily marijuana use, past year	ACASI, <.10
			ACASI/PAP	Cocaine/crack use, past year	ACASI, >.10
			ACASI/PAP	Marijuana use, L	ACASI, <.10
			ACASI/PAP	Threaten someone, past year	ACASI, <.01
			ACASI/PAP	Physical fight, past year	ACASI, >.10
			ACASI/PAP	Pulled knife/gun, past year	ACASI, <.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
Wright et al. (1998)	P	3,169	CASI/PAP	Illicit drug use	CASI, >.05
	US urban & suburban residents, 1995–1996				
Supple et al. (1999)	P	1,072	CASI/F-P	Marijuana use, L	CASI, >.05
	US urban & suburban residents, 1995-1996		CASI/F-P	Marijuana use, R	CASI, <.05
			CASI/F-P	Marijuana use, >12x ever	CASI, >.05
			CASI/F-P	Illicit drug use, L	CASI, <.05
			CASI/F-P	Illicit drug use, R	CASI, <.05
Bason (2000)	P	365	Mail/Tel	Marijuana use, R	Tel, >.05
	US college students, 2000		Mail/Tel	Cocaine use, R	Mail, >.05
			Mail/Tel	Hallucinogens	Tel, >.05
			Mail/Tel	Designer drugs	Tel, >.05
		276	Web/Tel	Marijuana use, R	Tel, >.05
			Web/Tel	Cocaine use, R	Web, >.05
			Web/Tel	Hallucinogens	Tel, <.005
			Web/Tel	Designer drugs	Tel, >.05
		289	TACASI/Tel	Marijuana use, R	Tel, <.05
			TACASI/Tel	Cocaine use, R	TACASI, >.05
			TACASI/Tel	Hallucinogens	()
			TACASI/Tel	Designer drugs	Tel, >.05
		319	Web/Mail	Marijuana use, R	Web, >.05
			Web/Mail	Cocaine use, R	Web, >.05
			Web/Mail	Hallucinogens	Web, >.05
			Web/Mail	Designer drugs	Mail, >.05
		332	Mail/TACASI	Marijuana use, R	Mail, >.05
			Mail/TACASI	Cocaine use, R	Mail, >.05
			Mail/TACASI	Hallucinogens	TACASI, >.05
			Mail/TACASI	Designer drugs	TACASI, >.05
		243	Web/TACASI	Marijuana use, R	Web, >.05
			Web/TACASI	Cocaine use, R	Web, >.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			Web/TACASI	Hallucinogens	TACASI, >.05
			Web/TACASI	Designer drugs	TACASI, >.05
Gribble et al. (2000)	P	2,343	TACASI/Tel	Crack, past 6 months	TACASI, .026
	Men who have sex w/men in 4 US cities, 1996–1998		TACASI/Tel	Inhalants, past 6 months	TACASI, .004
			TACASI/Tel	Downers, past 6 months	TACASI, .001
			TACASI/Tel	Opiates, past 6 months	TACASI, .039
			TACASI/Tel	Marijuana, past 6 months	Tel, >.10
			TACASI/Tel	Psychedelics, past 6 months	Tel, >.10
			TACASI/Tel	Meth, past 6 months	TACASI, >.10
			TACASI/Tel	Other amphet, past 6 months	Tel, >.10
			TACASI/Tel	Ecstasy, past 6 months	Tel, >.10
			TACASI/Tel	Cocaine, past 6 months	TACASI, >.10
			TACASI/Tel	Party drugs, past 6 months	TACASI, .17
			TACASI/Tel	Rcvd drugs/money for sex, R	TACASI, .009
			TACASI/Tel	Gave drugs/money for sex, R	TACASI, .018
Corkrey & Parkinson (2002)	P	406	TACASI/Tel	Marijuana use, L	TACASI, >.10
	Australia		TACASI/Tel	Marijuana use, past year	TACASI, >.10
	residents, 2000		TACASI/Tel	Monthly marijuana use	TACASI, >.10
			TACASI/Tel	Amphetamine use, L	Tel, >.10
			TACASI/Tel	Amphetamine use, past year	TACASI, <.05
			TACASI/Tel	Monthly amphetamine use	TACASI, <.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			TACASI/Tel	Heroin use, L	TACASI, >.10
			TACASI/Tel	Heroin use, past year	TACASI, >.10
			TACASI/Tel	Monthly heroin use	TACASI, >.10
McCabe et al. (2002)	P	3,567	Web/Mail	Marijuana use, R	Mail, <.0005
	US college students, 2001				
Newman et al. (2002)	NP	1,417	ACASI/FTF	Marijuana use, R	ACASI, .731
	Syearinge prog participants in 4 U.S. cities, 1997–1998		ACASI/FTF	Non-Rx methadone use, R	ACASI, .004
Couper et al. (2003)	NP	195	ACASI/CASI	Drive under infl, past year	CASI, >.05
	Convenience sample (newspaper ads, flyers)		ACASI/CASI	Marijuana use, L	CASI, >.05
			ACASI/CASI	Cocaine use,L	CASI, >.05
Knapp & Kirk (2003)	NP	231	Web/Mail	Marijuana use, L	Web, >.10
	US college students, 1999		Web/Mail	Been in jail, L	Mail, .07
			Web/Mail	Used CC w/o permission	Web, >.10
			Web/Mail	Accepted money for sex	Mail, >.10
		295	Mail/TACASI	Marijuana use, L	Mail, >.10
			Mail/TACASI	Been in jail, L	Mail, .10
			Mail/TACASI	Used CC w/o permission	Mail, >.10
			Mail/TACASI	Accepted money for sex	Mail, >.10
		178	Web/TACASI	Marijuana use, L	Web, >.10
			Web/TACASI	Been in jail, L	TACASI, .46
			Web/TACASI	Used CC w/o permission	Web, >.10
			Web/TACASI	Accepted money for sex	TACASI, >.10
McCabe (2004)	P	3,606	Web/Mail	Marijuana use, L	Mail, <.05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
	US college students, 2001		Web/Mail	Marijuana use, past year	Mail, >.10
			Web/Mail	Ecstasy use, L	Web, <.05
			Web/Mail	Ecstasy use, past year	Web, >.10
			Web/Mail	Narcotics use, L	Mail, >.10
			Web/Mail	Narcotics use, past year	Mail, >.10
			Web/Mail	LSD use, L	Mail, >.10
			Web/Mail	LSD use, past year	Mail, >.10
			Web/Mail	Psychedelics, L	Web, >.10
			Web/Mail	Psychedelics, past year	Mail, >.10
			Web/Mail	Inhalant use, L	Web, >.10
			Web/Mail	Inhalant use, past year	Mail, >.10
			Web/Mail	Cocaine use, L	Web, <.001
			Web/Mail	Cocaine use, past year	Web, >.10
			Web/Mail	Amphetamines, L	Mail, <.05
			Web/Mail	Amphetamines, past year	Mail, <.10
Turner et al. (2005)	P	2,228	TACASI/Tel	Marijuana use, R	TACASI, <.001
	USA & Baltimore, 1999–2000		TACASI/Tel	Marijuana use, past year	TACASI, <.001
			TACASI/Tel	Marijuana use, past 3 years	TACASI, <.001
			TACASI/Tel	Marijuana use, L	TACASI, <.05
			TACASI/Tel	Cocaine use, R	TACASI, <.05
			TACASI/Tel	Cocaine use, past year	TACASI, <.05
			TACASI/Tel	Cocaine use, past 3 years	TACASI, <.001
			TACASI/Tel	Cocaine use, L	TACASI, <.05
			TACASI/Tel	Drug injection use, past year	TACASI, <.01
			TACASI/Tel	Drug injection use, past 5 years	TACASI, <.01

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			TACASI/Tel	Drug injection use, L	TACASI, <.05
Wang et al. (2005)	P	1,918	Web/PAP	Ecstasy use	Web, >.05
	Students in Taipei, 2003		Web/PAP	Ketamine use	Web, >.05
			Web/PAP	Marijuana use	Web, >.05
			Web/PAP	Amphetamine use	Web, <.05
			Web/PAP	Illicit drug use, L	Web, >.05
Brenek et al. (2006)	NP	4,506	CASI/PAP	Drunk driving, R	CASI, <.05
	High school students in 8 states, 2004		CASI/PAP	Carried a gun, R	CASI, >.05
			CASI/PAP	Weapon carrying, R	CASI, >.05
			CASI/PAP	Marijuana use, L	CASI, >.05
			CASI/PAP	Marijuana use, before 13	CASI, >.05
			CASI/PAP	Marijuana use, R	CASI, >.05
			CASI/PAP	Cocaine use, L	CASI, >.05
			CASI/PAP	Inhalant use, L	CASI, >.05
			CASI/PAP	Meth use, L	CASI, >.05
			CASI/PAP	Ecstacy use, L	CASI, >.05
			CASI/PAP	Steroid use, L	PAP, >.05
Hamby et al. (2006)	NP	160	CASI/PAP	Physical assault perp.	CASI, <.05
	US college students		CASI/PAP	Sexual coercion perp.	PAP, <.01
			CASI/PAP	Injury perp.	CASI, <.10
Lucia et al. (2007)	NP	1,203	Web/PAP	Marijuana use, L	PAP, >.05
	Students in Lausanne, Switzerland, 2004		Web/PAP	Heroin, L	Web, >.05
			Web/PAP	Cocaine, L	PAP, >.05
			Web/PAP	Ecstasy, L	PAP, >.05
			Web/PAP	Hallucinogens	PAP, >.05
			Web/PAP	Amphetamines	Web, >05

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			Web/PAP	Driving w/o license	Web, <.01
			Web/PAP	Shoplifting	Web, >.05
			Web/PAP	Breaking into car	Web, >.05
			Web/PAP	Theft at school	Web, >.05
			Web/PAP	Theft at home	Web, <.05
			Web/PAP	Vehicle theft	PAP, >.05
			Web/PAP	Assault	PAP, >.05
			Web/PAP	Threat w/weapon	PAP, >.05
			Web/PAP	Racket	PAP, >.05
			Web/PAP	Robbery	()
			Web/PAP	Arson	PAP, >.05
			Web/PAP	Selling soft drugs	PAP, <.05
			Web/PAP	Selling hard drugs	PAP, >.05
			Web/PAP	Graffiti	PAP, >.05
			Web/PAP	Vandalism	PAP, <.01
			Web/PAP	Theft from person	PAP, >.05
van de Looij-Jansen and de Wilde (2008)	NP	531	Web/PAP	Vandal & steal, past year	?, >.05
	Schools in Rotterdam, Netherlands, 2005		Web/PAP	Carrying a weapon	W, <.05
			Web/PAP	Marijuana use, R	?, >.05
Eaton et al. (2010)	NP	5,227	Web/PAP	Drunk driving	Web, .01
	85 schools in 15 states, 2008		Web/PAP	Carried a gun	Web, .02
			Web/PAP	Weapon carrying	Web, .01
			Web/PAP	Weapon carrying @ school	Web, .005
			Web/PAP	Marijuana use, L	Web, .12
			Web/PAP	Marijuana use, before 13	Web, .07
			Web/PAP	Marijuana use, R	Web, .02
			Web/PAP	Marijuana use @ school	Web, .005
			Web/PAP	Cocaine use, L	Web, .145
			Web/PAP	Cocaine use, R	Web, .0175

(continued)

(continued)

Study	Sample description	Sample size	Modes compared	Illegal behavior	Better Mode, significance
			Web/PAP	Inhalant use, L	Web, .085
			Web/PAP	Heroin use, L	Web, .145
			Web/PAP	Meth use, L	Web, .12
			Web/PAP	Ecstacy use, L	Web, .02
			Web/PAP	Steroid use, L	Web, .155
			Web/PAP	Injection drug use, L	Web, .125

Abbreviations
FTF Face-to-face interview (includes CAPI)—interviewer speaks questions, respondent speaks responses
F-P Face-to-face interview—interviewer speaks questions, respondent records responses on paper or Computer
PAP Paper and pencil self-administered questionnaire
Mail Respondent receives and returns survey by mail
Tel Telephone interview (e.g. CATI)
CASI Computer-assisted self-interviewing
ACASI Audio computer-assisted self-interviewing
TACASI Telephone audio computer-assisted self-interviewing (includes IVR—interactive voice response)
W Web or Internet mode
L Lifetime use
R Use in past month (recent)
() = No difference between modes
? = Not clear which mode elicited more responses
P Probability sample
NP Nonprobability sample
HH Household
NHSDA National Household Survey of Drug Abuse
NLSY National Longitudinal Survey of Youth

coverage error, sampling error, and nonresponse as well as measurement or response error. If these sources of error were independent of one another, each could be addressed separately without regard to the others, but unfortunately they are sometimes interrelated. Using a survey mode that yields more valid responses to self-report questions about deviant behavior might also be afflicted by worse nonresponse (a lower percent of those selected for the sample agreeing to participate), producing more potential for sample bias. Or, a mode effective in eliciting admissions of unlawful conduct, such as Internet surveying, may only be feasible with the portion of the sample possessing the necessary technology, i.e., a computer and home Internet access. This can (if not corrected in various expensive ways) result in coverage error—i.e., some segment of the target population is not covered by the sample frame. Specifically, coverage may be poor for lower income persons—the very individuals most likely to commit common law crimes (also see Mesch Chap. 18, and Manzo and Burke Chap. 19 for the availability and characteristics of those participating in e-mail and web-based surveys). Some modes that are relatively effective in reducing response error for questions on sensitive topics may not be feasible except in circumstances that preclude use of probability sampling, as is the case with nearly all of the group-administered pencil-and-paper surveys of school samples that are so common in criminology. And of course, some effective modes may simply be more expensive than a researcher can afford. In sum, our exploratory findings pertain to one important dimension of surveys on crime, but not the only one. Scholars need to consider these findings in conjunction with a concern for the other sources of survey error.

We also want to stress the tentative nature of our rankings of survey modes. Because there have been so few direct comparisons of modes, relative to all the possible pairwise comparisons that might be made, the exact ranking of modes is based on a modest empirical foundation, and future research may well necessitate revision. Even those pairs of modes that have been directly compared have, in some cases, only been assessed in one or two studies, or only in connection with one or two offenses. Finally, most findings in this area pertain to illegal drug use, and it is possible they do not apply to other kinds of criminal behavior. Future research in this area needs to be directed at a wider array of criminal acts, to provide a broader base for judging the effectiveness of survey modes in eliciting reports of these behaviors.

24.7

Appendix Table A Individual findings of experimental studies of the effect of survey mode on the reporting of illegal behaviors (in chronological order)

References

Aquilino, W. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *The Public Opinion Quarterly, 58*, 210–240.

Bason, J. J. (2000). *Comparison of telephone, mail, web, and IVR surveys of drug and alcohol use among University of Georgia students*. Paper presented at the annual meeting of the American Association of Public Opinion Research, Portland, OR.

Beebe, T. J., Harrison, P. A., McRae, J. A. Jr., Anderson, R. E., & Fulkerson, J. A. (1998). An evaluation of computer-assisted self-interviews in a school setting. *The Public Opinion Quarterly*, *62*, 623–632.

Brenek, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., et al. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *The Public Opinion Quarterly*, *70*, 354–374.

Clark, J. P., & Tifft, L. L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review*, *34*, 516–523.

Corkrey, R., & Parkinson, L. (2002). A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions. *Behavior Research Methods, Instruments, and Computers*, *34*, 354–363.

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of Audio-CASI on self-reports of sensitive behavior. *The Public Opinion Quarterly*, *67*, 385–395.

Eaton, D. K., Brener, N. D., Kann, L., Denniston, M. M., McManus, T., Kyle, T. M., et al. (2010). Comparison of paper-and-pencil versus web administration of the Youth Risk Behavior Survey (YEARBS): Risk behavior prevalence estimates. *Evaluation Review*, *34*, 137–153.

Gribble, J. N., Miller, H. G., Cooley, P. C., Catania, J. A., Pollack, L., & Turner, C. F. (2000). The impact of T-ACASI interviewing on reported drug use among men who have sex with men. *Substance Use and Misuse*, *35*, 869–890.

Hamby, S., Sugarman, D. B., & Boney-McCoy, S. (2006). Does questionnaire format impact reported partner violence rates? An experimental study. *Violence and Victims*, *21*, 507–518.

Harrison, L. (1995). The validity of self-reported data on drug use. *Journal of Drug*, *25*, 91–111.

Johnson, T. P., Hougland, J. G., & Clayton, R. R. (1989). Obtaining reports of sensitive behavior: A comparison of substance use reports from telephone and face-to-face interviews. *Social Science Quarterly*, *70*, 174–183.

Kleck, G., Tark, J., & Bellows, J. J. (2006). What methods are most frequently used in research in criminology and criminal justice? *Journal of Criminal Justice*, *34*(2), 147–152.

Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: Does methodology matter? *Computers in Human Behavior*, *19*, 117–134.

Li, S.-M., Yu, Rwei-Ling, S., Hu, H.-C., & Huang, J.-S. (2003). Areca quid chewing by Taiwanese adolescents: Application of the Attitudes Social Influence Self-efficacy (ASE) model. *Addiction*, *98*, 1723–1729.

Lucia, S., Herrmann, L., & Killias, M. (2007). How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology*, *3*, 39–64.

McCabe, S. E. (2004). Comparison of Web and mail surveys in collecting illicit drug use data: A randomized experiment. *Journal of Drug Education*, *34*, 61–72.

McCabe, S. E., Boyd, C. J., Couper, M. P., Crawford, S., & D'Arcy, H. (2002). Mode effects for collecting alcohol and other drug use data: Web and U.S. mail. *Journal of Studies on Alcohol*, *63*, 755–761.

Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble. J., Cooley, P., & Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health*, *92*, 294–297.

O'Reilly, J. M., Hubbard, M. L., Lessler, J. T., Biemer, P. P., & Turner, C. F. (1994). Audio and video computer-assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, *10*, 197–214.

Rogers, S. M., Miller, H. G., & Turner, C. F. (1998). Effects of interview mode on bias in survey measurements of drug use: Do respondent characteristics make a difference? *Substance Use and Misuse*, *3*, 2179–2200.

Schober, S., Caces, M. F., Pergamit, M., & Branden, L. (1992). Effect of mode of administration on reporting of drug use in the National Longitudinal Survey. In C. F. Turner, J. T. Lessler & J. C. Gfroerer (Eds.), *Survey measurement of drug use*: *Methodological studies* (pp. 267–276). Rockville, MD: National Institute on Drug Abuse

Supple, A. J., Aquilino, W. S., & Wright, D. L. (1999). Collecting sensitive self-report data with laptop computers: Impact on the response tendencies of adolescents in a home interview. *Journal of Research on Adolescence*, *9*, 467–488.

Thornberry, T. P., & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. In *Criminal justice 2000*: *Measurement and analysis of crime and justice* (Vol. 4, pp. 33–83). Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Tourangeau, R., Jared, B. J., William, F. P., & Kenneth, R. (1997). Design and results of the women's health study. In Harrison, L., & Hughes, A. (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 344–365). Washington, DC: U.S. Department of Health and Human Services.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *The Public Opinion Quarterly*, *60*, 275–304.

Tourangeau, R., & Ting, Y. (2007). Sensitive questions in surveys. *Psychol Bulletin*, *133*(5), 859–883

Turner, C. F., Lessler, J. T., & Devore, J. W. (1992). Effects of mode of administration and wording on reporting of drug use. In C. F. Turner, J. T. Lessler & J. C. Gfroerer (Eds.), *Survey measurement of drug use*: *Methodological studies*. Washington, D.C.: DHHS Pub. No. (ADM)92-1929.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867–873.

Turner, C. F., Villarroel, M. A., Rogers, S. M., Eggleston, E., Ganapathi, L., Roman, A. M., et al. (2005). Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of Telephone Audio-CASI. *Addiction*, *100*, 1432–1444.

van de Looij-Jansen, P. M., & de Wilde, E. J. (2008). Comparison of Web-based versus paper-and-pencil self-administered questionnaire: Effects on health indicators in Dutch adolescents. *Health Services Research*, *43*, 1708–1721.

Wang, Y.-C., Lee, C.-M., Lew-Ting, C.-Y., Hsiao, C. K., Chen, D.-R., & Chen, W. J. (2005). Survey of substance use among high school students in Taipei: Web-based questionnaire versus paper-and-pencil questionnaire. *Journal of Adolescent Health*, *37*, 289–295.

Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *The Public Opinion Quarterly*, *62*, 331–353.

Part VI

Survey Designs and Construction

Issues in Survey Design: Using Surveys of Victimization and Fear of Crime as Examples

25

Sue-Ming Yang and Joshua C. Hinkle

25.1 Introduction

Surveys are one of the most used tools in social science research due to their versatility, efficiency, and generalizability (Bachman and Schutt 2008). The history of surveys can be dated back as far as ancient Egypt (Babbie 1990). Early surveys tended to be used for political purposes as a tool to understand citizens' attitudes (e.g., Marx 1880), a task they are still widely used for today. Surveys are versatile and can be used to study a wide variety of topics and, compared to many other types of research, collecting survey data is quite efficient and less time-consuming. Additionally, when combined with a random sampling strategy, results from survey research tend to have high generalizability (Schuman and Presser 1981; Tourangeau et al. 2000). While surveys clearly have many practical benefits for social science research, this is not to say survey construction and administration is a minor undertaking. The use of self-report surveys in criminal victimization has dramatically changed the definition of crime and the availability of information about crime (Cantor and Lynch 2000, p. 86). Prior to the launch of the National Crime Victimization Survey (NCVS), information about crime was primarily obtained through official crime information collected by the FBI in their Uniform Crime Report. The application of victimization surveys within our field has allowed a more direct measure of the prevalence rates of crime and victimization.

With the rise of victimization surveys as a measure of crime, it is important to realize that the design of surveys can affect the results. Deming (1944) first cautioned scholars that the usefulness of surveys could possibly be undermined by 13 types of errors, thus, "…the accuracy supposedly required of a proposed survey is frequently exaggerated…and is in fact often unattainable" (p. 369). Among these potential errors, he argued that imperfections in the design of questionnaires are one major source of bias. Ever since Deming raised the flag on the utility and accuracy of surveys in obtaining results, more scholars have turned their attention to different components of survey errors (Cochran 1953; Kish 1965; Dalenius 1974). Andersen et al. (1979) systematically summarized studies

We would like to thank Charles Hogan and Daniel Jarvis for their hard work in obtaining and coding surveys of fear of crime and victimization. We also want to thank Lior Gideon for his thoughtful comments on an earlier draft.

S.-M. Yang (✉)
Department of Criminology,
National Chung Cheng University,
168 University Rd., Min-Hsiung,
Chia-Yi, Taiwan
e-mail: smyang@ccu.edu.tw; yangsueming@gmail.com

J. C. Hinkle
Department of Criminal Justice and Criminology,
Georgia State University, P.O. Box 4018, Atlanta,
GA 30302, USA
e-mail: jhinkle@gsu.edu

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_25, © Springer Science+Business Media New York 2012

examining survey errors associated with individual non-response, measurement, and processing. His results were published in a book titled *Total Survey Error: Applications to Improve Health Surveys*. The term "total survey error" has been used ever since to indicate all sources of errors associated with both observational and non-observational errors in survey research (see Groves and Lyberg 2010 for a comprehensive review of the history of total survey error).

Gradually, the field of criminology and criminal justice also began to notice the potential impacts survey designs can impose on study outcomes. For example, findings from the NCVS in 1975 demonstrated that if a fear of crime question was asked prior to the screening interview, the likelihood of reporting victimization increased drastically (Murphy 1976; Gibson et al. 1978). This result signifies a "warm-up" effect, where the initial question reinforced the answers given for subsequent questions (Lynch 1996). Cantor and Lynch (2000) cautioned victimization scholars who use survey instruments that "[d]ata from victim surveys are heavily influenced by their design" (p. 128) and pointed out that a better understanding of the impact of survey design is imperative for advancing the field. In order to achieve high reliability and validity, it is important to take caution in every step of survey research.

The example cited above is just one illustration of how survey designs can result in different research outcomes. Response effects have many plausible sources including problems in understanding the questions, difficulties in recalling information, producing appropriate answers, and other cognitive processes (Groves and Lyberg 2010; Walonick 1993; Tourangeau et al. 2000). Specifically, researchers could influence survey results in a variety of ways related to how they choose to construct their questionnaires (Kury 1994). In this chapter, we will discuss some important, but often overlooked, topics related to survey research. Three sources of possible errors in survey design will be reviewed and discussed: (1) design of response options; (2) wording of the questions; (3) question-order (context) effects.

The first relates to what options survey respondents are given to answer the questions. In particular we will focus on whether or not respondents are given a neutral (or middle) option on Likert scale-type questions, and whether a "don't know" (DK) option is provided. The second issue relates to wording of the questions. The ways in which a question is worded can bias findings, or even lead to a situation where the question is not actually measuring the construct that the researcher was intending to measure. Finally, the third issue deals with bias that may be introduced by the order in which the questions are asked. That is to say that sometimes asking certain questions first may influence the answers given to later questions. This occurs when the questions in the survey are interrelated but not independent—responses to later questions in the survey are therefore conditioned by answers to the former ones. Thus, the order of the questions would result in different responses (Walonick 1993; Bishop 1987; Tourangeau and Rasinski 1988; Ayidiya and McClendon 1990; Lorenz et al. 1995; Kury 1994). This chapter will discuss each of these three survey design issues in detail and will use examples from victimization and fear of crime surveys as illustrations. Thirty-five surveys of victimization and fear of crime collected as part of an ongoing study (Yang and Hinkle, In Progress) were examined and used as examples for this chapter (see Appendix A).[1] A handful of state-level crime victimization surveys are used in addition to these as examples as well.

25.1.1 I. Response Options

The formatting of the response options is one element of survey design that may affect the outcome. For instance, one could get very

[1] As part of this ongoing study, these surveys were identified through database searches for studies of fear of crime and victimization. To be eligible for inclusion, the studies must have tested the relationship between victimization and fear of crime (i.e., at least reported the correlation between the two). However, that did not have to be the main focus of the study, it was just necessary for the published results to statistically examine the relationship between fear and victimization in some way.

different estimates of fear of crime from answers to open-ended versus close-ended questions. Fattah (1993) pointed out that "the element of suggestibility inevitably present in close-ended questions is probably responsible for yielding a higher percentage of respondents declaring being afraid than when open-ended questions are used" (p. 53). Other recent research has also found discordance between fear of crime (and other emotions like anger about crime) when comparing close-ended survey measures to responses from the same respondents in open-ended interviews (see Farrall et al. 1997; Farrall and Ditton 1999). Like Fattah, these studies also tend to find higher levels of fear and other emotional responses in data collected with close-ended surveys than in open-ended interviews. This is just one example of how the design of a survey can change the findings to a great extent. Considering that most victimization surveys use close-ended questions, our discussions below will also focus on response options related to close-ended questions only.

Among close-ended questions, the response options provided naturally affect the results. This issue is obvious, as the majority of survey questions are multiple-choice type questions which ask respondents to choose an answer from a list provided. For example, the most common fear of crime question over the years has been some variation of a question like "How safe would you feel walking alone outside at night in your neighborhood?" Response options provided to this question tend to be ratings of safety along the lines of "very safe," "somewhat safe," "somewhat unsafe," or "very unsafe." The first impact of response options provided is that they clearly limit the responses survey participants can give. As such, it is important to select meaningful and appropriate response options for each question.

Beyond that broad concern, the three most commonly seen issues relating to response options involve offering the middle alternative or not, providing a don't know (DK) category, and the arrangement of response options. The impact of these factors is often difficult to detect, because all the response options tend to remain the same across questions within the same survey. Nonetheless, this does not mean that the impact of these issues should be ignored. The issues related to the arrangement of response options (e.g., logical order) have already been reviewed extensively in every research methods textbook, and many other works, and are thus well understood (e.g. see, Maxfield and Babbie 2007; Bachman and Schutt 2010; Kury 1994; Schuman and Presser 1981; Tourangeau et al. 2000).[2] As such, we will only focus on the first two issues and discuss how they might affect scholars who study criminal justice topics.

Starting with the middle alternative issue, scholars have different opinions on whether to offer a middle alternative when asking attitudinal questions. Scholars who support providing a middle option believe that very few people are truly indifferent about any issue (Sudman and Bradburn 1982). Thus, offering a middle option should not change the survey results in any substantial way. On the other hand, some critics worry that respondents might not be willing to reveal their true feelings if the middle option is available as an easy way out. For example, Bishop (1987) has argued that offering middle alternatives to subjects would result in a substantial change in conclusions compared to data collected with the exclusion of a middle alternative, because people tend to choose middle alternatives when they are available. This tendency is especially true when the middle alternative is shown as the last response option of the question. Ayidiya and McClendon (1990) also found that more people tend to choose middle option when provided. However, they pointed out that the overall conclusions drawn from the research are unlikely to be affected as the ratio of respondents choosing either end of the response spectrum does not change. That is, the availability of a middle alternative was found to

[2] We are not saying that the study of response sequence is unimportant. For example, research has also found that simple changes in the sequencing of close-ended responses produced very different results (see Kury and Ferdinand 1998). However, it is more common for one survey to arrange all response options in a similar fashion across questions, and thus the effects are hard to detect.

draw about the same amount of people who would otherwise have chosen a positive or negative answer. As such, offering a middle alternative has no substantial impacts on conclusions drawn from the survey results. Nonetheless, many researchers or practitioners still prefer not to provide middle alternatives as response options. In general, most victimization surveys that are adapted from NCVS (like the Utah Crime Victimization Survey 2006, Minnesota Crime Survey 1999, and Maine Crime Victimization Survey 2006) tend not to include middle alternatives in their response options.

Moving on to the second response option issue, the provision of a DK option is another common choice survey designers regularly face. A common worry regarding the inclusion of a DK option is that it will cause people to refrain from disclosing their true feelings. However, the research findings regarding the inclusion of a DK option are very similar to those for the middle alternatives. Studies show that providing respondents with an option of answering DK does reduce the number of people who would otherwise choose other substantive responses. Even excluding the DK category, the overall pattern is not affected (Sudman and Bradburn 1982). That is, providing respondents with DK as an option does not necessarily pull respondents out of other possible response categories. Thus, the concern related to the DK option biasing survey results might not be warranted. For example, the Utah Crime Victimization Survey 2006 began including DK as a possible response option for many questions, for which it was not offered in previous versions of the survey. A breakdown of the collected data shows that only less than 2% of respondents (out of 1,199 subjects) chose DK rather than other possible options when asked to describe their overall perception of fear. For questions that were related to their perception of safety in the community, the percentage of respondents selecting DK was even lower. Based on the response patterns across four biannual surveys, providing the DK option did not really pull those potential "fence-sitters" (or "floaters" as Schuman and Presser 1981 call them) from their original response category. Rather, only less than 1% of respondents chose DK when they were asked to report their fear of crime and likelihood of victimization.

A review of surveys on victimization and fear of crime shows no consensus on whether to include DK as a response option (see Appendix A for surveys used in the review). For instance, a victimization survey in Maine (2007 version) did not include the DK as a possible option for response. A crime victimization survey in Alaska allowed the respondents to answer DK; however, the DK responses were not coded in the database and were simply treated as missing values. In the Minnesota Crime Survey (1999 version), a DK option was not provided to the questions related to perceptions of fear of crime; however, respondents could choose "unknown" when answering factual questions related to victimization details.

One additional concern that has been raised is that the offering of the middle category and the DK option can interact with each other. Meaning, the offering of a middle option might divert people who would otherwise have chosen the DK option, because choosing the middle option seems to be a more substantive take on the issue than giving the DK answer. Thus, for scholars who use surveys it is important to design survey responses in ways which will generate the most complete and accurate information. Doing so requires tailoring things like response options to the topic at hand. For questions that ask for pure information (e.g., questions asking "how many past victimization experiences" the respondent had or asking about their level of "fear of crime"), you might not want to provide the DK option as the respondents should be able to make a judgment or recall their personal experience. However, when questions of interest are related to knowledge that may not be accessible in every case—such as details regarding victimization, or value judgments—then it might be reasonable to allow respondents to choose the middle alternative or answer with DK. In such cases, a DK response may be the true answer of a respondent who lacks knowledge related to the given question. Thus not providing DK as an option may force them to merely guess at a response.

In sum, the choice of whether to include middle or DK responses depends upon the focus of the particular study and the nature of inquiry. Research has generally shown that including these response options does not bias conclusions in any substantive way. People choosing them were found to have been equally likely to have chosen positive or negative response options otherwise (see Schuman and Presser 1981). However, there are other concerns related to response options. For example, it is important to take into account potential interaction between middle and DK responses, and whether to treat DK responses as missing values. The latter could be a concern for data analysis as removing those cases risks lowering sample size and lessening statistical power. That said, the number one concern of surveys is collecting accurate information, thus researchers should not exclude DK responses solely to force respondents to choose answers for analysis' sake. Researchers must examine each question on a case-by-case basis and decide whether DK and middle response options are necessary to collect accurate information.

25.1.2 II. Question Wording

A second major issue in survey design comes from the wording of the questions. Schuman and Presser (1981) suggest that question wording can affect results in two different ways: it could affect the marginal distribution (i.e., response tendency), or it could affect the size or direction of the relationships between responses given to questions. More specifically, the wording of questions often serves as a way to frame the issue of inquiry, and thus the way in which a question is worded could influence the responses to the given question (Kahneman and Tversky 1984).

As an example of the influence wording can have, customers are much more positive about ground beef labeled "75% lean" rather than "25% fat," despite the fact they mean exactly the same thing (Levin 1987). Another classic example of the potential influence of wording choices on survey responses comes from studies that asked respondents whether they agreed or disagreed with banning public speeches against democracy. These studies found that Americans were much more likely to agree with statements that did "not allow" the speeches than they were to agree with statements that "forbid" them (see Rugg 1941; Schuman and Presser 1981). While "forbid" and "not allow" are synonymous at face value, these studies clearly show they are not equal when it comes to generating responses with a survey question—at least in the case of the speech against democracy example. Numerous studies have replicated the wording effect and consistently found that question wording affects people's decisions (for example, see Levin et al. 1998; Walonick 1993; Heaps and Henley 1999). Additionally, question wording effects could be conditioned by background factors such as the education level of respondents (Schuman and Presser 1981, p. 6), which could further affect relationships between attitudinal variables and demographic variables and thus bias conclusions drawn from the data.

A good practical example of the importance of question wording comes from recent survey research on fear of crime. The most commonly used survey measures of fear of crime are variations of the fear question used by the Bureau of Justice Statistics' National Crime Survey. These questions ask respondents to report their levels of fear of crime with items along the lines of "How safe do you, or would you, feel walking alone outside at night in your neighborhood." Response options are typically "very unsafe, somewhat unsafe, somewhat safe, and very safe." In short, these questions assume that asking someone "how safe" they feel is analogous to asking them "how fearful" or "how afraid" they are. However, as respondents are easily affected by the exact wording used in survey questions just like the previous "forbid" and "not allow" example shows, such assumptions are not always correct.

These types of questions have recently come under attack, with critics arguing they are poor measures of fear of crime due to issues with their wording. Some have argued that this question taps more into perceptions of risk/safety, than

fear of crime (e.g., Ferraro 1995; Farrall 2004; Farrall and Gadd 2004a, b). The critique seems reasonable as the wording of these questions clearly asks people to rate their perceived level of safety and makes no mention of "fear" or being "afraid." Thus, the key concern here is whether asking people "how safe" they feel leads to different responses than if the questions had asked them "how afraid" or "how fearful" they were.

Research thus far has suggested that these wording differences do indeed lead to different response distributions of fear of crime. In the past, studies using these types of questions have consistently found that between a quarter to two-thirds of citizens are fearful about crime (see Farrall 2004). This paints a dire picture of modern society being an era where a significant portion of the population is living in fear of crime and are staying home behind locked doors (especially at night!). However, some have argued that the high prevalence of fear of crime is likely just an artifact of question wording problems embedded in the standard fear of crime measure. Thus, Farrall and colleagues have suggested that the use of such questions has led to an overstating of the actual level of fear of crime in society (Farrall 2004; Farrall and Gadd 2004a, b).

As one example, Farrall and Gadd (2004a, b) asked one sample of 995 individuals a safety question—"How safe do you feel walking alone in this area after dark?" The question had typical response options of "very safe," "fairly safe," "a bit unsafe" and "very unsafe." To demonstrate the effects of question wording, another sample of 977 individuals were given a new fear question—"In the past year, have you ever felt fearful about becoming a victim of crime?" This question was a simple yes/no question, and if the respondent answered yes, they were asked two follow-up questions dealing with the frequency of fear of crime ("How many times have you felt like this in the past year?") and the magnitude of fear ("And on the last occasion, how fearful did you feel?"). The magnitude question had response options of "not very fearful," "a little bit fearful," "quite fearful," and "very fearful." In short, their study found that while both types of questions showed similar proportions of the sample had felt fearful or unsafe, the responses to frequency and magnitude questions suggested that only looking at those proportions could overstate the severity of the fear of crime problem in society. Specifically, responses to the surveys showed that one-third of the sample who were asked the safety question reported feeling unsafe (a bit unsafe and very unsafe combined). This proportion is fairly typical of research on "fear" which uses such safety questions. Turning to their specific fear question, they also found that 37% of the sample reported having felt fearful of becoming a crime victim in the past year on the initial yes/no question. Thus, when looking only for a simple wording change, there does not appear to be much bias in the marginal distribution as relatively similar number of people reported feeling "fearful" as those who reported feeling "unsafe" in the other sample.

However, when they turned to the follow-up questions, they found that questions worded only to ask if a person has felt afraid or unsafe may overstate the level of fear in daily life. Specifically, 20% of the sample experienced only one fearful event in the past year, while nearly half of the same had only 1−4 fearful experiences during that time frame. Furthermore, the magnitude question revealed that those who had reported at least one fearful event did not rate the experience as particularly frightening, with only 15% of the sample reporting their experience to have been "quite fearful" or "very fearful." Perhaps more importantly, the study also found several other differences between the two types of fear measures when examining responses by demographic groups. As is typical in past fear of crime research using the perceived safety at night question, the data showed that older people and females are more fearful (see Ferraro 1995 for a review of studies on gender/age and fear). However, using their new fear measure the age gap disappeared and the gender gap was reduced from females being 28% more "unsafe" than males to only 13% more fearful when looking at their new fear question.

This is a good illustration of the importance of carefully considering the impacts of question wording. While sometimes the bias is readily apparent, as in the aforementioned examples involving "forbidding" or "not allowing" speeches or forming opinions about ground beef that is "75% fat" or "25% lean," they may also be more subtle and only reveal themselves upon closer examination. In the case of the fear of crime example, one could easily mistakenly conclude that asking respondents "how safe" or "how fearful/afraid" they are has little consequence since Farrall and Gadd (2004a, b) found that nearly identical percentages (33% and 37%, respectively) reported being unsafe or fearful. However, further inspection showed that such generally worded questions likely overstate levels of fear in a sample by ignoring frequency and magnitude of fear of crime events. Most people who reported feeling fearful at least once had only experienced 1−4 fearful events in the past year and on average they did not rate the event(s) as particularly frightening.

More importantly, their examination of demographic variables showed important differences in levels of fear compared to levels of safety by age and gender as discussed above. This suggests that even though similar proportions of the sample were identified using either measure, these two questions still may measure two distinct constructs. Indeed other studies have found this to be the case. For example, in a review of past work in this area, Ferraro and LaGrange (1987) found that general correlations between perceptions of safety/risk and fear of crime ranging from 0.32 to 0.48. Thus, while there is some relationship between the two, the low correlations and differential demographic effects on responses to fear of crime and perceived safety questions suggest that they are different constructs. Thus, one has to carefully decide on the wording of questions in order to reflect what they truly want to measure in surveys. The findings also point to problems with the past research on fear of crime on using measures of perceived safety/perceived risk while interpreting the findings under a general framework of fear.

Moreover, the above discussion focused only on *one* particular critique of the wording of the standard NCVS "fear" question and variants of it. This illustrates the complexities of question wording issue. Fear of crime is by no means a clearly operationalized concept. In addition to the "fear" versus "safety" issue discussed above, research has shown that the fear of crime must be distinguished from more global fears and that levels of fear vary across different types of crime. Thompson et al. (1992) utilized three measures of the fear of crime: "global fear," "fear of property crime", and "fear of violent crime." They found that fear varied across these different measures and argued that different methods of measurement could be a source of error. In another study, Keane (1992) drew a distinction between "concrete fear" and "formless fear" by using the standard NCVS fear question to assess these concepts.

> For each type of crime listed below, please indicate how likely you think it is to happen to you during the next year (or in the future). If you feel certain that it will not happen to you, then circle the number 0 beside the crime. If you feel certain that it will happen to you, then circle the number 10.

This item was followed by a list of specific crimes. Keane's (1992) findings of fear varying by crime type point out that the contradictory results in fear of crime research are at least partially due to different methods of operationalizing the concept of fear of crime.

In short, these studies, along with others, suggest that fear of crime is not a uniform construct and that many of the inconsistencies in fear of crime studies may come from different ways of wording the questions. In particular, several have suggested that it is preferable to ask about fear of specific types of crime rather than just "crime" as an omnibus construct given that levels of fear (or perceptions of safety and risk) have been found to vary across types of crime (see Farrall et al. 1997; Ferraro and LaGrange 1987). As such, these differences of magnitudes of fear by crime type are ignored if one only asks about "crime" rather than asking about specific types of crime such as robbery, rape, burglary, etc.

Another issue regarding wording of survey questions deals with problems that can arise if questions are too vague. Vagueness in question wording can have general effects on the validity and reliability of survey data as well. Several of the critiques of the standard NCVS fear question have pointed out problems related to its vagueness (e.g., see Ferraro and LaGrange 1987). In particular, the question is criticized as being too vague in three ways. First, in many cases, the reference to "crime" tends to be left out in the question wording; as such, it merely asks respondents how safe they feel walking alone at night in their neighborhoods. Thus, it becomes entirely subjective for the respondents to decide on whether or not to associate their fear with crime or something else. Second, the contextual reference is not always specified and often is not well defined. For example, the term "neighborhood" could mean different things to different people. Thus findings that levels of fear vary among respondents living in the same neighborhood may be an artifact of them imagining different geographic areas. As such, validity of fear of crime measurements could likely be enhanced by providing specific wording about crime and using less ambiguous geographic areas as the reference for the questions. Lastly, it is also noted that the time frame of the scenario—"at night"—is also vague and misses the possibility that levels of fear may be different, for example, right after sundown than at 1 a.m.

Lack of clear context in question wording can also lead to biased results. Yang and Wyckoff (2010) in a study surveying college students' victimization experiences and fear of crime found that questions with clearer time and geographic reference (e.g., walking on campus at night) tended to extract more specific responses, while questions with no reference were more likely to be subject to question order effects (which we will discuss in detail in the following section).

Another critique of the standard NCVS fear question is the reliance on hypothetical situations (Ferraro and LaGrange 1987). Specifically, these critiques note that questions often ask respondents how safe they "would" feel if walking alone at night, rather than how safe they actually feel when doing so (and some variants of the question mix the two together and ask "How safe do you, or would you, feel…"). It is argued that this risks overstating levels of fear by asking people to imagine situations they were not actually in. The critics also note that this problem of using hypothetical scenarios may be worsened by asking about an activity—like walking alone at night—that many people may seldom or never engage in. Thus it is suggested that more accurate measures of fear should aim to ask about episodes of fear of crime that were actually experienced. For instance, asking the respondents whether they were ever fearful of being a crime victim in the past year as Farrall and Gadd (2004a, b) did, or grounding the question in actual experience by starting it with "In your everyday life…" as suggested by Ferraro and LaGrange (1987).

25.1.2.1 Examples of Fear of Crime Wording in Actual Surveys

A key concern that arises after this discussion is just how much current research on fear of crime has possibly been plagued by surveys using questions subject to the wording problems outlined above. As part of an ongoing study (Yang and Hinkle, in progress), we have collected surveys from studies exploring the relationship between fear of crime and victimization (see Footnote 1 and Appendix A). Currently we have identified 35 unique surveys used across the studies we have collected to date, and an examination of these surveys allows us to gauge how many studies of "fear" of crime have really asked about fear of crime and how many have asked about perceptions of safety/risk or other constructs like "worry" about crime.

Out of these 35 surveys, only 14 (40%) include a measure of "real fear"—defined as the question specifically including the terms "fear," "fearful," "afraid", or "scared." Of these 14 items, six asked respondents how fearful or afraid they were walking alone at night, four asked about fear of becoming a victim of a series of specific crimes, two asked about fear of crime/becoming a victim in general (no reference to specific crime types), one asked students about fear of being attacked in

school and one asked respondents whether people in their neighborhoods could leave property out without fearing that it would be damaged or stolen.

The key takeaway is that 60% of the surveys collected above did not include a real measure of emotional fear, despite these surveys being identified through a specific search for academic studies of fear of crime and victimization. The next step is to see what types of questions were asked in place of true fear of crime measures. The most commonly asked type of questions were measures of perceptions of safety. A total of 21 surveys included some form of perceived safety questions—15 of these were studies that did not ask a real fear question, while six of these were surveys that asked a safety question in addition to the fear question. The most common safety items were variations of the standard NCVS question which asked respondents how safe they felt walking alone at night. There were a total of 13 surveys including such a question. In terms of the geographic context of these safe walking at night questions, six referred to the respondent's neighborhood and two provided no geographic reference, and are thus subject to the issue of vagueness discussed earlier. The other five safe walking at night items referred to how safe the respondents' felt in their "area" or "the environment in which you live in." The other eight safety questions included three items which generally asked if respondents felt their neighborhood was safe and five items that asked about specific circumstances such as how safe they felt when home alone at night or whether they avoided unsafe areas during the day because of crime.

The other two types of questions we identified as related to the "fear" issue were measures of perceived risk of victimization and level of "worry" about crime or victimization. There were 10 surveys which included measures of perceived risk. Nine surveys featured questions asking respondents how likely they felt they would become victims of crime in the future, and one question asked them to rate the level of risk in the area in which they lived. Perceived risk questions tended to be asked in combination with fear or safety questions, with only three surveys including a risk measure with no other measures of fear, safety, or worry.

There were also 10 surveys which included measures of respondents' "worry" about crime. Seven questions asked respondents to rate their level of worry about specific types of crime (one of which asked how "anxious" they were rather than how "worried"), one referred to worry about crime in general, one asked how worried they would be staying home alone at night, and one asked how worried they were about their personal safety when walking alone at night. As with risk, questions about worry tended to be asked in combination with the other constructs examined here, with only three surveys including a question about worry.

In sum, our review of "fear" questions used shows a wide variety of question wordings used to measure the construct of fear of crime in studies of fear and victimization. Only 14 out of 35 studies included measures of fear that specifically asked about being fearful or afraid. Not surprisingly, given the impact the NCVS has on the development of other surveys, the most commonly used questions asked respondents how safe they felt walking alone at night. A number of questions also asked respondents to rate their perceptions of risk of becoming victims of crime or their levels of worry about crime and victimization. These questions tended to be asked in combination with safety or fear items; however, there were three cases for each where perceived risk or worry was the only "fear"-related question asked. In addition to having four potentially different constructs (fear, safety, risk, and worry) measured by these items (all of which were interpreted as a measure of fear of crime by at least some studies using these surveys), there are also numerous wording differences within each category as outlined above. With so many variations in how "fear" of crime is measured in surveys, it is not surprising that findings on the topic have long been equivocal and widely debated.

25.1.3 III. Question Ordering

Another equally important issue in survey construction is question order effects. This is a form of bias which may be introduced by the temporal

order in which the material presented potentially leading to different responses. The cause of the question order effects is not due to the ambiguity of questions themselves discussed above. Even questions which avoid the problems with question wording and response options discussed earlier may be affected by question order effects. This is because the order in which questions are asked may influence responses given to items which are placed later in the survey. For instance, keeping with the fear of crime and victimization example we have been using, it is possible that asking victimization questions first could influence the level of fear reported, or vice versa (see Yang and Wyckoff 2010 for review).

Question order effects (sometimes referred to as context effects) are due to the nature of language and how people make sense of conversation. Specifically, Schuman and Presser (1981) argue that "… words and sentences take part of their meaning from the context in which they occur" (p. 30). For surveys, one important context is the order in which questions are asked. The experiences a person is asked to recall earlier in a survey, or the scenarios one is asked to consider first, form the context for answering later questions. The earlier questions may bring to mind memories or imagined situations that in turn influence responses to subsequent questions because those may be memories/situations they would not have considered otherwise. By thinking about the information once in previous questions, this information becomes readily available to the respondents (the recency principle). McClendon and O'Brien (1988), in proposing the recency principle, noted that the location of specific questions (i.e. past victimization experience) relative to general questions (i.e., opinion on how safe your neighborhood is) also makes a difference. They argued that when the general questions are asked last there are larger effects for the specific items closest to the general question and smaller effects for those further away.

The mechanism of the question order effect can be explained through a chain of human cognitive and memory retrieval processes. When we answer a question, the cognitive process involves many stages: interpretation of the questions; retrieval of the relevant information from long-term memory; making judgments based on the information retrieved; and finally, deciding on a proper answer (Tourangeau and Rasinski 1988). The same procedure is used when people complete a survey, a respondent will retrieve information from his or her long-term memory to render a judgment and then provide an answer. However, if this respondent has already answered a question on a similar topic, this previous judgment, already used once and now stored in the short-term memory, will take priority over other information in the long-term memory and will be drawn upon to provide an answer to this new question (Matlin 1994; McClendon and O'Brien 1988; Wyer 1980).

Studies have found that subjects tend to rely on their most recent experience when making judgments regarding their attitudes and beliefs, rather than using other information that may be equally or more relevant (Bem 1972; Bem and McConnell 1970; Wyer 1980). In the context of survey research, Bradburn and Mason (1964) referred to this as the "saliency effect," which occurs when a set of questions concerning a general subject are asked early in the survey. The early and repeat appearance of questions examining the same topic will give salience to the topic and sensitize the subject to the issue, and thus increase the subject's likelihood of showing approval, concern, or worry in subsequent questions about the topic (also see Johnson et al. 2006). Bradburn and Mason (1964) suggested three other possible question order effects including redundancy, consistency, and fatigue. However, these other types of question order effects are not relevant to the focus of the current chapter and thus will not be reviewed here (for more information, see Bradburn and Mason 1964, p. 58; Rossi et al. 1983).

Question order effects are much more common and salient than is often thought. Schuman and Presser (1981) pointed out that in some cases, order effects can elicit as large as 20% marginal differences in the responses given. Moreover, question order effects are complex phenomena and the direction of effects varies. Different arrangement of questions could either

create more consistent results or more contrasting findings, depending on the content of questions. Among the discussions of question order effects, two notions are tested most often and considered important: the assimilation effect (or consistency effect) and the contrast effect (Babbie 1990; Bishop et al. 1985; Schuman et al. 1983b; Schwarz et al. 1991; Strack et al. 1988; Tourangeau et al. 2000). Assimilation (consistency) effects represent a situation where questions presented earlier reinforce respondents to interpret or respond to the target questions in a way similar to the preceding questions. Contrast effects, on the other hand, represent a situation where the order of questions diversifies the responses to target questions relative to the preceding questions by driving them to give responses divergent from those given to the earlier questions (Tourangeau et al. 2000; Schuman and Presser 1981).

25.1.3.1 Assimilation Effects

As mentioned above, assimilation effects occur when the target questions bear information that is unfamiliar to the respondents. Thus, the preceding questions provide respondents with some contextual information and help respondents to interpret the meaning of the unfamiliar questions that follow. Subjects tend to make a connection between the target questions and nearby questions they had just answered in order to determine the nature of questions about unfamiliar topics. Tourangeau and Rasinski (1988) demonstrated such an effect using the Monetary Control Bill as an example. What they found is when there was a block of inflation questions preceding the target question relating to the Monetary Control Bill, respondents tended to associate the two and answer the question assuming it was inflation related. Upon further examination, it seems that the assimilation effect only occurred when the contextual questions were positioned in a cluster rather than scattered throughout the survey. The assimilation effect has important implications for people who want to examine less popular concepts in criminal justice, such as newly implemented policies or practices that might be unfamiliar to general public.

Assimilation effects can also occur with questions regarding value judgments. This type of effect has also been referred to as "part–part consistency," which involves two questions with different contexts but similar logic. Thus, subjects' judgment of one condition will affect their judgment of another. Hyman and Sheatsley (1950) conducted a widely cited split-ballot experiment demonstrating this type of effect and how the order in which questions are presented can have impacts on the results. They presented the following two questions in alternate order to different respondents. They then compared responses of the two halves of the sample to see if the question ordering affected the results.

> **Communist reporter question:** Do you think the United States should let communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?
> **American reporter question:** Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?

In this study, Hyman and Sheatsley found that American respondents were more likely to support letting communist reporters into the United States after having answered the American reporter question first. On the contrary, they were less likely to allow American reporters into communist countries after having answered the Communist reporter question first. Essentially, respondents tried to be fair by keeping the same standards when answering the questions. Schuman et al. (1983a) replicated Hyman and Sheatsley's experiment and used a randomly assigned form in a telephone survey. Their findings confirmed Hyman and Sheatsley conclusion regarding the assimilation effect. Though these types of opposing questions are less often seen in victimization surveys, it is nevertheless important to carefully examine the placement of questions which bear parallel structure in a survey.

25.1.3.2 Contrasting Effects

As described earlier, contrast effects represent a situation where the order of questions results in more contrasting opinions, rather than more consistent opinions, on ratings of general

attitudes. A major cause of contrasting effects is presenting questions concerning a specific domain (e.g., past personal victimization experiences) prior to later questions regarding a general attitude (e.g., how safe is your neighborhood), and thus having the specific questions influence responses to the general questions. Questions with general focus usually emphasize the rating of an omnibus construct or experience, while specific questions focus on a specific dimension of the same overall construct or experience (DeMoranville and Bienstock 2003). As described in cognitive process theory, when answering a general question, respondents need to first define the meaning of the question to help them narrow down its focus (Bradburn and Danis 1984; McClendon and O'Brien 1988). Specific questions are usually written with a very clear focus and context, which makes it easier for respondents to answer without drawing upon other cues. Thus, when both types of questions are presented in the same survey, respondents tend to use specific questions to give context to the general questions, and to direct their attention in order to accurately interpret the general questions (Benton and Daly 1991; McFarland 1981; Schuman and Presser 1981; Sigelman 1981; Wyer 1980). The decision to use both types of questions in the same survey can be a strategic design to assure response accuracy, like the design of the Conflict Tactics Scales (CTS) for intimate partner violence studies (Ramirez and Straus 2006).

In CTS, the earlier questions about intimate partners were worded in a neutral way to increase the likelihood of subjects disclosing their violent actions against their partners, which were assessed in the later questions. Nonetheless, it is important to note that surveys may invoke this principle unintentionally, when the contextual effects created by the questions preceding the attitudinal questions actually affect responses. In both scenarios, the survey results may be altered; however, the priming effect is intentional in the first approach but is unintended in the second scenario and, thus, is more problematic.

Many factors determine whether the question order effect provoked should be in the direction of an assimilation or contrasting effect. Reviewing earlier studies, Tourangeau et al. (2000) summarize two principles determining the direction of question order effects. First, when questions require deep processing of information, then the design of a survey tends to produce contrast effects rather than assimilation effects. Second, vaguely defined scales, extreme scenarios identified in questions, and a clear focus on both contextual and target questions all promote contrast effects (pp. 220, 221). The opposite conditions—clearly defined scales, common situations, and questions without clear focus—promote assimilation effects (Tourangeau et al. 2000; Herr et al. 1983).

25.1.3.3 Question Order Effects in Victimization Research

In victimization research, it is unclear whether having a significant event in the past, such as a victimization experience, affects subjects' responses about perceptions of safety or fear of crime at a later time. Compared to other factors, salient life events or fundamental beliefs are most likely to affect people's decision-making processes. For example, social psychologists have found that events which are memorable, informative, and are recalled with strong emotion have the strongest impact on people's probability judgments (Tyler and Rasinski 1984, p. 309). Thus, by focusing on these salient events, people tend to overestimate the probabilities that related events will actually occur in the future (Carroll 1978). As such, respondents who are focusing on past victimization experiences from first answering such questions may then overestimate likelihood of future victimization and thus report higher levels of fear of crime than they would have absent the victimization questions having been asked first.

In order to explore the impacts of question order effect in victimization surveys, Yang and Wyckoff (2010) conducted a randomized experiment using a telephone survey which included common victimization and fear of crime questions which was administered to a sample of college students. They adapted items from Smith and Hill's (1991) survey testing fear of crime and victimization on campus.

The survey used in the study asked questions *specifically* about personal experiences with, experiences of acquaintances relating to, or knowledge of, victimization on campus, and asked *generally* about whether the respondents felt safe on campus under different circumstances. The experiment randomly assigned respondents to two different groups; the control group received a survey that placed the general questions of perceptions of safety before the specific questions regarding victimization, while the treatment group received a survey where the specific victimization questions preceded the general safety questions.

Overall, Yang and Wyckoff did not find a significant difference between the two groups on four questions relating to perceived safety. However, further examination showed that questions with more detailed reference provided (e.g., walking on campus at night) tended to extract more consistent responses while questions with no reference were more likely to be subject to contrast effects. Additionally, respondents' gender and victimization experience showed significant differences on perceived safety in which female subjects and prior victims perceived the college campus to be less safe than their counterparts. For the two safety questions that contained specific details (walking alone in the day time or after dark), there was no interaction effect between gender/victim status and question arrangement. For the other two questions that asked for a general assessment of safety of the college campus, the study found interaction effects where nonvictims and females demonstrated the expected question order effects.

That is, presenting victimization questions first reduced females and non-victims' perceptions of safety. Females, who are also less likely to be crime victims per crime statistics, were more likely to be influenced by the early presence of victimization questions and report lower levels of perceptions of safety (Baumer 1979; del Carmen et al. 2000; Ferraro 1996; Gordon and Riger 1989; Hale 1996; McConnell 1997; Pain 1995; Rountree and Land 1996; Warr 1985). Nonvictims, who did not have personal victimization experiences, were also more likely to feel unsafe when asked the victimization questions first. Thus for both females and nonvictims the findings suggested that presenting victimization questions first served as a reference which caused them to think about the possibility of being victimized on campus and subsequently led to lower ratings of perceptions of safety. As expected, past victims tended to report lower ratings on perceived safety compared to nonvictims. However, when the victimization questions were presented first, their levels of perceived safety actually increased to a level comparable to that of the nonvictims. Yang and Wyckoff (2010) argued for the potential for interaction effects between respondent status and question ordering—perhaps people with different experiences with victimization use different anchors to make judgments in determining their levels of fear of crime.[3]

This conditional question order effect can perhaps explain the inconsistent findings in studies of the relationship between past victimization and one's level of fear. Among the extensive literature focusing on fear of crime, some studies argue that having been victimized increases fear of crime as past victims tend to be more distrustful, more cautious and more fearful (e.g. Balkin 1979; Skogan 1987; Smith and Hill 1991). Rountree and Land (1996) found that previous burglary victimization experiences increased one's fear of future burglary victimization and produced a general unsafe feeling toward one's living environment. Though this finding sounds reasonable, others have concluded just the opposite. Studies by Sparks et al. (1977) and Smith (1976) actually found a negative relationship between victimization and fear of crime; that is, some people show a reduced fear of crime after being victimized. Mayhew (1984), however, using the 1982 British Crime Survey, found only a tenuous relationship between victimization and fear of crime. The source of the inconsistent findings could be variations between samples, differences in survey design (Lynch 1996;

[3] A similar mechanism was found in Smith and Jobe's (1994) study, in which they argued that generic knowledge of food consumption worked as an anchor to produce the base rate while the episodic memories help to adjust the base rate.

Murphy 1976)[4] or variations in the definition/measurement of "fear of crime."

Others have further pointed out differential question order effects by survey administration methods. Willits and Ke (1995) argued that response to mail questionnaires would be less likely to show question order effects than telephone surveys, since respondents have more time to consider their answers as well as the opportunity to preview and review the items. Strack et al. (1988), on the other hand, found that the order effect reduced in telephone surveys under the circumstance that subjects were told about the areas to be covered by the survey in advance. Willits and Ke (1995) further concluded that though placing specific items first induced the responses to the general questions, browsing the questionnaire before answering any question nonetheless reduced the potential order effect. Therefore, it is reasonable to believe that research using mail surveys may reach different conclusions than when using telephone surveys.

In our research in progress, we are examining victimization surveys closely to see if survey design could contribute to the inconsistent findings on the relationship between fear and victimization. Out of the 35 victimization surveys we identified in our ongoing study (see Appendix A), we are currently able to determine the specific order the victimization questions and the "fear" questions were asked for 27 of them.[5] Of these 27, 16 (59%) ask the fear/safety/risk/worry item first, and 11 (41%) ask the victimization questions first. These different practices across surveys offer some direction for further examination to see to what extent the inconsistent findings in research testing the link between victimization and fear of crime may be the result of survey design regarding question ordering.[6]

25.2 Conclusions

Surveys have become commonplace in the field of criminology and criminal justice and are being used as tools in studies of a variety of topics such as evaluating police services and interventions, detecting trends in community well-being (e.g., general social surveys), and measuring the prevalence of victimization for both the general population (e.g., NCVS) and subgroups such as students. With so much of the knowledge base in our field being driven by findings from survey data, it is crucial to be sure our measures are accurate and reflect what we think they are measuring. As such, in order to gather accurate information it is extremely important to ensure that questions are designed in ways that actually measure the phenomenon of interest, and do so with high validity and reliability.

To understand some common problems in survey designs, three possible issues were identified and discussed in this chapter: response option effects, question wording effects, and question order effects. In terms of response options, most victimization surveys tend to omit middle alternatives to attitudinal questions based on a belief that the offering of middle alternatives might draw respondents away from their actual attitudes. The Utah Victimization survey results show that the percentage of respondents who chose middle alternatives when provided is actually very small. Thus, offering middle alternatives does not necessarily lead to biased results in surveys of victimizations and fear of crime. On the other hand, the option DK or "unknown" is generally included in victimization surveys. The finding regarding the effect of including the DK option is similar to that for offering middle alternatives. That is, though some respondents will choose DK or middle

[4] The order in which questions were presented varies from study to study. For example, the British Crime Survey presents fear of crime questions (general questions) before the specific victimization questions, while Rountree and Land (1996) provided the specific victimization questions first. Smith and Hill (1991), however, presented questions without specifying the order of presentation of the questions.

[5] For some surveys we have merely identified them and been able to code question wording from articles or reports published analyzing their data, but cannot yet code question ordering as we are still attempting to obtain the full survey instruments.

[6] Yang and Hinkle (in progress) are currently conducting research to further examine whether the empirical evidence supports this speculation.

alternatives when the options are provided, the proportions of respondents drawn to these options is pretty evenly divided among subjects who would otherwise show positive or negative attitudes. Thus, the overall ratio of "pros" and "cons" should not be affected dramatically and should not cause much bias to research findings.

The effects of question wording are best illustrated by the fear of crime example. This is a great example demonstrating the importance of question wording as it shows that a large portion of research in this area is hampered by using variations the standard NCVS "how safe do you feel walking alone at night" question. This question has been shown to have a host of wording problems that call into question its validity and thus challenge the conclusions we can draw from the body of research in this area. For instance, even the long-accepted finding that the elderly are more fearful of crime is now suspect given that at least one study which asked specifically about fear rather than safety did not find an age effect (Farrall and Gadd 2004a, b). Thus, it is important for researchers to closely examine the wording of questions to make sure they capture intended information. For instance, the examples above show that asking people "how safe" they feel is not the same as asking them "how fearful" or "how afraid" they are of crime. As such, researchers must decide when designing their survey whether they are really interested in measuring fear of crime or perceptions of safety (or other constructs such as perceptions of victimization risk or worry about crime) and design their questions accordingly.

Question order effects, on the other hand, are perhaps the one survey design issue that is most often overlooked, but could have important impacts on survey results. As described in the previous section, the order in which questions are asked in a survey could lead to either assimilation effects which amplify the response tendency or contrast effects which lead to divergent results among respondents. Moreover, question order effects are more likely to occur with regard to questions involving general attitudes with no specific reference provided. Yang and Wyckoff (2010) also found that respondents' characteristics can also influence their susceptibility to order effects (such as their gender, education level, etc.). Thus, it is important to take into account both question ordering itself, as well as its potential interaction with respondents' characteristics, when examining data for any question order bias. An examination of actual victimization surveys showed that there is no consistent practice on whether to present victimization questions or fear/perceived safety/risk questions first in surveys. Thus, the lack of knowledge on question order effects could partly explain the lack of consensus in research on whether past victimizations are related to higher levels of fear of crime. Overall, the review of literature shows that questions with more specific reference or context, clearer wording, and clear scales are less likely to generate biased results.

The examples cited throughout this chapter are of course limited to survey research on fear of crime and victimization we identified, but they nonetheless provide a clear illustration of the current practices in victimization survey design and thus highlight the importance of being cautious when designing surveys. Moreover, they reveal a number of common issues with survey design which are relevant to surveys on every topic in criminology and criminal justice and show the importance of considering factors such as what response options to provide, how to word questions, and how question ordering may affect responses when attempting to construct surveys. Paying attention to such details is a crucial part of generating the type of quality data on crime and related issues needed to continue advancing our field.

Appendix A: List of 35 Example Surveys

The Table provides a list of the 35 surveys of fear of crime and victimization used in the examples in this chapter (see Footnote 1 for more details on the collection of these surveys). These do not include the state-level surveys used as examples as those are self-explanatory. Many of these 35 surveys were unnamed and are just referred to by the author's name or a descriptive name we gave them. Where possible we provide a link to a website for the survey (i.e., for major surveys like the ICVS or British Crime Survey), for all others we provide a reference to a publication that uses data from the survey.

Survey name	Citation for single use surveys
Beck and Robertson survey	Beck, A., & Robertson, A. (2003). Crime in Russia: Exploring the link between victimization and concern about crime. *Crime Prevention and Community Safety: An International Journal, 5*(1), 27–46
British crime survey	http://www.homeoffice.gov.uk/science-research/research-statistics/crime/crime-statistics/british-crime-survey/
Cambodian UNICVS	Broadhurst, R., & Bouhours, T. (2009). Policing in Cambodia: Legitimacy in the making*? Policing and Society, 19*(2), 174–190
Canadian general social survey	http://www.statcan.gc.ca/dli-ild/data-donnees/ftp/gss-esg-eng.htm
Canadian urban victim survey	Weinrath, M., & Gartrell, J. (1996). Victimization and fear of crime. *Violence and Victims, 11*(3), 187–197
Caribbean victimization survey	Painter, K. A., & Farrington, D. P. (1998). Criminal victimization on a Caribbean island. *International Review of Victimology, 6*(1), 1–16
Chiricos, Padgett and Gertz survey	Chircos, T., Padgett, K., & Gerta, M. (2000). Fear, TV news and the reality of crime. *Criminology, 38*(3), 755–785
Chockalingam and Srinivasan survey	Chockalingam, K., & Srinivasan, M. (2009). Fear of crime victimization: A study of university students in India and Japan. *International Journal of Victimology, 16*(1), 89–117
Community living and integration survey	Ditton, J., & Chadee, D. (2005). People's perceptions of their likely future risk of criminal victimization. *British Journal of Criminology, 46*(3), 505–518
Survey of community, crime and health	http://www.icpsr.umich.edu/icpsrweb/NACDA/studies/04381
Dukes and Hughes survey	Dukes, R. L., & Hughes, R. H. (2004). Victimization, citizen fear and attitudes towards police. *Free Inquiry in Creative Sociology, 32*(1), 50–57
European union international crime survey	http://www.europeansafetyobservatory.eu/euics_fi.htm
European social survey	http://www.europeansocialsurvey.org/
General social survey	http://www.norc.uchicago.edu/GSS+Website/
International crime victimization survey (ICVS)	http://rechten.uvt.nl/icvs/
International labor organization survey	Dammert, L., & Malone, M. F. T. (2003). Fear of crime or fear of life? Public insecurities in Chile. *Bulletin of Latin American Research, 22*(1), 79–101
Keilman and David crime survey	Kleinman, P. H., & David, D. S. (1973). Victimization and perception of crime in a ghetto community. *Criminology, 11*(3), 307–343
Miethe seattle crime survey	Roundtree, P. W., & Land, K. C. (Rountree and Land 1996). Perceived risk versus fear of crime: Empirical evidence of conceptually distinct reactions in survey data. *Social Forces, 74*(4), 1353–1367

(continued)

Survey name	Citation for single use surveys
Italian multipurpose survey	Miceli, R., Roccato, M., & Rosato, R. (2004). Fear of crime in Italy: Spread and determinants. *Environment and Behavior, 36*(6), 776–789
National crime victimization survey (NCVS)	http://bjs.ojp.usdoj.gov/index.cfm?ty=dcdetail&iid=245
NCVS (school crime supplement)	http://bjs.ojp.usdoj.gov/index.cfm?ty=dcdetail&iid=245
Nikolic-Ristanovic crime survey	Nikolic-Ristanovic, V. (1995). Fear of crime in Belgrade. *International Review of Victimology, 4*(1), 15–31
Observatory of the North-West survey	Amerio, P., & Roccato, M. (2007). Psychological reactions to crime in Italy: 2002–2004. *Journal of Community Psychology, 35*(1), 91–102
Oklahoma city survey	Baba, Y., & Austin, D. M. (1989). Neighborhood environmental satisfaction, victimization and social participation determinants of perceived neighborhood safety. *Environment and Behavior, 21*(6), 763–780
Queensland crime victims survey	http://www.police.qld.gov.au/programs/cvs/
Reid, Roberts and Hillard survey	Reid, L. W., Roberts, J. T., & Hillard, H. M. (1998). Fear of crime and collective action: An analysis of coping strategies. *Social Inquiry, 68*(3), 312–328
Scottish crime and victimization survey	http://www.esds.ac.uk/findingData/scsTitles.asp
Smith and Torstensson survey	Smith, W., & Torstensson, M. (1997). Gender differences in risk perception and neutralizing fear of crime. *British Journal of Criminology, 37*(4), 608–634
Stiles, Halim and Kaplan survey	Stiles, B. L., Halim, S., & Kaplan, H. B. (2003). Fear of crime among individuals with physical limitations. *Criminal Justice Review, 28*(2), 232–253
The Fear of crime in America	Ferraro, K. F. (1995). *Fear of crime: Interpreting victimization risk*. Albany, NY: State University of New York Press (Note: See Appendix A for survey details)
Theall, Sterk and Elifson crime survey	Theall, K. P., Strek, C. E., & Elifson, K. W. (2009). Perceived neighborhood fear and drug use among young adults. *American Journal of Health and Behavior, 33*(4), 353–365
Trinidad crime survey	Chadee, D., & Ditton, J. (1999). Fear of crime in Trinidad: A preliminary empirical research note. *Caribbean Journal of Criminology and Social Psychology, 4*(1), 112–129
Tseloni and Zarafonitou crime survey	Tseloni, A., & Zarafonitou, C. (2008). Fear of crime and victimization: A multivariate multilevel analysis of competing measurements. *European Journal of Criminology, 5*(4), 387–409
Violence against women survey	http://www.nij.gov/nij/pubs-sum/172837.htm
Women's safety survey	http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/F16680629C465E03CA256980007C4A81/$File/41280_1996.pdf

References

Andersen, R., Kasper, J., & Frankel, M. R. (1979). *Total survey error: Applications to improve health surveys*. San Francisco, CA: Jossey-Bass.

Ayidiya, S., & McClendon, M. (1990). Response effects in mail surveys. *Public Opinion Quarterly, 54*(2), 229–247.

Babbie, E. (1990). *Survey Research methods*. Belmont, CA: Wadsworth, Inc.

Bachman, R., & Schutt, R. K. (2008). *Fundamentals of research in criminology and criminal justice*. Thousand Oaks, CA: Sage.

Bachman, R., & Schutt, R. K. (2010). *The practice of research in criminology and criminal justice* (4th ed.). Thousand Oaks, CA: Sage.

Balkin, S. (1979). Victimization rates, safety, and fear of crime. *Social Problems, 26*, 343–358.

Baumer, T. L. (1979). Research on fear of crime in the United States. *Victimology, 3*, 254–263.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology, 14*, 23–31.

Benton, J. E., & Daly, J. (1991). A question order effect in a local government survey. *Public Opinion Quarterly, 55*, 640–642.

Bradburn, N., & Danis, C. (1984). Potential contributions of cognitive research to survey questionnaire design. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 101–129). Washington, DC: National Academy Press.

Bradburn, N. M., & Mason, W. M. (1964). The effect of question order on responses. *Journal of Marketing Research, 1*(4), 57–61.

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly, 51*(2), 220–231.

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1985). The importance of replicating a failure to replicate: Order effect on abortion items. *Public opinion quarterly, 49*, 105–114.

Cantor, D., & Lynch, J. P. (2000). Self-report surveys as measures of crime and criminal victimization. *Criminal justice 2000: Measurement and analysis of crime and justice* (Vol. 4, pp. 263–315). Washington, DC: United States Department of Justice.

Carroll, J. S. (1978). The effect of imaging an event on expectations for the event. *Journal of Experimental Social Psychology, 14*, 88–97.

Cochran, W. G. (1953). *Sampling techniques*. New York: Wiley.

Dalenius, T. (1974). Ends and means of total survey design. Report in ''*Errors in Surveys*.'' Stockholm: Stockholm University.

del Carmen, A., Polk, O. E., Segal, C., & Bing, R. L. I. I. I. (2000). Fear of crime on campus: Examining fear variables of CRCJ majors and nonmajors in pre- and post-serious crime environments. *Journal of Security Administration, 21*(1), 21–47.

Deming, W. E. (1944). On errors in surveys. *American Sociological Review, 9*, 359–369.

DeMoranville, C. W., & Bienstock, C. C. (2003). Question order effects in measuring service quality. *International Journal of Research in Marketing, 20*, 217–231.

Farrall, S. (2004). Revisiting crime surveys: Emotional responses without emotions? *International Journal of Social Research Methodology, 7*, 157–171.

Farrall, S., & Ditton, J. (1999). Improving the measurement of attitudinal responses: An example from a crime survey. *International Journal of Social Research Methodology, 2*(1), 55–68.

Farrall, S., & Gadd, D. (2004a). Rearch note: The frequency of fear of crime. *British Journal of Criminology, 44*, 127–132.

Farrall, S., & Gadd, D. (2004). *Fear today, gone tomorrow: Do surveys overstate fear levels?* Unpublished manuscript. Retrieved April 15, 2008, from http://www.istat.it/istat/eventi/2003/perunasocieta/relazioni/Farral_abs.pdf.

Farrall, S., Bannister, J., Ditton, J., & Gilchrist, E. (1997). Questioning the measurement of the 'fear of crimefear of crime:' Findings from a major methodological study. *British Journal of Criminology, 37*(4), 658–679.

Fattah, E. A. (1993). Research on fear of crime: Some common conceptual and measurement problems. In W. Bilsky, C. Pfeiffer, & P. Wetzels (Eds.), *Fear of crime and criminal victimisation*. Stuttgart: Ferdinand Enke Verlag.

Ferraro, K. F. (1995). *Fear of crime: Interpreting victimization risk*. Albany, NY: State University of New York Press.

Ferraro, K. (1996). Women fear of victimization: The shadow of sexual assault. *Social Force, 75*(2), 667–690.

Ferraro, K. F., & LaGrange, R. (1987). The measurement of fear of crime. *Sociological Inquiry, 57*(1), 70–101.

Gibson, C., Shapiro, G.M., Murphy, L.R., & Stanko, G.J. (1978). *Interaction of survey questions as it related to interviewer-respondent bias*: *Proceedings of the Section on Survey Research Methods*. Washington, D.C.: American Statistical Association.

Gordon, M., & Riger, S. (1989). *The female fear*. New York, NY: Free Press.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849–879.

Hale, C. (1996). Fear of crime: A review of the literature. *International Review of Victimology, 4*, 79–150.

Heaps, C. M., & Henley, T. B. (1999). Language matters: Wording considerations in hazard perception and warning comprehension. *The Journal of Psychology, 133*(3), 341–351.

Herr, P., Sherman, S. J., & Fazio, R. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology, 19*, 323–340.

Hyman, H. H., & Sheatsley, P. B. (1950). The current status of American public opinion. In J. C. Payne (Ed.), The teaching of contemporary affairs: Twenty-first yearbook of the National Council of Social Studies (pp. 11–34). New York: National Education Association.

Johnson, P. K., Apltauer, K. W., & Lindemann, P. (2006). Question-order effects in the domains of language and marriage. *Columbia Undergraduate Science Journal, 1*(1), 43–49.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–350.

Keane, C. (1992). Fear of crime in Canada: An examination of concrete and formless fear of victimizationvictimization. *Canadian Journal of Criminology*, *34*, 215–224.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Kury, H. (1994). The influence of the specific formulation of questions on the results of victim studies. *European Journal on Criminal Policy and Research, 2*(4), 48–68.

Kury, H., & Ferdinand, T. (1998). The victim's experience and fear of crime. *International Review of Victimology, 5*(2), 93–140.

Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society, 25*, 85–86.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.

Lorenz, F., Saltiel, J., & Hoyt, D. (1995). Question order and fair play: Evidence of evenhandedness in rural surveys. *Rural Sociology, 60*, 641–653.

Lynch, J. P. (1996). Clarifying divergent estimates of rape from two national surveys. *The Public Opinion Quarterly, 60*(3), 410–430.

Marx, K. (1880). Introduction to the French edition of Engels'Socialism: Utopian and Scientific. F. Engels, Socialisme utopique et socialisme scientifique, Paris.

Matlin, M. W. (1994). *Cognition* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Maxfield, M. G., & Babbie, E. R. (2007). *Research methods for criminal justice and criminology*. Belmont, CA: Wadsworth Publishing.

Mayhew, W. P. (1984). *The effects of crime: Victims, the public and fear*. Strasbourg, France: Council of Europe.

McClendon, M. J., & O'Brien, D. J. (1988). Question-order effects on the determinants of subjective well-being. *Public Opinion Quarterly, 52*, 351–364.

McConnell, E. H. (1997). Fear of crime on campus: A study of a southern university. *Journal of Security Administration, 20*(2), 22–46.

McFarland, S. G. (1981). Effects of question order on survey responses. *Public Opinion Quarterly, 45*, 208–215.

Murphy, L. R. (1976). *Effects of attitude supplement on NCS-cities sample victimization data*. Washington, DC: U.S. Bureau of the Census, Demographic Surveys Division, National Crime Surveys Branch.

Pain, R. (1995). Elderly women and fear of violent crime: The least likely victims? A recommendation of the extent and nature of risk. *British Journal of Criminology, 35*(4), 584–598.

Ramirez, I. L., & Straus, M. A. (2006). The effect of question order on disclosure of intimate partner violence: An experimental test using the conflict tactics scales. *Journal of Family Violence, 21*(1), 1–9.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.). (1983). *Handbook of survey research*. Orlando, FL: Academic Press.

Rountree, P. W., & Land, K. C. (1996). Burglary victimization, perceptions of crime risk, and routine activities: A multilevel analysis across Seattle neighborhoods and census tracts. *Journal of Research in Crime and Delinquency, 33*, 147–180.

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly, 5*, 91–92.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Orlando, FL: Academic Press Inc.

Schuman, H., Presser, S., & Ludwig, J. (1983a). Context effects on survey responses to questions about abortion. *Public Opinion Quarterly, 45*(2), 216–223.

Schuman, H., Kalton, G., & Ludwig, J. (1983b). Context and contiguity in survey questionnaires. *Public Opinion Quarterly, 47*(1), 112–115.

Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55*, 3–23.

Sigelman, L. (1981). Question-order effects on presidential popularity. *Public Opinion Quarterly, 45*, 199–207.

Skogan, W. E. (1987). The impact of victimization on fear. *Crime and Delinquency, 33*(1), 135–154.

Smith, D. L. (1976). The aftermath of victimization: Fear and suspicion. In E. Viano (Ed.), *Victims and society*. Washington, DC: Visage Press.

Smith, L. N., & Hill, G. D. (1991). Victimization and fear of crime. *Criminal Justice and Behavior, 18*, 217–239.

Smith, A., & Jobe, J. (1994). Validity of reports of long-term dietary memories: Data and a model. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 121–140). New York: Springer.

Sparks, R. F., Glenn, H. G., & Dodd, D. G. (1977). *Surveying victims*. Chichester: Wiley.

Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology, 18*, 429–442.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions*. San Francisco, CA: Jossey-Bass Inc.

Thompson, C. Y., Bankston, W. B., & St. Pierre, R. (1992). Parity and disparity among three measures of fear of crime: A research note. *Deviant Behavior: An Interdisciplinary Journal, 13*, 373–389.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299–314.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tyler, T. R., & Rasinski, K. (1984). Comparing psychological images of the Social perceiver: Role of perceived informativeness, memorability, and affect in mediating the impact of crime victimization. *Journal of Personality and Social Psychology, 46*(2), 308–329.

Walonick, D. S. (1993). Do researchers influence survey results with their question wording choices? (Doctoral Dissertation). Walden University.

Warr, M. (1985). Fear of rape among urban women. *Social Problems, 32*(3), 238–250.

Willits, F. K., & Ke, B. (1995). Part whole question order effect: Views of rurality. *Public Opinion Quarterly, 59*, 392–402.

Wyer, R. S, Jr. (1980). The acquisition and use of social knowledge: Basic postulates and representative research. *Personality and Social Psychology Bulletin, 6*, 558–573.

Yang, S., & Wyckoff, L. A. (2010). Perceptions of safety and victimization: Does survey construction affect perceptions? *Journal of Experimental Criminology, 6*(3), 293–323.

What Would You Do? Conducting Web-Based Factorial Vignette Surveys

26

Hadar Aviram

Much of the skill and challenge of conducting surveys involves the need to engage respondents deeply with the survey topic. Factorial surveys, or vignette surveys, offer a unique opportunity to do so. Rather than asking respondents about their lives and experiences, vignettes solicit their thoughts and opinions on hypothetical scenarios, which are crafted and manipulated by the researcher to control for specific variables. While this unique research design overcomes some challenges presented by other surveys, it presents challenges of its own, and requires care and skill. This chapter aims to provide a tentative step-by-step guide to conducting vignette surveys on the web, offering advice on (and some examples of) the design, recruitment, and analysis of such surveys.

26.1 Why Vignettes?

Vignettes are "artificially constructed case descriptions, presented for respondents to consider and to report what they would have done in the circumstances" (Sapsford 1999, p. 142). The appeal of vignette surveys lies in their unique design. They are based on a clean experimental premise: The researcher presents respondents with a hypothetical scenario and asks them for their opinions or reactions to the scenario. Not all respondents, however, are exposed to the same scenario. The researcher creates multiple subtle variations on the scenario, controlling for the variables she is interested in. These variations are randomly assigned to respondents, thus allowing statistical testing of the impact of such variations on respondents' reactions to the scenario (Raffery 2001).

The usage of vignettes offers several advantages as a research design, which stems from its experimental nature. First, it allows the researcher a high degree of control over the variables tested. Since all control over the text of the scenario lies with the researcher, and the scenarios are randomly assigned, running regressions and testing hypotheses do not require the assumptions involved in analyzing panel data and the like. The fact that the scenarios are similar save for the intentional variations creates "clean" comparisons, which would require a variety of statistical assumptions if respondents were asked about actual past events.

Second, vignettes allow access to invisible domains of human experience. Confronting respondents with hypothetical situations allows researchers to target groups they would not have otherwise been able to identify. For example, it is much easier to interview respondents about their willingness to engage in civil litigation (Aviram and Leib 2010), tax evasion (Thurman 1989), organizational dishonesty (Elis and Simpson 1995; Jehn and Jonsen 2010),

H. Aviram (✉)
Hastings College of the Law,
University of California, 200 McAllister Street,
San Francisco, CA 94102, USA
e-mail: hadar.aviram@gmail.com

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_26, © Springer Science+Business Media New York 2012

end-of-life planning (Denk et al. 1997), or police reporting (Aviram and Persinger 2012), than to locate people who have already done so. Moreover, in situations in which there are real dilemmas, vignettes can capture professionals who would choose different courses of action across an array of choices (Taylor 2006; Hughes and Huby 2002; Ludwick and Zeller 2001).

A third and related advantage pertains to the ability to raise sensitive issues through hypothetical scenarios. This is particularly important in studies that examine issues of victimization and vulnerability (Denk et al. 1997; Leighton 2010; Jehn and Jonsen 2010; Aviram and Persinger 2012; Alexander and Becker 1978). Careful crafting of a vignette allows creating some distance between the scenario and the respondent, providing the latter with a safer, more comfortable cognitive environment for his or her reply.

Finally, as a cross between experimental designs and survey research (Dülmer 2007), factorial vignette studies are a less costly alternative in fields in which conducting a full-blown experiment would be too expensive (Shooter and Galloway 2010).

There are, however, some drawbacks and caveats associated with the experimental nature of vignette surveys. These stem from the fact that respondents are commenting on hypothetical situations, not actual experiences, which create difficulties extrapolating as to their realistic responses (Abbott and Sapsford 1993). It is not always possible to guess how this might affect response bias, but some concerns may involve the fact that respondents might be more emotionally removed from situations they have not experienced in person, and therefore might appear to be more proactive, or purport to respond in ways that they perceive as more ethical or more courageous. Also, the success of a vignette survey depends on the respondents' imaginations and their ability to place themselves in the hypothetical situation, which, of course, differs among respondents. These concerns mean that vignettes must be crafted with much care and are crucial for the successful conduction of the survey, as well as the ability to learn and generalize from it.

26.2 Potential Topics for Vignette Surveys

Vignette surveys are potentially useful in a myriad of fields. They are particularly appealing for studying situations that present moral and ethical dilemmas. Vignette research designs have been used extensively in sociology, social psychology, medical sciences, public policy, behavioral economics, empirical philosophy, criminology, and criminal justice.

Scholars, whose interests lie at the crossroad of law and social science, use vignettes to tap into the decision-making process whether to pursue legal action to remedy a situation or a dispute. This chapter discusses vignette survey technique using two examples from recent studies conducted by the author of this chapter: A vignette survey measuring people's perceptions and responses to domestic violence situations (Aviram and Persinger 2012) and a vignette survey examining the impact of relationships on the decision to invoke legal action (Aviram and Leib 2010). In both cases, respondents were directed to a web survey in which they were confronted with a hypothetical scenario; the aim was to test which variables accounted for respondents' likelihood of resorting to the law to resolve their problems.

In the domestic violence study, respondents read a scenario describing a domestic violence incident that occurred to a friend of theirs. The scenarios varied by the gender of the parties (male aggressor, female victim; female aggressor, male victim; same-sex male couple; same-sex female couple), by the extent of reciprocity in the violence, and, for the same-sex scenarios, by the existence of an outing threat by the aggressor. Respondents then answered questions about possible solutions to the problem in general, and their inclination to report the incident to the police.

In the friendship and disputes study, each respondent was confronted with five different scenarios describing everyday disputes. Each dispute varied in the relationship between the parties: good friend, co-worker, or acquaintance.

Respondents were asked questions pertaining to their perception of the dispute (a personal disappointment, a moral failure, or a legal violation), and were provided a scale for reporting the extent to which they would be willing to pursue different types of recourse, legal or non-legal, to the dispute.

26.3 Step I: Configuring the Research Design: What Do You Want to Study?

The first step in crafting a vignette survey is deciding on the variables to be tested. The dependent variables—the respondents' opinions or courses of action—are the questions that respondents will be asked following the scenario. In the domestic violence study, we were mainly interested in whether respondents would report domestic violence incidents to the police, but we also wanted to know whether they would recommend doing so to others, as well as the extent to which they would support alternative courses of action, such as anger management for the aggressor, therapy for the victim, group counseling, separation, or reconciliation. The answers to these questions, we hypothesized, would depend upon a variety of demographic and factual variables. Our focus, however, was on a particular aspect of domestic violence: its occurrence in unconventional gender settings. We wanted to know whether respondents would be more likely to report incidents occurring in a traditional gender setting—a male aggressor and a female victim—than incidents occurring in same-sex relationships or situations in which the aggressor was a woman and the victim was a man. Anecdotal evidence suggested that domestic violence was not a rare occurrence in these unconventional settings, but that the disturbing problem of underreporting such incidents was more severe when the setting deviated from the traditional scenario. For our purposes, this meant that the domestic violence scenario used would include all possible gender combinations. Rather than one independent variable accepting four values (one for each combination), we decided to include two independent variables: gender of the aggressor and gender of the victim.

We were interested in other issues as well, such as violent behavior in the past, seriousness of injuries, and the like, but for reasons explained below, budgetary constraints required focusing on variables related to the gender settings, such as mutual violence (which, as we hypothesized, might lead to blur the aggressor/victim divide and the perception of the incident as domestic abuse rather than "just fighting") and, for same-sex scenarios, explicit "outing" threats on the part of the aggressor, which on one hand might lead to a more serious perception of the incident, and on the other hand might dissuade respondents from reporting such incidents, due to fear that the act of reporting would lead to substantial harm to the victim beyond the abuse.

In the friendship study, we had two dependent variables: the perception of the dispute and the potential recourses (including, but not limited to, legal action). Our independent variable was the relationship between the parties to the dispute. Based on insights from law and society scholarship, the expectation was that people would be less inclined to perceive disputes with good friends as legal matters (they might be more likely to perceive them as personal disappointments or moral failures) and also enthusiastic about litigating such disputes. However, we also suspected that the nature of the relationship, and not just its intensity, might yield differences in perceptions and recourses. Therefore, a nominal independent variable with three values: good friend, co-worker, and acquaintance were chosen.

Considering the statistical analysis later on, it is usually best to pick independent variables that accept two or, at most, three values. This guideline is particularly helpful when dealing with multiple independent variables (and is not unrelated to budgetary limitations, as explained below). The analysis is greatly simplified if the study is configured as containing multiple independent variables of binary values, rather than fewer variables with more values. However, in cases in which various values are

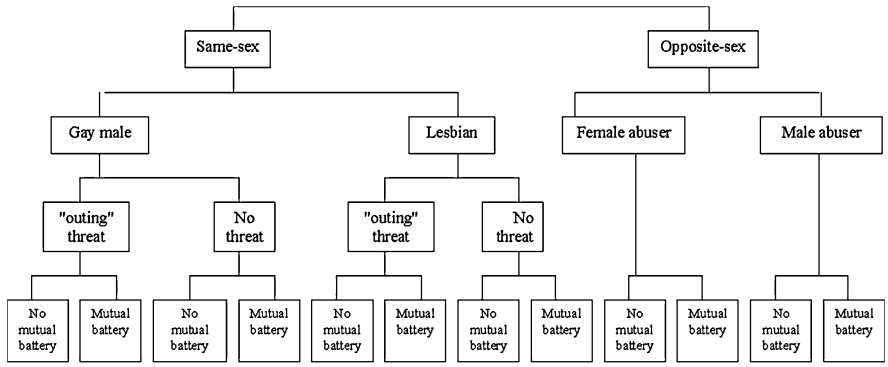


Fig. 26.1 Vignette structure for domestic violence (adopted from Aviram and Persinger 2012)

necessary, dummy variables can later be used to reconfigure categorical variables, though this requires additional steps in "cleaning" the database, and discerning the exact effects requires more sophisticated statistical analysis.

26.4 Step II: Get Real: Designing Around Budget Constraints

Budgetary constraints are always a serious issue in survey research, impacting various decisions, including sample size and length of survey (Salant and Dillman 1994). In the case of web surveys, while a surveyor's participation on the phone or in person is unnecessary, there are costs associated with respondent recruitment and programming. Some programming expertise is particularly valuable in vignette surveys, whose structure requires randomization of options, ordering, and other special features. If the researcher does not possess programming expertise, assistance in programming should be budgeted for as well. The unique feature of vignette surveys, however, is that the number of variations adopted in the hypothetical scenario has an exponential effect on the number of subjects required for an adequate statistical analysis. Consider, for example, the structure of our domestic violence survey portrayed in Fig. 26.1.

Parametric and non-parametric tests require a minimum *n* in each category. When taking into account the demographic questions included in the survey, the *n* increases. As can be seen in Fig. 26.1, each of the added variables doubled the number of scenarios, thus doubling the number of respondents needed (save for the "outing" threat variable, which only doubled the number of respondents assigned the same-sex scenarios). In the domestic violence survey, budget limitations meant foregoing the ability to control for previous violent incidents, the respective size of the parties, and the nature of the relationship between the parties.

Budget limitations impacted our design for the friendship study, too. We wanted to examine the impact of friendship on pursuing legal action in a set of five different scenarios: An unreturned loan, a rental dispute, a quibble over parking tickets, the theft of a business idea, and a dispute over season tickets jointly purchased. Presenting our respondents with only one randomly selected scenario would mean increasing our *n* five-fold. Our decision, therefore, was to present all scenarios to each respondent, randomizing the order in which they were received, and compensate for this choice at the analysis phase.

26.5 Step III: Crafting Good Vignettes

Having formed hypotheses and decided on variables within the research budget, the time comes to flesh out the hypothetical scenario for the survey. A few guiding principles might be helpful when doing so.

First, it is important to create realistic scenarios. In order for the vignette to be a useful and valid simulator of actual human behavior, the situation described need not be too much of a stretch for the respondents. There are various possible sources for everyday hypothetical situations. Previous research on the topic might provide some insights into the scenarios worth examining. For example, in our the friendship study, the season tickets scenario originated from first-hand evidence in one coauthor's previous work, in which he encountered a surprising amount of litigation around this issue. Similarly, the scenario in which one party lends a car to the other party only to later receive $500 worth of parking tickets emerged from real-life stories that Aviram and Leib (2010) heard anecdotally from friends (incidentally, my coauthor lent me his car when he went on sabbatical).

It is important, however, not to assume that a wide spectrum of respondents will connect with a situation that the research team happens to think common and mundane. Socioeconomic and cultural assumptions sometimes find their way into the crafted scenario, and care must be taken to avoid them. Since respondents will differ from each other, the scenario should be generally enough so as not to alienate any group of respondents. Pilot surveys, as well as circulating drafts of the scenarios among colleagues prior to deploying the survey, can be helpful in eliminating such biases.

In the domestic violence survey, we took great care not to imply a particular demographic milieu. Our basic scenario was as follows:

> *Your best friend, Jenny, has been in a two-year relationship with Ken.*
>
> *One night Jenny asks you to come pick her up because she and Ken had a fight. When you arrive Jenny is scraped up, has a black eye and a sprained shoulder. After you leave the couple's home, concerned, you ask what happened.*
>
> *Jenny tells you that she and her boyfriend, Ken, attended a party where they had too much to drink. Jenny explains that after the party Ken accused her of flirting with another man, yelled at her, calling her a liar and a cheater. Jenny admits that Ken was so angry that he lost control and shoved her down the stairs in their home, which resulted in the bruises and injured shoulder. According to Jenny, afterward, Ken was very sorry for pushing her and asked her not to tell anyone what had happened.*
>
> *You have noticed unexplained bruising on Jenny's forearms in the past. Jenny says that sometimes she is afraid of Ken, but she says it's her own fault for provoking him.*

In crafting the scenario, we were careful to avoid the topic of marriage (so as to avoid controversy in the same-sex scenarios), as well as to avoid a detailed layout of the couple's home, which could be a socioeconomic marker (the stairs could be in a house, or lead to an apartment). In the heterosexual scenarios we avoided the word "partner".

A variation of the scenario, occurring between lesbian partners, with an "outing" threat and mutual violence, read as follows:

> *Your best friend, Jenny, has been in a two-year relationship with Kathy.*
>
> *One night Jenny asks you to come pick her up because she and Kathy had a fight. When you arrive Jenny is scraped up, has a black eye and a sprained shoulder. Kathy's cheek is slightly bruised. After you leave the couple's home, concerned, you ask what happened.*
>
> *Jenny tells you that she and her partner, Kathy, attended a party where they had too much to drink. Jenny explains that after the party Kathy accused her of flirting with another woman, yelled at her, calling her a liar and a cheater. In response to these accusations, Jenny says she slapped Kathy on the cheek. Jenny admits that Kathy was so angry that she lost control and shoved her down the stairs in their home, which resulted in the bruises and injured shoulder. Jenny says, afterward, Kathy was very sorry for pushing her, but told her that if she told anyone what happened, Kathy would tell everyone that Jenny was a lesbian.*
>
> *You have noticed unexplained bruising on Jenny's forearms in the past. Jenny says that sometimes she is afraid of Kathy, but she says it's her own fault for provoking Kathy.*

When generating variations on the vignette, it is important to make sure that the story is coherent and believable in all its variations. The domestic violence stories had to be vivid and plausible across all gender situations and not appear outlandish with the outing and mutual violence variations. It is important to keep in mind that, for some respondent, even the possibility of having a best friend who is gay or

lesbian required an imagination leap; such situations are hardly avoidable, and these limitations must be kept in mind when analyzing and interpreting the findings.

Note that our vignettes did not present respondent with first-person experiences of victimization, but rather with an event that occurred to a "best friend". The sensitivity involved in victimization reports required a bit of distancing within the vignette, so as to decrease reluctance. Similar techniques are important when respondents are asked about ethical behavior (conflict, lying, deceit) in organizational or professional contexts (Jehn and Jonsen 2010).

Crafting the vignettes for the friendship study required similar attention to detail. Since the main variable was the degree of friendship between the parties to the dispute, ideally we would have liked to compare disputes with friends to disputes with strangers. However, stories involving transactions with friends seldom seemed plausible, everyday life occurrences when involving strangers. We therefore had to compromise for other relational levels, and ended up comparing good friends to co-workers and acquaintances. In addition, we wanted gender to be immaterial to decisions about dispute resolution, and were concerned about the fact that some respondents might find it difficult to imagine good friends of the opposite sex. We therefore chose gender-ambiguous names for the protagonists in our scenario. The five vignettes read as follows:

You loan Terry, a good friend/co-worker/acquaintance, $10,000, and have not been paid back on the due date. Terry loaned you this amount several years ago and you paid back in full.

One evening over drinks you tell Jamie, a good friend/co-worker/acquaintance, about a business idea you had. After a few weeks, you find out that Jamie has stolen your business idea, run with it, and hasn't shared any profits with you.

You give Chris, a good friend/co-worker/acquaintance, a very good deal on renting your apartment or house while you are away for a few months. You later find out that Chris sublet the place to somebody else at market rate, substantially above what Chris paid you, and therefore, made a profit. You and Chris never discussed the possibility of subletting.

You loan Ash, a good friend/co-worker/acquaintance, your car while you are away for a few months. When you return, you find out that Ash has accumulated $500 worth of parking tickets and didn't tell you about it.

You and Joey, a good friend/co-worker/acquaintance, have been alternating purchasing season tickets for one another for a number of years. The tickets cost several thousand dollars. This year, it is Joey's turn to purchase the tickets, but Joey doesn't want to buy them with you anymore. You feel it would be more fair to end the arrangement next year, when you would be "even" with each other.

Several things should be pointed out. First, all five situations represent realistic examples provided by the dispute and litigation literature. Second, we chose gender-neutral names for the protagonists, in the hopes of eliminating gender bias in dispute resolution and creating more flexibility for the respondents in imagining the relationship with the other party. Third, we took care to avoid any terminology evoking law or litigation, so as not to prime respondents toward legal perceptions or recourse options. And fourth, the less sensitive nature of this study did not require recurring to third-person techniques.

Finally, when constructing vignettes, it is important to keep in mind our own possible biases. Even the most careful vignette may reflect the researchers' personal and cultural beliefs and values (Rosenthal 1970). While administering the survey on the web removes the problematic situational effects, the story should be one with which a broad array of respondents, from diverse demographic and cultural backgrounds, can engage. This is one of many good reasons why vignette surveys should ideally be tested in a pilot study on a smaller, yet diverse, sample.

26.6 Step IV: Designing, Programming, and Administering the Survey

The success of a factorial survey depends on its programming and flow as much as on the content of the vignettes themselves. Even in the late 1970s, when surveys were administered by hand, survey design and structure was an important

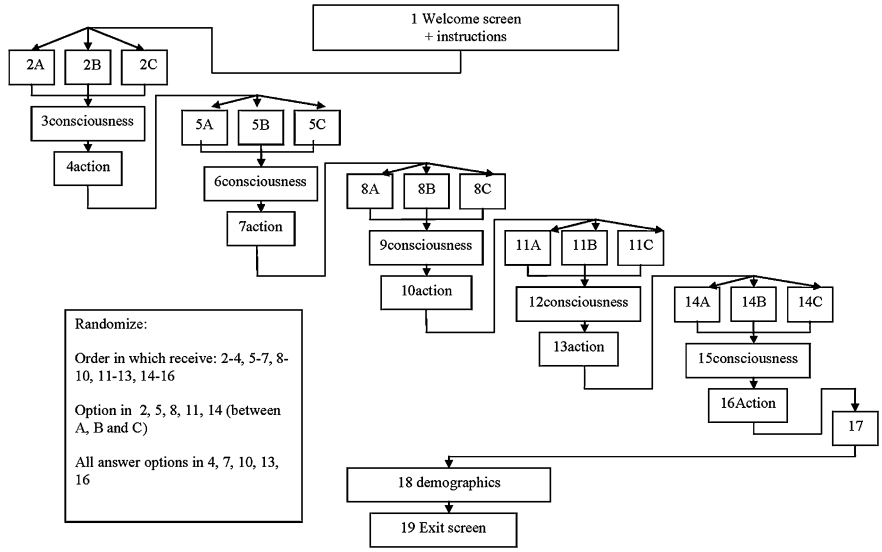


Fig. 26.2 Programming algorithm (adopted from Aviram and Leib 2010)

consideration (Sonquist and Dunkelberg 1977). There is a variety of software tools available for structuring surveys; some are available for free on the web, some come with software packages at research institutions, and some can be purchased, as a standalone product or as part of a programming service agreement. The choice of software platform depends, of course, on survey budget, but there are other considerations involving the software's capability to create proper randomization and its user friendliness.

One important factor to keep in mind is randomization and order bias (Schuman and Presser 1981). While the impact of preceding questions on later questions is always an issue in a survey, it is a source of particular concern in factorial surveys, because the thematic closeness in questions, all of which relate to the same vignette, may bias respondents to craft their answers so that they conform to each other. The implications of question-order effects are that many components of a given survey should be randomized: The order in which respondents receive vignettes, if the study uses several vignettes, some of the questions within each vignette, and the answers in multiple-choice vignettes.

Several studies encourage writers to use only one vignette per respondent and control factors within it (Rossi and Nock 1982; Shooter and Galloway 2010; Elis and Simpson 1995). This approach is methodologically "cleaner" than presenting each respondent with a sequence of vignettes: First, the possible impact of previous vignettes on subsequent ones is eliminated, and second, if data analysis requires comparing the vignettes to each other, all differences save for the designed variations are controlled for. However, sometimes budgetary considerations constrain the sample size and require asking respondents questions about several vignettes; after all, any additional variable in a vignette requires multiplying the number of respondents. This course of action, while more cost-effective, is riskier in the sense that the mere repetition of the scenario-answering experience may tire or bias respondents. It also makes the survey longer, a problematic consideration with regard to response rates (Reece and Siegal 1986). The gravest risk, however, is that later, in the analysis phase, the researcher will want to compare between different vignettes, answered by the same respondent. This creates serious sample distribution distortions, which are discussed further in the friendship survey. In our friendship study, for budgetary reasons, we compromised and presented respondents with all five of the stories (in randomized order, of course). Figure 26.2 presents the programming algorithm for the survey:

Since vignette studies often ask people about their opinions on sensitive, personal matters, it is important to create an environment that grants the same amount of legitimacy to each option. Experiments on survey wording have found that offering descriptions of two antagonized perspectives yields less bias than offering a one-sided statement and asking respondents to agree or disagree with it (Schuman and Presser 1981, p. 86). Also, merely asking respondents to "agree" or "disagree" with a given position may generate an agreement bias (Schuman and Presser 1981, p. 290). These are of particular importance in factorial studies, which are by nature used to gauge values and reactions and should be administered in an environment that fosters the individual's engagement with his or her personal values rather than conformism to survey expectations.

26.7 Step V: Respondent Recruitment and Sampling

Sample representation is of utmost concern to anyone conducting a survey. This is, of course, true for vignette surveys as well. While the survey is designed to gauge the significance of variables inserted into the vignette, demographic variables will invariably be included in any logit, probit, or OLS model, and it is therefore advantageous to obtain a sample as diverse and inclusive as possible. This is less of a challenge when conducting a study of decision making within a given professional group (Stanton and Rogelberg 2001, for examples see Hughes and Huby 2002; Taylor 2006), but requires more attention when addressed at the general population, measuring public opinions and expectations (Thurman 1989; Denk et al. 1997; Leighton 2010). As is the case in all web surveys, respondents are generally more educated and affluent than the general population (Stanton and Rogelberg 2001).

For those opting out of the less costly and problematic option of surveying students (Dalziel 1996; Peterson 2001), there are two main options. The first is working with a respondent recruitment firm. The costs involved with a professional service are considerable; however, this option may prove cost-effective if the service also provides survey software and programming. Professional recruitment is usually relatively swift, yielding a complete database within a few days.

It is important to keep in mind that professionally recruited subjects are usually modestly remunerated for their services on the panel. Payment in itself is not necessarily a problem, if it is commensurate with the value of respondents' time and effort (Reece and Siegal 1986), but several scholars raise the concern that "professional respondents" may substantially differ from one-shot respondents, in terms of survey fatigue and in other important ways (Baldinger 2001; Stanton and Rogelberg 2001). These concerns are mitigated by most reputable professional online survey panel companies through tight screening processes and rigorous control over survey distribution (Baldinger 2001).

The less expensive option is self-recruitment, which entails providing the survey link openly on the web and advertising online. While cheaper, this is more time-consuming and requires careful attention to survey variables when planning survey deployment. In the domestic violence survey, sexual orientation would have been an important variable in determining respondents' reactions to unconventional gender settings. We therefore oversampled GLBT respondents by posting links to our survey on GLBT Google and Craig's List groups. To avoid priming, we avoided addressing domestic violence groups, or indicating the topic of the survey in our advertisements.

One potential concern with self-recruitment in factorial studies stems from unexpected online reactions to specific vignettes. Any respondent, confronted with only one version of the vignette and unaware of the tested variables, may republish the link in other forums, compromising the diversity of the sample. In our domestic violence survey, a blogger who randomly happened to receive our female-toward-male violence variation concluded that the survey was "about male victims" and posted the link on a

fathers' rights forum. We had to intercede, exchange emails with the blogger, and politely ask that the link be removed. Such interventions are not always possible, and should be kept in mind when making decisions about respondent recruitment.

A final word pertains to sampling technique. For a couple of decades, the common wisdom in the field was that the sample for a factorial survey should be randomly drawn, in which case "each randomly drawn vignette sample is a reduced representative sample of the whole vignette sample from which it originates" (Dülmer 2007, p. 384). However, in a recent study, Dülmer experimentally shows that, under some circumstances, quota sampling is a superior method. While some of the challenges of random sampling are relaxed with wide web distribution, quota sampling can help in situations in which some variable combinations are impossible (such as, in our domestic violence study, the "outing threat" variable, which does not work with the opposite-sex scenarios). Choosing a quota sample might be a viable option in some circumstances, and is easily negotiable with most professional survey panel companies.

26.8 Step V: Statistical Analysis

Rather than providing a comprehensive guide to statistical analysis of surveys, this subchapter highlights the issue of hierarchical models and nesting, which is an important issue to keep in mind in vignette settings in particular. Most vignette survey studies will yield a database that includes, as independent variables, not only the variables in the vignette, but also a series of demographic variables pertaining to the respondents. As the analysis progresses, one is likely to run a number of tests on the data. A simple analysis assumes that respondents are interchangeable, and includes only vignette characteristics in the regression. A more complex analysis assumes and includes interactions between vignette and respondent, and typically the best regression model will be more inclusive.

A particular type of problem arises when, due to budgetary constraints, each respondent was presented with more than one vignette. This is not an issue if regression models are run for each type of vignette separately. However, sometimes researchers will want to include the type of vignette as an additional independent variable. In the friendship study, the plan was to include an analysis that compared the five scenarios—the loan, the business idea theft, the rental, the parking tickets, and the season tickets—to each other, and include this comparison in our regressions. Beyond the obvious design problem (vignettes differ from each other in a variety of ways that were not controlled for in the original design), there is also a sample distribution problem. Changing the unit of analysis from respondent to scenario does not simply increase the sample size from 1,000 to 5,000 observations; it creates a sample with 1,000 sets of five observations each. This creates a twofold problem related to the resulting change in sample distribution: First, the vignettes will have correlated error terms due to idiosyncrasies of the respondent, and second, the vignettes are hierarchically nested within respondents (Hox et al. 1991, p. 497; Raffery 2001).

The best way to confront the sample distribution problem is to take it into account in the research design and present all respondent with only one vignette, creating variations within the vignette. However, if this is not possible, there are several statistical solutions, and some or all of them should be pursued. First, it is highly recommended to run a Kolmogorov–Smirnov test on the original and modified samples, to check whether the databases' distributions differ significantly. Second, the modified database should be cluster-analyzed, thus controlling for intrarespondent effects. And third, hierarchical or multilevel analysis is recommended, including hierarchical models. Hox et al. (1991) have made some suggestions for such models. At the time of this chapter's writing, for Stata users, the GLLAMM patch addresses this exact problem (Rabe-Hesketh and Skrondal 2004).

26.9 Step VI: Interpreting the Results: Generalizations and Limitations

Even the most carefully crafted study becomes weaker with careless inference and over claiming. As with any research project, design and recruitment choices made due to budgetary constraints need to be explained and justified, so that the audience is aware of the limitations.

When framing data discussion and results, it is important to remember the limitations of vignette surveys. A hypothetical scenario, realistic and mundane as it may be, is not the real situation (Abbott and Sapsford 1993). It is risky, therefore, to leap from respondents' answers to the survey to strong predictive language about what respondents might do in real life. There are, for example, reasons to assume that people may report less temptation to cheat on their taxes (Thurman 1989) or engage in deceitful corporate behavior (Elis and Simpson 1995; Jehn and Jonsen 2010). In the domestic violence context, we assumed a bias on behalf of police reporting, and in the friendship context, we assumed a litigiousness bias. After all, in a hypothetical setting, people are not directly confronted with the practical, financial, and emotional costs of pursuing legal action.

That said, it is perfectly acceptable to explain, both in the methodological chapter and in the discussion, the steps taken to improve the experiment's external validity. The more the vignette simulates real-life situations, the easier it is to justify extrapolation and prediction. Moreover, even if a vignette survey does not offer a crystal ball of prediction, it does offer keener insights into people's consciousness and value systems than the traditional opinion survey. Ultimately, readers' confidence in inferences may vary with their enthusiasm about experimental methods and survey research; but even readers who are skeptical about the merits of generalizing from hypotheticals may learn something new about how people perceive and approach potential dilemmas.

Factorial surveys are wonderful research tools that combine an experimental design with the power (and reduced costs) of a survey. When meticulously planned and executed, and carefully interpreted, they can teach us much about opinions, values, and perceived courses of action.

References

Abbott, P., & Sapsford, R. (1993). Studying policy and practice: The use of vignettes. *Nurse Researcher, 1*, 81–91.

Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly, 42*, 93.

Aviram, H. & Leib, E. J. (2010). *Are friendship and the law separate spheres? A vignette survey study*. Paper presented at the 5th Conference on Empirical Legal Studies, New Haven.

Aviram, H., & Persinger, A. M. (2012). Domestic violence in unconventional gender settings: A vignette survey study. *Hastings Women's Law Journal, 29*(2), 152 (forthcoming).

Baldinger, A. L. (2001). *Online research and data quality: Stepping toward truth. Council of American Survey Research Organizations Journal, 2001*, 93–96.

Dalziel, J. R. (1996). Students as research subjects: Ethical and educational issues. *Australian Psychologist, 31*, 119–123.

Denk, C. E., Benson, J. M., Fletcher, J. C., & Reigel, T. M. (1997). How do Americans want to die? A factorial vignette survey of public attitudes about end-of-life medical decision-making. *Social Science Research, 26*(1), 95–121.

Dülmer, H. (2007). Random or quota design? Experimental plans in factorial surveys. *Sociological Methods Research, 35*(3), 382–409.

Elis, L. A., & Simpson, S. S. (1995). Informal sanction threats and corporate crime: Additive versus multiplicative models. *Journal of Research in Crime and Delinquency, 32*(4), 399–424.

Hox, J. K., Kreft, I. G. G., & Hermkens, P. J. (1991). The analysis of factorial surveys. *Sociological Methods Research, 19*(4), 493–510.

Hughes, R., & Huby, M. (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing, 37*(4), 382–386.

Jehn, K. A., & Jonsen, K. (2010). A multimethod approach to the study of sensitive organizational issues. *Journal of Mixed Research Methods Research, 4*(4), 313–341.

Leighton, S. (2010). Using a vignette-based questionnaire to explore adolescents' understanding of mental health issues. *Clinical Child Psychology and Psychiatry, 15*(2), 231–250.

Ludwick, R., & Zeller, R. A. (2001). The factorial survey: An experimental method to replicate real world problems. *Nursing Research, 50*(2), 129–133.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *The Journal of Consumer Research, 28*(3), 450–546.

Rabe-Hesketh, S., & Skrondal, A. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton: CRC.

Raffery, A. A. (2001). Statistics in sociology, 1950–2000: A selective review. *Sociological Methodology, 31*, 1–45.

Reece, R. D., & Siegal, H. A. (1986). *Studying people: A primer in the ethics of social research*. Macon: Mercer University Press.

Rosenthal, R. (1970). The social psychology of the behavioral scientist: On self-fulfilling prophecies in behavioral research and everyday life. In E. R. Tufte (Ed.), *The quantitative analysis of social problems*. Reading: Addison-Wesley.

Rossi, P. H., & Nock, S. L. (Eds.). (1982). *Measuring social judgment: The factorial survey approach*. Beverly Hills: Sage.

Salant, P., & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley.

Sapsford, R. (1999). *Survey Research*. London: Sage.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude suveys*. New York: Academic Press.

Shooter, W., & Galloway, S. (2010). The use of factorial surveys in leisure research. *Journal of Leisure Research, 42*(4), 641–652.

Sonquist, J. A., & Dunkelberg, W. C. (1977). *Survey and opinion research: procedures for processing and analaysis*. Englewood Clifs: Prentice Hall.

Stanton, J. M., & Rogelberg, S. G. (2001). Using Internet/Intranet web pages to collect organizational research data. *Organizational Research Methods, 4*(3), 200–217.

Taylor, B. J. (2006). Factorial surveys: Using vignettes to study professional judgement. *British Journal of Social Work, 36*(7), 1187–1207.

Thurman, Q. C. (1989). General prevention of tax evasion: A factorial survey approach. *Journal of Quantitative Criminology, 5*(2), 127–146.

Part VII

Special Issues in Survey Methodology

Comparability of Survey Measurements 27

Daniel L. Oberski

Comparative surveys nowadays provide a wealth of survey data on a diverse range of topics covering most countries in the world. The online companion[1] to the ''SAGE handbook of public opinion research,'' for example (Donsbach and Traugott 2008), lists some 65 cross-national comparative social surveys that have been conducted around the world since 1948. Besides these general social surveys, many surveys on specific topics such as education, old age and retirement, health, working conditions, and literacy, to name just a few, are carried out continually.

Surveys may be conducted for different purposes. One purpose is estimation of population means, totals, and marginal distributions; another is the estimation of relationships between variables. Van de Vijver and Leung (1997) called studies with these goals respectively "level" and "structure" oriented. A comparative survey will then have as its goal to *compare* such level and/or structure parameters.

However, it is well known that even estimates from surveys carried out with the utmost care and attention to quality will contain some amount of survey errors (Groves 2004). A simple division can be drawn between *errors due to the selection of sample units*, and *errors due to the measurement instrument*. These error sources may have an effect on the estimates in the form of both bias and variance.

In comparative surveys, then, the estimates to be compared may each be influenced by survey errors, leading to the possibility that the comparison to be made is invalidated. That is, the estimates from different surveys might not be *comparable*.

The problem of comparability does not only occur when comparing the results of different surveys, but will also apply to the comparison of subpopulations in the same survey.

This chapter discusses the problem of comparability from the point of view of total survey error. The problem is illustrated with examples, and comparability issues related to errors due to both the selection of sample units and the measurement instrument are discussed. The discussion is, of necessity, limited and brief. For further details the reader is referred to the literature at the end of the chapter.

Figure 27.1 shows a typical application of comparative surveys: groups, in this case European countries, are compared on their means. Figure 27.1 compares the percentage of adults in each country that indicate participating in some type of educational or training activity, as estimated by Eurostat's 2007 Adult Education Survey.

D. L. Oberski (✉)
Joint Program in Survey Methodology,
University of Maryland, College Park, MD, USA
e-mail: daniel.oberski@gmail.com

[1] http://www.gesis.org/en/services/data/portals-links/comparative-survey-projects/.

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_27, © Springer Science+Business Media New York 2012

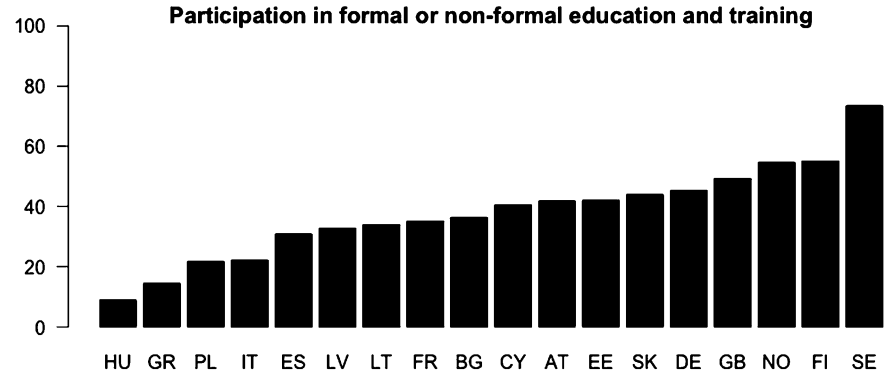


Fig. 27.1 Comparative statistics by country on life-long learning in Europe. Shown are percentages of adult population participating in education, from Eurostat's 2007 Adult Education Survey. *Source* Boateng (2009, Table 1)

Boateng (2009:2) presented these data, remarking that "Total rates of participation vary between countries and the data shows the Nordic countries and the UK/GB (Great Britain) having high rates of participation. Low rates of participation are found in Hungary, Greece, Poland, and Italy." (p. 3). This certainly appears to be the case when looking at Fig. 27.1. However, how might the estimated means shown there have been different if nonresponse were much higher in Greece and Hungary than it was in Finland and Sweden? Would the comparison still yield the same conclusion? Suppose nonresponse in Greece were predominantly due to noncontacts—for example, because those in training programs are more difficult to contact—while nonresponse in Finland is more related to refusals of the lower educated?

It is clear, then, that although the differences shown in Fig. 27.1 are statistically significant and appear to confirm a priori expectations one might have on differences between Nordic and other countries, there are possibly other *nonsubstantive* explanations for the findings shown. Nonresponse is only one of those factors that form possible alternative explanations for differences: among others are differences in frame errors, translation errors, differences in understanding of the concept of "job training" across countries, differences in interviewer training, differences in the printed answer scales, and so on. In effect, *any difference in systematic survey error across countries potentially threatens the comparison*.

Whether these issues truly affect the comparison is not known, and it is by no means the intention here to suggest the Adult Education Survey provides incomparable numbers. The problem of comparability will arise on any occasion where groups are being compared on a survey statistic.

Besides means, measures of relationship between variables are also of central interest. Boateng (2009:3), for example, mentions "varying gender differences [in educational participation]" across countries. Public health researchers have studied social class differences in self-rated health. Differences across countries in these health inequalities were studied by von dem Knesebeck et al. (2006).

Figure 27.2 shows country rankings on these health inequalities over educational levels. To estimate inequality and simplify comparison, the slope of a linear regression of self-rated health on education levels in each country is shown,[2] based on data from the European Social Survey round 4 (see Stoop and Harrison, Chap. 15 for more detail on the ESS). The size of the coefficients is shown as filled circles with two standard error intervals. Lower (more negative) values indicate larger inequalities in health across levels of education. It can be seen that Germany is by these estimates the most equitable, while Greece has the most inequality. Policy

[2] For a complete description of the study design and the original questionnaires, please see http://ess.nsd.uib.no/.

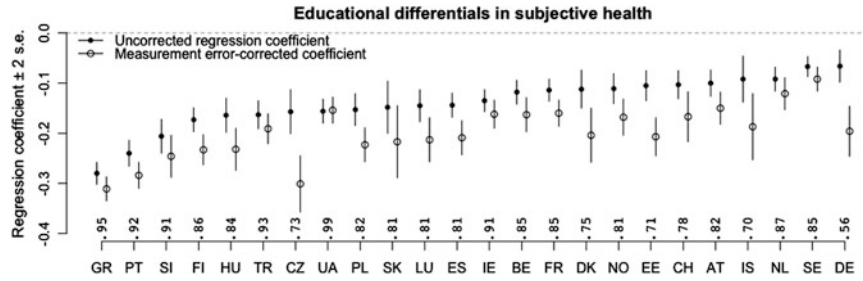


Fig. 27.2 The estimated difference in self-rated health per education level by different countries (*closed circles*). Shown also is the same difference corrected for measurement error in the education variable (*open circles*). The estimated reliabilities used to correct the regression coefficients in each country are printed at the *bottom*

makers might be tempted to emulate Germany's health policies as a shining example of health equality in Europe.

Similarly to the comparative analysis of the proportion of life-long-learners discussed above, one may wonder whether these differences are due to a substantive reason such as countries' health policies, or whether there might be some other possible explanation for differences. Again, differential nonresponse might play a role. However, in the case of relationship parameters that would imply a high-order interaction among country, nonresponse bias, and the two variables of interest (Groves and Couper 1998). This explanation is still possible, then, but less plausible than it was for the comparison of means.

Measurement error, on the other hand, is well-known to have a direct and strong effect on simple regression coefficients (e.g. Fuller 1987). These are attenuated downward by unreliability in the independent variable, in this case education. Presumably, if there are strong differences between countries in the reliability of education level, the rank ordering of the countries may change accordingly.

Level of education is an objective variable, but it is not free of measurement error. Ganzeboom and Schröder (2009) found reliabilities of level of education between 0.7 and 0.9 (p. 9). The bottom of Fig. 27.2 displays the estimated reliability of the level of education variable, showing that reliability varies considerably across countries (see also Oberski et al. 2008). For example, the reliability is estimated as close to 1 in the Ukraine but is very low in Germany (only 0.56).

Unreliability can have a strong biasing effect on the regression coefficient. Having obtained these reliability estimates, one can then correct the regression coefficients used to rank the countries for measurement error, obtaining corrected coefficients. These are shown in Fig. 27.2 as open circles, again with two standard error intervals. Particularly, the rank of countries where educational level is estimated with lower reliability is affected. Germany, for instance, moves from first place to the middle, while Greece is relieved from its role as the most unequal country in Fig. 27.2 by the Czech Republic. Correction for measurement error dramatically affects the conclusions drawn from the comparison of these relationships.

The above examples demonstrate that comparability of survey measurements is an issue that must be carefully considered before substantive conclusions may be drawn from comparative surveys. The problems are not limited to the comparison of means, but extend to the comparison of relationships.

Nor is the issue of comparability limited to "subjective" variables; comparisons across groups of an objective variable such as whether a person is in a training program may also be threatened by issues of comparability.

Another example that clarifies this is the comparability across time of the US Bureau of Labor Statistics' Current Population Survey, or the National Election Study which have

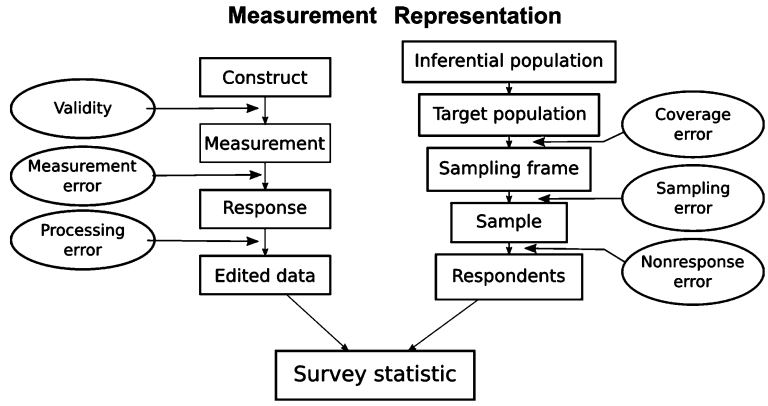


Fig. 27.3 Total survey error framework, adapted from Groves (2006)

switched from face-to-face to telephone to mixed mode data collection including web surveys over the years (Cohany et al. 1994; Malhotra and Krosnick 2007). Survey mode has been shown to affect estimates and therefore it is a question to what extent the time series can be compared. This also demonstrates that the issue of comparability is not limited to cross-country comparisons, but may also extend to comparison over time, different social groups, and so forth.

Two issues are of prime concern. First, what are possible sources of incomparability? This is discussed in the following section in the framework of Total Survey Error (TSE) (Groves 2004). Second, in what way should survey methodologists and users consider comparability? This question is addressed in the remainder of the chapter.

27.1 Incomparability as Differential Total Survey Error

Incomparability of statistics across groups arises from systematic error or "bias." A framework for describing such errors in surveys is that of TSE (Groves 2006), shown in Fig. 27.3. The figure shows the survey process and sources of errors that may arise during this process. The errors may be random or systematic. Each of these errors is described in detail elsewhere in the present volume.

Particular to comparative surveys is the existence of at least one other survey statistic, with which the statistic shown in Fig. 27.3 is to be compared. This statistic is subject to different survey errors arising from the sources shown there. If the overall statistics are differently affected, this will cause "bias" in the comparison. That is, besides the substantive differences of interest, the comparison will also be partly affected by differences in systematic survey errors.

At this high level of abstraction, the problem is simple: differences in systematic TSE. And so is the solution: systematic TSE must either be reduced to zero or kept equal or close to equal across groups. This observation is the basis for the field of comparability of survey statistics.

More specifically, it can be seen that Fig. 27.3 has a "representation" and a "measurement" side. These correspond respectively to the goals of obtaining a sample statistic that is close to the population value, and measuring a quantity that corresponds to that intended by the survey researcher.

Systematic representation errors involve undercoverage, overcoverage, and nonresponse bias. These errors are so mathematically similar to each other that they will be discussed jointly in the following section. Systematic errors in measurement arise from invalidity, processing, and measurement error. Measurement error is generally understood to include response error, interviewer-induced response effects, social desirability, method effect, response styles, and random error (unreliability).

It should be noted that errors that are, in TSE terms, random with respect to one statistic, usually the mean or total, are systematic with

respect to another. For example, purely random measurement error will not affect means, but it will bias correlations and regression coefficients. Conversely, nonresponse bias may, in a particular instance, affect means but not correlations. Of importance for comparability are only those errors that may cause systematic deviations. From this point of view, sampling error, for example, is not relevant, because it is standard practice to account for sampling error in any statistical analysis.

Systematic errors due to survey mode have received quite a bit attention in the literature (Dillman et al. 2008). In the TSE framework adopted here, a survey mode is a combination of different aspects such as wording, design, interviewers, coverage, nonresponse propensities, and so on. Differences in mode are likely to increase potential incomparability since many sources of error are changed at the same time. However, we do not consider it separately from its constituent errors here.

In the field of cross-cultural psychology, a different classification of biases was given by Van de Vijver and Leung (1997). These authors distinguished between "construct," "item," and "method" biases. In the TSE framework, construct bias corresponds to differences in invalidity (top left of Fig. 27.3), and item bias corresponds to differences in measurement error. "Method" bias, in their formulation, is essentially the rest of Fig. 27.3.

27.2 Comparability

Making comparisons possible starts by "meeting the coming disease" of bias and preventing it. This involves such actions as putting in place adequate translation and adaptation procedures for cross-cultural studies (Harkness 2003) ensuring uniformity of fieldwork arrangements (Jowell 2007), and defining concepts that will be similarly understood in different groups (Hui and Triandis 1985). Such actions are vital to the goal of bringing the ideal of comparability closer. However, they do not guarantee that this ideal is attained. Just as even the best survey has errors, even the most carefully designed, implemented, and monitored comparative survey will have some degree of bias in its comparisons; there will always be some possibility of incomparability.

There are four basic possible approaches to this problem.

First, the researcher may simply ignore the problem. This amounts to an implicit or explicit assumption that any bias in the comparison will be small. That is, that any biases will not be so large as to "explain away" the conclusions drawn. Such an assumption can be warranted if prior research suggests it. In other cases it will amount to a hope.

In psychometrics, comparability is formulated in terms of "item bias": the difference, *for equal values of the true score*, in expected value of the variable given the comparison group (Mellenbergh 1989). This idea is the basis for the rich literature on differential item functioning or "DIF" (Holland and Wainer 1993).

The second possibility is then that, by some means, the hypothesis is tested that some form of bias will occur. Item bias is usually associated with the comparison of parameters of item response theory (IRT) models over groups, to which we will return later in the chapter. This includes the approach taken in the invariance testing literature (Meredith 1993; Steenkamp and Baumgartner 1998; Millsap et al. 2007) discussed later in this chapter. If the test indicates that the measures are not comparable across group, the researcher will then avoid the comparison in question altogether. This approach corresponds to what will here be termed a "strong" interpretation of comparability: that any possibility of bias (in the population) is sufficient to invalidate comparisons.

Another interpretation of the idea of comparability is that a comparison is warranted if it is very unlikely that the bias due to differential errors will be so large that it can change conclusions. It might be called the "weak" interpretation of comparability. Weak comparability is implicitly assumed if the researcher ignores the issue of comparability. However, given certain research designs, it may actually be

investigated. This, then, is the third possible approach to the problem of incomparability: to perform analyses that make plausible the assumption that any biases will not be so large as to "explain away" the conclusions drawn. Parts of the rest of this chapter will give particular attention to this approach.

A fourth and final approach is to attempt to directly estimate the survey errors, so they may be corrected for. Any differences across groups can then no longer threaten the conclusions drawn from comparisons. This approach was taken in the example of the educational differential in subjective health above: since it was known that differential measurement error in the education variable over countries could invalidate comparisons, each regression coefficient was corrected for the estimated reliability before the final ranking was made. The correction approach based on models that estimate item bias is known as "equating" in the psychometrics literature (Holland 1982).

In psychophysics, a notion similar to "item bias" is termed "differences in response function" and has been the subject of extensive research in that field since its inception (Stevens 1975). Psychophysics is concerned with the relationship between a stimulus and a person's perception of that stimulus. Classic examples are hues of light, highness of musical notes, and creaminess of butter. Many psychophysical experiments used response scales with arbitrary reference points, such as squeezing a ball, producing a louder or softer sound, or freely giving a number (with no restrictions) to allow the subject to indicate their perception of the stimulus.

The interpersonal differences in scale precluded direct comparison between people of the obtained scores, since, for instance, older people will squeeze less, and whether one decides that 100 or 1,000 means "very creamy" is entirely up to the respondent. Classical psychophysics therefore developed the notion of the response function, and methods to estimate it for individual respondents so as to correct for incomparability (Stevens 1975). The method was applied to survey research by Saris (1988), and also forms the basis for the "anchoring vignettes" approach developed by King et al. (2004).

The last three approaches—strong comparability, weak comparability, and correction of statistics—all require some form of modeling. In what follows different models that can be used for these purposes will be discussed and demonstrated. The first concern, however, should always be to reduce the possibility of incomparability as much as possible.

27.3 Prevention of Incomparability

When feasible, incomparability should be prevented as much as possible by reducing TSE, and by making errors likely to be similar across groups. Sources of survey error and suggested methods to reduce them are discussed elsewhere in this volume. Here, a short description of two methods useful for developing comparable questions are discussed: translation and quality control by coding of question characteristics. Another method is cognitive interviewing (Willis 2005).

27.3.1 Translation or Adaptation

One particular source of incomparability specific to comparative studies is that of translation or adaptation. Translation or adaptation usually refers to questions being asked in different languages, as is the case in cross-national research. But it may also refer to the comparison of other groups: questions designed for adults are not usually directly applicable to children, for instance.

Harkness (2003) gave an overview of different translation procedures intended to yield comparable questions, describing particularly the Translation, Review, Adjucation, Pretesting, and Documentation (TRAPD) procedure developed for large cross-national surveys such as SHARE and ESS. For more information we refer to Harkness et al. (2010), where many other issues relating to cross-national comparative surveys are also discussed.

The difficulties inherent in translation or adaptation procedures can be demonstrated with the following item from the ESS, cited by Zucha in Hoffmeyer-Zlotnik and Harkness (2005):

> Please tell me how important each of these things should be in deciding whether someone born, brought up, and living outside [country] should be able to come and live here. Firstly, how important should it be for them to… be wealthy?

The stimulus "be wealthy" was translated in Italian as "…avere una buona salute," which means "…be healthy." Clearly this is simply a mistake in the process; it does go to show, however, that even with extremely careful procedures in place, mistakes are still made.

In French and Spanish "be wealthy" was translated as "être riche"/"ser rico", which means "be *rich*." These languages do not have a separate word for "wealthy," but only the word "rich," which has a related but different meaning. The word "wealthy" cannot be translated into these languages as closely as "rich" could be.

This example demonstrates that translation is not just an issue of finding the corresponding word in the other language. It entails translating the *concept intended* as closely as possible (Harkness 2003). How close the translation is able to get to the intended meaning; in turn, depends on the question itself. In the example, if the original question writers had written "rich" instead of "wealthy." there would have been no conceptual difference for these languages.[3]

Another example of this phenomenon is the word "fair." used in social trust scales: "people try to be fair." There appears to be no very close translation of this concept in many European languages. As solutions, "honest" (Dutch, Spanish, Swedish) and "behaving correctly" (French) are found in the ESS. Both come close to the intended concept but are not as close as they would have been if the original questions had used the word "honest."

Adaptation issues also arise when the mode of survey administration is different for the groups to be compared. For example, questions with many categories in self-administered modes when asked verbally are often broken into multiple steps. This may have an effect on the responses. There is no simple solution to this problem, and researchers may have to weigh the possibility of incomparability against the information loss incurred if the number of categories were reduced.

27.3.2 Question Comparability Quality Control Using a Question Coding System

An additional check on question comparability may be provided by comparing different characteristics of the questions in the original and adapted versions. One such coding system (SQP) will be discussed here. Another coding system, QAS, was developed by Lessler and Forsyth (1996).

Zavala (2011) gives the results of coding question characteristics with the SQP coding system (Oberski et al. 2011; see also Saris and Gallhofer 2007) and comparing the results across translations in the European Social Survey round 5.

Table 27.1 shows the differences found over three countries for two items. The countries and items have been relabeled, since this table is given merely as a demonstration of the type of problem that can be detected with this method.

Table 27.1 shows that problems that can be detected by this procedure are not semantic in nature; that type of equivalence must be established by another method, such as the translation procedure or cognitive interviewing. Rather, the differences are on aspects of the question design that are known from the survey methodology literature to affect question quality and response patterns (Alwin 2007; Saris and Gallhofer 2007). Common are differences in the visual layout of the response scales, and the omission of interviewer or respondent instructions. Since the

[3] There might of course still be difficulty in translating into other languages. Source questions formulated in the English language, which is often claimed to have more words than any other natural language, would appear to be particularly prone to this type of problem.

Table 27.1 Unwarranted differences in two survey question's design characteristics across three countries in the ESS

	Item 1	Item 2
Country B	• Missing scale definition	• Missing question introduction
	• Missing respondent instructions "please use this card"	• Missing respondent instructions "please use this card"
	• Showcards have different layout	• Showcards have different layout
Country C	• Missing respondent instructions "please use this card"	• Missing respondent instructions "please use this card"
	• Series of separate questions changed to battery of questions	• Series of separate questions changed to battery of questions
	• Showcards had numbers before the categories, while source had no numbers	• Showcards had numbers before the categories, while source had no numbers
Country D	• Missing question introduction	• Showcards had numbers before the categories, while source had no numbers
	• Showcards had numbers before the categories, while source had no numbers	• Showcards had boxes around the categories, while source had no boxes
	• Showcards had boxes around the categories, while source had no boxes	• Showcards have vertical instead of horizontal layout
	• Showcards have vertical instead of horizontal layout	

layout of response options affects the response (Christian and Dillman 2004) such differences form a threat to equivalence of the question and were corrected before the questionnaires were fielded.

Rigorous translation and pretest procedures such as the ones described here can help to reduce incomparability. However, even when such procedures have been followed it is still possible that the questions are not comparable. For this reason, models were developed that allow for the estimation and correction of the degree of incomparability across groups.

27.4 The Representation Side

The "representation" sources of TSE shown in Fig. 27.3 arise from the process that selects respondents into the sample. As will be shown below, if this selection process produces bias that differs over the groups being compared, a bias in the comparative statistic may occur. Since the literature on this topic focuses strongly on bias due to nonresponse, this section will follow that convention and discuss bias on the representation side in terms of nonresponse error. However, it may be kept in mind that the same arguments will apply to coverage error and other possibly differential selection processes such as self-selection into convenience samples, self-selected Internet panels, and so forth.

27.4.1 Incomparability due to Nonresponse

Nonresponse bias is well-known to affect survey statistics (Groves and Couper 1998) and, due to rising nonresponse rates, has become a prime concern among survey methodologists. In this section we develop the effect of nonresponse bias on the *comparison* of survey statistics. If this comparison is affected by nonresponse bias, comparability may be threatened.

Researchers may be interested in the comparison of means or totals, but also of odds ratios, correlations, regression coefficients, etc., across groups. Interest will typically focus on the difference between the groups, $T^{(1)} - T^{(2)}$ in some statistic. In each group, there may be nonresponse bias, so that the statistic observed by using only respondent data in each group, say

$T_R^{(g)}$, equals $T_R^{(g)} = T^{(g)} + b^{(g)}$, where $b^{(g)}$ is the nonresponse bias in group g.

Differential nonresponse may make the groups incomparable, in the sense that differences between the groups might be due to differences in nonresponse bias between the groups. This can readily be seen by expressing the differences between groups as:

$$T_R^{(1)} - T_R^{(2)} = T^{(1)} - T^{(2)} + \left(b^{(1)} - b^{(2)}\right) = (\text{true difference}) + (\text{difference in bias}).$$

It can be seen that the threat to comparability is the difference in nonresponse bias between the groups (Groves and Couper 1998, pp. 8–9). This difference in nonresponse bias, in turn, depends on the differences $T_R^{(g)} - T_N^{(g)}$ between respondents and nonrespondents with respect to the statistic, and the response rates in the groups:

$$b^{(1)} - b^{(2)} = (\bar{\rho}^{(1)} - 1)\left(T_N^{(1)} - T_R^{(1)}\right) + (\bar{\rho}^{(2)} - 1)\left(T_N^{(2)} - T_R^{(2)}\right),$$

where $\bar{\rho}^{(g)}$ is the response rate in group g. The bias in the difference will be zero if the response rates in both groups are 1. If the response rates are not 1 but equal, there may still be bias if the differences between respondents and nonrespondents are not equal across groups. Conversely, even if these differences between respondents and nonrespondents are equal, there may still be bias if the response rates are not equal. There will be zero bias only if the differences between respondents and nonrespondents are equal and the response rates are also equal. Strong bias can be obtained if the biases are opposite to each other and the response rates are substantial. It can also be shown that higher response rates, similar response rates, lower nonresponse biases, and similar nonresponse biases will tend to lower the relative bias of the difference.

The problem may be more pronounced when comparing means than when comparing correlations, regression coefficients, and other relationship parameters. Goudy (1976) and Voogt (2004) argued the necessary assumptions to be more plausible when comparing (logistic) regression coefficients than when comparing means. It is also a common assumption in much of experimental psychology that nonresponse bias will tend not affect estimates of relationships. Whether this is actually true or not remains mostly a matter of assumption (Groves and Peytcheva 2008, p. 182).

It is clear, then, that unless one is willing to make strong assumptions, nonresponse bias will generally cause groups to be incomparable in the strong sense. This will be especially true when comparing the results obtained from different surveys, such as is usually the case in cross-national research.

Perhaps, however, in a particular study, the biases due to nonresponse are not so strong so as to be able to "explain away" the found differences. One method of examining this possibility of weak comparability is to attempt to estimate nonresponse bias directly (Groves 2006; Stoop et al. 2010). Another is to attempt to work out the *maximum possible bias* in the difference. Recent work by Schouten, Cobben, and Bethlehem (2009) has provided results that allow for estimation of a maximum bias in the comparison of means. This maximum bias can then be used to see whether such a bias could in fact form an alternative explanation for the observed difference.

27.4.2 Maximal Absolute Bias and the R-Indicator

Survey response can be viewed as a stochastic process (Bethlehem 1988; Groves and Couper 1998:11-2) with every potential respondent having a probability of participation, denoted as ρ, sometimes called the response propensity score. The decision to participate or not is determined by a person's ρ, with 1 being a certain respondent and 0 a certain nonrespondent. The response rate is then just the average, denoted $\bar{\rho}$, of the response propensities.

Schouten et al. (2009) developed an indicator of the representativeness of a survey based on the response propensities. They termed this quantity the "R-indicator" The R-indicator, $R(\rho)$, has a rather simple definition: it is defined

as $R(\rho) = 1 - 2\sqrt{\text{var}(\rho)}$. The R-indicator will lie in the (0, 1) interval, 0 meaning maximum unrepresentativeness, and 1 meaning perfect representativeness. Note that the R-indicator does not depend on the statistic or the target variable of interest, but only on the variance of the propensity scores among sampled units.

If an estimate of the propensity score can be obtained from auxiliary data, the R-indicator can then be estimated simply via calculating the variance of the estimated propensity scores (Shlomo et al. 2008). Estimates of the R-indicator do assume, however, that response propensity is not related to the outcome statistic of interest after conditioning on the auxiliary variables. That is, that the nonrespondents are Missing at Random (MAR) given the auxiliary variables. The disadvantage of this approach is, therefore, that its success depends wholly on the selection of adequate auxiliary variables.

By intuition, it seems reasonable that nonresponse bias, and therefore bias in the difference between groups, would depend on the R-indicator. And, indeed, this can be shown to be the case. In a single survey, the *maximum absolute bias* in a mean will be bounded by $|b| \leq \frac{[1-R(\rho)]\sigma_y}{2\bar{\rho}}$, where σ_y is the standard deviation of the target variable. Thus, the more unrepresentative the survey is, and the lower the response rate, the stronger the bias can potentially be.

That maximum bias depends on the R-indicator and response rate is an incredibly useful result, since the maximum absolute bias gives an upper bound on the amount of damage nonresponse can do to the estimate. Given an estimate of the R-indicator, it can be investigated, for instance, whether an obtained mean or total still differs significantly from some predetermined value, even after taking the maximum bias into account. If it does, then nonresponse bias becomes less plausible as a possible alternative explanation for observed differences.

The concept of maximum bias is here straightforwardly extended to the comparison of groups. When a difference in means is of interest, the maximum bias of the difference is bounded by

$$|b^{(1)} - b^{(2)}| \leq \frac{1}{2}\left(\frac{\sigma_y^{(1)}[1-R(\rho^{(1)})]}{\bar{\rho}^{(1)}} + \frac{\sigma_y^{(2)}[1-R(\rho^{(2)})]}{\bar{\rho}^{(2)}}\right).$$

This worst case will occur only when the nonresponse biases in the two groups are maximal *and in opposite directions*. When the biases can be assumed to be in the same direction, the maximum bias of the difference will be bounded by $\max\left(\frac{\sigma_y^{(1)}[1-R(\rho^{(1)})]}{2\bar{\rho}^{(1)}}, \frac{\sigma_y^{(2)}[1-R(\rho^{(2)})]}{2\bar{\rho}^{(2)}}\right)$.

The above is merely a reformulation of the theoretical results given earlier. Its advantage, however, is that, under the assumption of MAR given the covariates, it does not depend on unknown quantities: it can be estimated in a given study. In a particular study where two groups are compared on their means, weak comparability can be made more plausible if good estimates of the R-indicators can be obtained. An illustration of this approach is given below.

Table 27.2 gives the sample sizes, response rates, and estimated R-indicators for the Belgian and Norwegian samples of the ESS, Round 3 (fielded in 2006). It can be seen that both countries had relatively high and similar values for both response rate and R-indicator. The values are neither perfect nor equal; however, as is to be expected in any applied study.

According to the strict definition of comparability, the Belgian and Norwegian samples are not comparable, because they do not have exactly the same response rate, R-indicator values, and (presumably) nonresponse bias. As shown by the equations above, this *might* lead to apparent differences that do not reflect population differences.

This shows that the strict definition of comparability is not very helpful in practical research, since some differences in response rates, and representativity will be rule rather than exception. The weak definition of comparability then compels us to ask whether such differences can plausibly explain the observed differences between the samples or not.

To demonstrate the application of this principle we compare Norway and Belgium on

Table 27.2 Sample sizes, response rates, and estimated representativity (R) indicators for the European Social Survey Round 3 in two countries

	Sample size	Response rate	R-indicator (estimate)
ESS 2006 (Belgium)	2,927	61.4%	0.807
ESS 2006 (Norway)	2,673	65.6%	0.762

Source Shlomo et al. (2008, p. 33)

average reported happiness, and the proportion of people who say they use the internet "every day." These comparisons are shown in Table 27.3, together with the maximal biases assuming bias is in the same or in opposite directions. The corresponding "worst-case scenario" estimates of differences between Belgians and Norwegians are displayed in the last four rows.

Table 27.3 shows that in the worst case scenario, the possibility cannot be ruled out that the differences in response rate and lack of representativity are the cause of the apparent differences between Belgium and Norway. This will only occur when the biases are maximal and in opposite directions; for example, when the imbalance is on gender *and* males are *more* likely to use the Internet in Norway, but *less* likely to do so in Belgium. Other scenarios where this might occur are of course also possible.

For the cross-country comparison of happiness, applying a correction for the maximum potential bias due to differential nonresponse drives home the point that incomparability due to nonresponse cannot be ruled out as an explanation for differences. Differential nonresponse even has the potential to make it look as though the average Norwegian is happier than the average Belgian, when in reality the opposite might be true.

Looking at daily Internet use, the differences between Belgium and Norway could be due to nonresponse if the biases are maximal and in opposite directions. However, if we can assume that they are in the same direction, then the maximal bias due to differential nonresponse is 0.09. In that case, even the worst-case bias adjustment will still yield the conclusion that Norwegians have a statistically significantly higher daily Internet use than Belgians. So a possible alternative explanation for these differences in terms of incomparability due to differential nonresponse bias is less plausible in comparison of Internet use.

The estimated R-indicator, when it can be applied, is a useful device for examining whether weak comparability with respect to nonresponse is plausible or not. It must be kept in mind that estimates of the R-indicator are only as good as the auxiliary variables that are used to estimate them. Furthermore, although some have contended that nonresponse is less likely to affect comparisons of relationships than it is to affect comparisons of means, work on comparing relationships across groups is, at the time of writing, still a field open for investigation.

In this section the theoretical effects of nonresponse on comparability were given. Through the R-indicator, it was possible to perform an analysis of weak comparability, under certain assumptions. The methods and results presented here were originally developed to deal with the issue of nonresponse, but can be equally well applied to assess the possible effects of over- or undercoverage or self-selection in nonrandom samples on comparability.

27.5 The Measurement Side

The previous section discussed the goal of obtaining a comparative sample statistic that is close to the "population" value. The second goal identified in Fig. 27.3 is that of measuring a quantity that corresponds to that intended by the survey researcher, the "measurement" side of total survey error.

Tourangeau et al. (2000) gave an account of the process of survey response, based on a body of experimental findings in cognitive psychology.

Table 27.3 Analyses of differences between Norway and Belgium in happiness and Internet use, and the same analyses correcting for maximum bias based on R-indicators and response rates in the two surveys

	"How happy are you?" (0–10)	Use internet "every day"
ESS 2006 (Belgium)	7.66 (1.6)	0.346 (0.48)
ESS 2006 (Norway)	7.93 (1.5)	0.493 (0.50)
Difference	−0.27* (0.06)	−0.147* (0.01)
Max. abs. bias of difference (opposite directions)	0.52	0.17
Max. bias-corrected estimate (opposite directions)	+0.25*	+0.03
Max. abs. bias of difference (same direction)	0.27	0.09
Max. bias-corrected estimate (same direction)	0.00	−0.06*

*p-value < 0.05. *Source* ESS data (2006). http://ess.nsd.uib.no/ess/round3/

In their process model, the respondent goes through the stages of comprehension, retrieval, judgment, and response. Response behavior is then tied to these different processes. For example, consider the question (p. 38):

> During the past 12 months, since January 1, 1987, how many times have you seen or talked to a doctor or assistant about your health? Do not count any time you might have seen a doctor while you were a patient in a hospital, but count all other times you actually saw or talked to a medical doctor of any kind.

The structure of this question is very complex, containing many subordinate clauses, and conditions. The complexity is in the concept being described, leading Tourangeau et al. (2000, p. 39) to remark that "processing this question into its underlying logical form is likely to impose quite a burden on working memory, one that may exceed the capacity of the respondent." They, therefore, argue that the step of comprehension may not correspond to that intended by the researcher.

Suppose that one wished to compare the number of doctor's visits made by people with high cognitive capacity to people with lower cognitive capacity—for example, intellectually disabled adolescents compared with their peers. Considering that the comprehension hurdle is likely to be cleared differently by these two groups, the response process is likely different as well. Thus, different answers to this question may be obtained for two people, one with a very low intelligence and the other with a high intelligence; *even if they visited the doctor the same number of times*. Such variation of responses for the same true score will be called "measurement error" here.

Differences in the response process across the groups to be compared may threaten comparability. Differences in comprehension may arise due to translation or cultural differences, a topic that has received some attention in the literature (Harkness et al. 2003). Differences in the retrieval of information may occur due to differences in the salience and frequency of an event asked about (Tourangeau et al. 2000, Chaps. 3–4), and in opinion research due to differences in the attitude "strength" or "crystallization" (Petty and Krosnick 1995), among other factors. An example of a cause of differences in judgment—the way in which retrieved information is combined—is differences in motivation; for example, if one group answers the question after a long list of filter questions and the other does not. Response mapping differences may occur if the meanings of the labels are not equal across countries (Harkness et al. 2003), not equally spaced (Oberski et al. 2008) or if there are differences in response style (Baumgartner and Steenkamp 2001; Harzing 2006). Finally, differences in the overall strategy of response may be related to differences in task difficulty, respondent ability, and motivation to perform the task (Krosnick 1991).

Factor analysis and IRT model the process by which the respondent arrives at an answer, given the "true" variable of interest. This relationship between the underlying score and the respondent's final answer at is called the "response

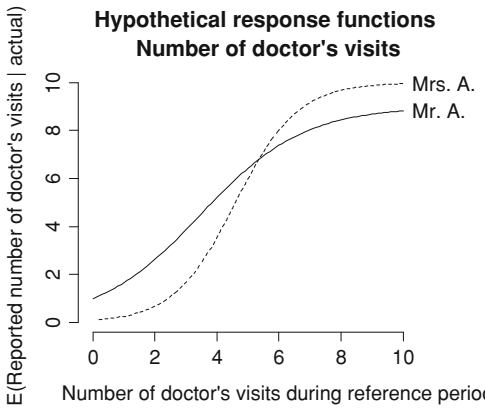


Fig. 27.4 Hypothetical response functions to a question querying number of doctor's visits for two respondents

function." Invariance (equivalence) testing and DIF are the comparison of these response functions across the groups to be compared.

For illustration, two hypothetical response functions are given in Fig. 27.4. In a survey, both Mrs. and Mr. A. are asked how many times they visited the doctor in a certain period. The plot shows the expectation of the number reported given the actual number of visits for both respondents. Of course, this expectation cannot be observed in practice; the curves represent a series of counterfactuals, the answers respondents would give in expectation if they had visited the doctor a certain number of times.

Figure 27.4 shows that Mr. A, if he cannot recall any visit in the reference period will still report one; this might happen, for instance, if he incorrectly places a prior visit in the reference period (forward telescoping). If he has had close to 10 visits, he might be ashamed to admit it and reports fewer. The effect is that Mr. A overestimates the low and underestimates the high numbers. The reverse is the case for Mrs. A.: she underreports low numbers, claiming "it was nothing," while as the numbers get higher she tends toward overestimation. She might include events that happened after the reference period, for instance (backward telescoping).

The final result of the differences in response functions shown in Fig. 27.4 is that, for example, if Mr. and Mrs. A. had both visited the doctor 10 times in the reference period, Mrs. A. will report 10 visits, while on average Mr. A. reports 8.8 visits. A researcher might be tempted to conclude that Mrs. A. is more ailed than her spouse. However, this conclusion would be incorrect. In general, *any conclusion resulting from the comparison of the reported number of visits of Mr. and Mrs. A can be potentially explained by the difference in their response functions.*

If Mr. and Mrs. A. are representative of their genders, for instance, a statistically significant difference in subclass means by gender of the number of doctor's visits, has a difference in response function as a possible alternative explanation. Invariance testing, differential item functioning, and anchoring vignettes are all methods of attempting to rule out this alternative explanation.

The logic behind these methods when applied to the comparison of means and relationship parameters is demonstrated by considering the comparison of a three-item "social trust" scale across countries in the ESS (e.g. Reeskens and Hooghe 2008).

The three items assumed to measure "social trust" in this questionnaire are:

PPLTRST	*Would you say that… most people can be trusted, or that you can't be too careful in dealing with people?* (0–10)
PPLFAIR	*…most people would try to take advantage of you if they got the chance, or would they try to be fair?* (0–10)
PPLHLP	*…most of the time people try to be helpful or that they are mostly looking out for themselves?* (0–10)

Figure 27.5 shows the country means on one of the three items, PPLHLP. Confidence intervals using the survey weights are also shown. It can be seen that Greece has the lowest mean on this item and Denmark the highest. Below we will illustrate comparisons by considering Greece (GR), Poland (PL), and the United Kingdom (GB). For example, the difference of 0.18 between the mean of the PPLHLP item in Greece and Poland is significant ($p = 0.037$). The differences with the United Kingdom are

also clearly statistically significant. However, it is not clear whether such differences can be explained by variation in response functions over the countries.

There are many models that have as a goal the estimation of the response function. The most commonly used are the multiple group factor model (or MGSEM), and Item Response Theory (IRT) models. These models can only be estimated by assuming an underlying latent variable "Trust," for which the three items PPTRST, PPLFAIR, and PPLHLP form a scale. This model is shown as a graph in Fig. 27.6. The graph shown there may represent different types of model: factor, IRT, and latent class models can all be applied. The key assumption is that of conditional independence of the items given the latent Trust variable (uncorrelatedness for factor models).

A literature has developed around the aim of establishing whether the measures can be compared across groups: the procedure of so-called "invariance" testing (Meredith 1993; Steenkamp and Baumgartner 1998; Millsap et al. 2007; Davidov et al. 2010). Invariance testing aims to assess whether (1) the model shown in Fig. 27.6 holds in all countries ("configural invariance"); (2) the slopes (loadings) are equal across countries ("metric invariance") and; (3) the intercepts are equal in addition to the slopes ("scalar invariance"). Under the model assumptions, this amounts to evaluating whether response functions vary over countries. When scalar invariance holds, the groups are said to exhibit "full score equivalence"; when only metric invariance holds, "scale equivalence."

Reeskens and Hooghe (2008) performed such invariance tests on the social trust data presented here using the model shown in Fig. 27.6. Employing multiple group structural equation models (linear factor analysis), they concluded that the three items are not scalar invariant across countries, but that there is metric invariance.[4] What this means is illustrated in the left-hand side of Fig. 27.7.

Figure 27.7 plots the expected value of the PPLHLP item for different values of the theoretical latent factor "Trust." In the case of the linear factor models employed by Reeskens and Hooghe (2008) this is simply a matter of plotting *(intercept)* + *(loading) Trust.* This has been done for the United Kingdom (GB), Greece (GR), and Poland (PL). That metric invariance holds can be seen because the lines are parallel. The lack of scalar invariance is reflected in the distance between the response functions for the different countries.

Overall, Fig. 27.7 shows that there is variation in estimated response functions across these three countries. In the factor model, a British person is expected to choose a value 0.7 higher than a Greek person, and a Greek person chooses values 0.6 above those chosen by Poles, even when their unobserved Trust scores are all equal. For instance, Fig. 27.5 showed a statistically significant difference between Poland and Greece of 0.18. But this is much less than the item bias of 0.6 shown in Fig. 27.7. Therefore the raw means plotted in Fig. 27.5 should not be taken at face value: the lack of scalar invariance causes the countries to be incomparable on their raw means.

The factor model commonly used in the invariance testing literature is not the only possible model to estimate response functions based on the graph in Fig. 27.6. Other possibilities are latent class analysis (Kankaraš et al. 2010), and IRT modeling with its theory of differential item functioning (Mellenbergh 1989; Millsap and Yun-Tein 2004). As an illustration of the differences and similarities of these approaches, the right hand side of Fig. 27.7 provides estimated item response functions for the PPLHLP item based on an IRT model.[5]

Figure 27.7 shows that the spaces between the IRT model response curves at the mean (zero) are approximately equal to the spaces

[4] A test of whether the factor model holds in each country is not possible in this case because with only three indicators the model without equality constraints has zero degrees of freedom.

[5] A two-parameter normal ogive model was estimated. Expected values were calculated by multiplying country-specific item characteristic curves with the scores 0–10 and summing over categories.

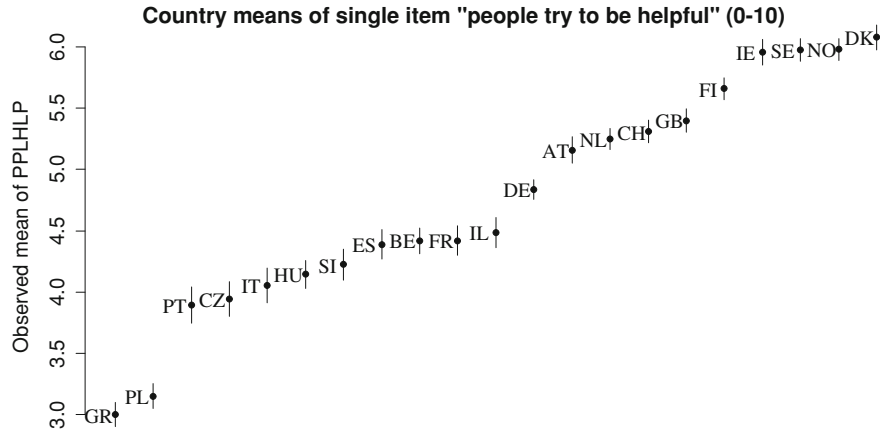


Fig. 27.5 Mean of the item PPLHLP in different countries of round 1 of the ESS, with 2 s.e. confidence intervals

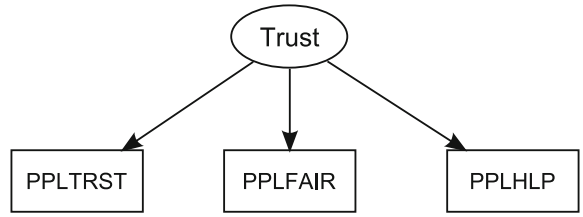


Fig. 27.6 Latent variable model of the three observed indicators

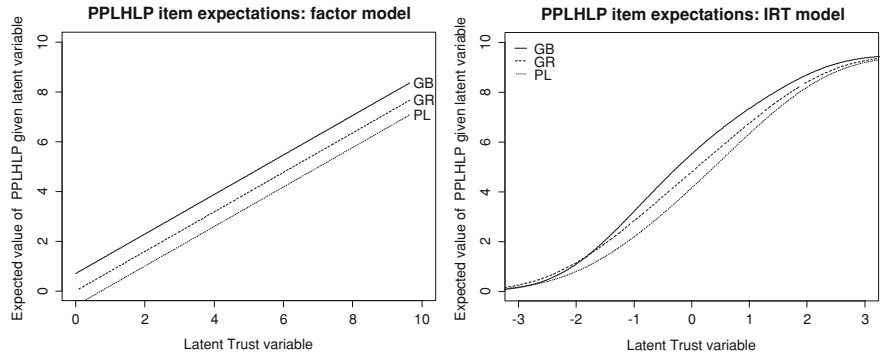


Fig. 27.7 Item response curves for United Kingdom (GB), Greece (GR), and Poland (PL). *Left* factor analysis; *right* IRT

between the parallel response curves produced by the factor model. Away from the mean, however, the IRT model predicts less score inequivalence than the factor model. In this model, whether two scores from different countries on the PPLHLP item can be compared would depend on the score on the underlying scale.

Another point of interest in this comparison is that even though only the intercept parameters (thresholds) were allowed to vary over countries, mimicking the metric invariance found in the factor analysis, the lines in Fig. 27.7 are not exactly parallel. This signals that lack of scalar invariance has more far-reaching consequences

under the IRT model than under the linear factor analysis model.

It should be remarked that Fig. 27.7 shows the lack of equivalence only for the single item PPLHLP. One might also compare means on the composite score of the three trust items together, as an estimate of the social trust construct. In this case the lack of invariance for one item does not automatically imply lack of invariance for the mean of the composite score (Byrne et al. 1989).

Scalar and metric invariance guarantee the comparability of means/totals and covariances, respectively. Regression coefficients, correlations, and other measures of bivariate association might still be incomparable, however. These quantities are affected by the *reliability* of the measures (Fuller 1987). Under the true score model implied by factor analysis, the relationship between the true and observed regression coefficient can be expressed as

$$\beta_{YX} = \beta_{yx} \cdot \kappa_X,$$

where Y and X are the observed variables, y and x are their error-free counterparts, and κ_X is the reliability "ratio" (Fuller 1987) of the observed independent variable.

If the reliability differs across groups, regression coefficients will be affected, since the observed coefficient equals the true coefficient multiplied by the reliability of the independent variable. A correlation or standardized regression coefficient is affected by the reliability of both variables. If reliabilities differ, observed regression coefficients and correlations will differ also even if the true difference across groups is zero.

Oberski et al. (2007) showed that there can be considerable differences in the reliability across countries. The same can be seen at the bottom of Fig. 27.2 (earlier in this chapter) for the reliability of the single variable level of education. Figure 27.8 shows the composite score reliability for the trust scale in different countries, with 2 s.e. intervals. It can be seen that reliabilities range between 0.6 and 0.8 over countries.

To illustrate the consequences of these differences, suppose a researcher regressed voting for right-wing political parties on the average of the three social trust items in Austria and France. Using data from the ESS round 2, in Austria the coefficient for the linear regression of voting for the FPÖ on the average of the three trust items is −0.0203 (0.004). In France the same coefficient for voting for the Front National is −0.0204 (0.005). These two coefficients are remarkably close to each other, suggesting that in both countries the relationship between social trust, and voting for right-wing parties is equal.

Figure 27.8 shows that the reliability of the trust score is 0.8 in Austria but only around 0.6 in France. This suggests that the *true* relationship is stronger in France than it is in Austria, possibly threatening the comparability of the regression coefficients. Correction for unreliability changes the regression coefficients to −0.025 and −0.033 for France and Austria, respectively. Even after correction for differential measurement error, then, the two coefficients are similar, and not statistically significantly different from each other. The same result is obtained using probit regression with correction for differential measurement error[6]: the p-values for the test of different regression coefficients across countries before and after correction are 0.987 and 0.364, respectively. Thus, in this case the differential measurement error shown in Fig. 27.8 did not change the result of the cross-country comparison. A case where the cross-country comparison *is* radically different after correction was already shown in Fig. 27.2 (earlier in this chapter).

One may also obtain the measurement error estimates from IRT models rather than factor analysis. In that case the total information curve gives the reciprocal of the conditional variance of the IRT trust score. This information curve is shown for six different countries in Fig. 27.9, plotted against the rescaled trust composite score.

The higher the curve in Fig. 27.9 at a point, the higher the reliability will also be (Mellenbergh 1994). For example, at the mean of the reference country (Austria), which is set to zero, the measurement quality is highest in the UK,

[6] The analysis and correction using probit models were done using Mplus 5.2.

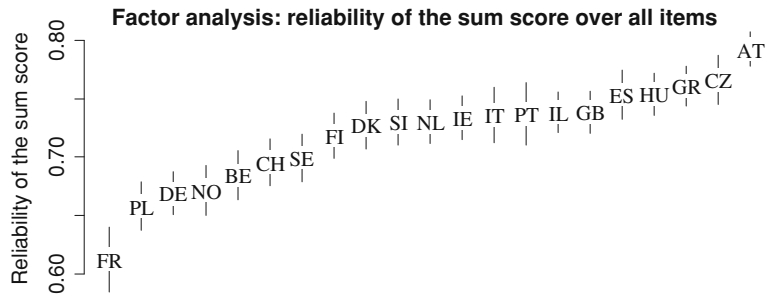


Fig. 27.8 Sum score reliabilities from the factor model with approximate 2 s.e. confidence intervals

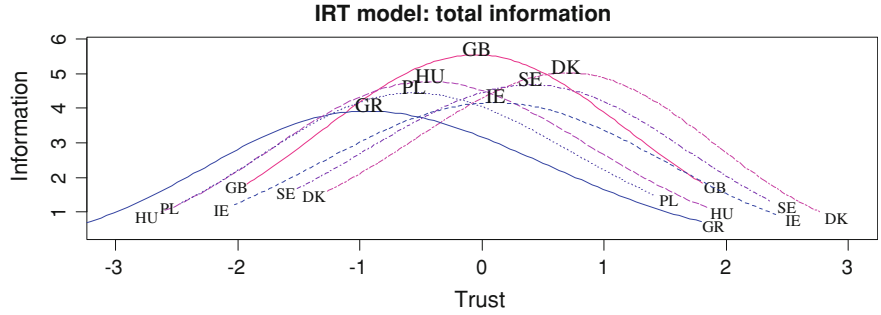


Fig. 27.9 Total information curves in different countries for the trust IRT score

and lowest in Greece. But starting at about 1.3 standard deviations of Trust below the reference country's mean, the measurement quality is highest in Greece.

Overall comparability of measures of association in IRT will depend not only on the differences between the curves, but also on the marginal distribution of the Trust variable in each country (e.g. Oberski 2011). For instance, if Greeks are less trusting than Britons, the average reliability may still be higher in Greece than in the UK. Therefore, the determination of comparability is more complicated than in linear factor models, but the same principles of weak and strong comparability apply.

27.5.1 Models with a Stochastic Systematic Component

So far only models with bias (intercepts) and random measurement error (slopes and/or variance parameters) were considered. Another possibility is that each respondent has their own method of answering questions: there may be a systematic "response style" or "method effect."

A response style occurs when a respondent tends toward certain choices over all questions regardless of the content (Billiet and McClendon 2000). Acquiescence, extreme response style, and middle response style are most often considered. A method effect occurs when a respondent tends toward certain choices over those questions asked by a certain method regardless of the content (Saris and Andrews 2011). For example, if acquiescence were to apply only to agree-disagree scales, it should be considered a method effect rather than a response style.

The common denominator of response style and method effect is that they represent a systematic response error on the part of the respondent that is *stochastic*, i.e., differing over respondents. I use the term "stochastic systematic error" to distinguish this type of error from relative bias, which, confusingly, is sometimes also called "systematic error" (Groves 2004).

The responses to two different variables answered by the same person will be dependent if that person has a stochastic systematic error: response style and method effects cause correlated measurement error. Besides causing possible bias in comparison of means; then, comparisons of *relationships* will also be affected by differing stochastic systematic errors. If the response style or method effect obeys a linear factor model, observed regression coefficients will equal

$$\beta_{YX} = \kappa_X \beta_{yx} + (1 - \kappa_X)\beta_{eYeX},$$

where β_{eYeX} is the regression coefficient that would be obtained by regressing the measurement error of Y on the measurement error of X. The coefficient β_{eYeX} will increase as the variance of style or method factors over survey respondents increase. It can therefore be seen that variations in response style or method factors bias the regression coefficient upwards in each group. Therefore, if there is more variation in response styles in one group than in the other, the regression coefficients will differ, threatening comparability.

In order to either establish comparability or correct for differing method or style influences, special models are necessary. For estimating the method factor influence, so-called "multitrait-multimethod" (MTMM) models can be used (Saris and Andrews 2011). Such models, as well as models with style factors are discussed in more detail by Billiet and Matsuo (Chap. 10).

Another method to estimate individual response functions is that of "anchoring vignettes" (King et al. 2004). An example of this method is given in Fig. 27.10. A question on degree of problems with mobility is asked each time of the same respondent using different stimuli. Afterward, the same question is asked of the respondents themselves. Respondents' answers to the "self" questions can then be corrected for the individual response functions, as estimated from the responses to the vignettes.

Under certain model assumptions, each individual response to the "self" questions can be corrected for that individual's response function. Therefore, cross-group comparison on the transformed scores cannot be threatened by measurement incomparability if the assumptions of the model hold.[7]

Figure 27.11 shows a very rudimentary example analysis. Response functions for 15 respondents are plotted, as estimated from their answers to the vignettes. Each of the five columns in Fig. 27.11 shows a respondent who gave the response 1, 2, 3, 4, or 5 to the "self" question, respectively. These responses have been "mapped back" onto the respondents' response functions, shown by the vertical dotted lines. The three rows represent respondents from three different continents. Thus, differences between the vertical lines over rows represent evidence of incomparability of the responses. For example, respondent 106246 in the top left corner and respondent 109579 just below that gave the same response (1), but this implies a much lower true opinion for the second respondent than for the first.

Figure 27.11 serves only as an illustration of the principle behind anchoring vignettes. In practice, response functions are estimated with a more advanced model than the rather simplistic linear regression employed here (King et al. 2004; Wand et al. 2007).

Factor analysis, IRT, MTMM models, response style models, and anchoring vignettes are all measurement models intended to estimate the response function. These models are relevant for cross-group comparability because they can estimate variation in response functions over the groups to be compared under the appropriate model assumptions. The models are used for different purposes:

1. Establishing equivalence through invariance testing or DIF analyses (strong comparability);
2. Evaluating whether inequivalence is large enough to invalidate cross-group comparability (weak comparability);
3. Correction for inequivalence, if possible, either through

[7] The principle of anchoring vignettes is identical to that of response function analysis in classical psychophysics.

Question: *Overall in the last 30 days, how much of a problem did [name/you] have with moving around?*

Response categories:

1. *None*
2. *Mild*
3. *Moderate*
4. *Severe*
5. *Extreme/Cannot Do*

Vignettes:

V1. [Mary] has no problems with walking, running or using her hands, arms and legs. She jogs 4 kilometres twice a week.

V2. [Rob] is able to walk distances of up to 200 metres without any problems but feels tired after walking one kilometre or climbing up more than one flight of stairs. He has no problems with day-to-day physical activities, such as carrying food from the market.

V3. [Anton] does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry the groceries and do some light household work.

V4. [Vincent] has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy.

V5. [David] is paralyzed from the neck down. He is unable to move his arms and legs or to shift body position. He is confined to bed.

Fig. 27.10 Anchoring Vignettes for mobility, used in World Health Survey instrument in 2002. *Source* http://gking.harvard.edu/vign/eg/?source_extra=mobility.shtml

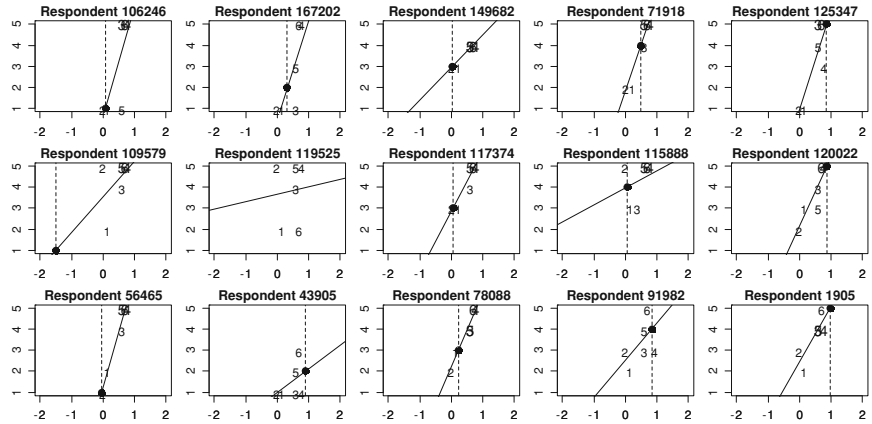


Fig. 27.11 Anchoring vignettes: estimated response functions for respondents from Eastasia (*top*), Eurasia (*center*), and Oceania (*bottom*). The *dotted line* shows the estimated "self" score when the response is mapped back onto the response function. Each column shows randomly chosen respondents giving answers in categories 1 through 5, respectively

a. estimation directly in the model, or
b. postestimation adjustments of comparison parameters (two-step approach).

The first two applications, establishing weak or strong comparability, were already discussed. Correction of the comparison for inequivalence can be done either by direct estimation in the model conditional on comparability (Byrne et al. 1989), or by adjustment of the comparison after the estimation, as was done in Fig. 27.2, for instance (see also Saris and Gallhofer 2007). The three applications are not mutually exclusive, but complement each other.

27.6 Conclusion

Survey measurements are subject to survey errors; thus, whenever different groups are compared on some survey statistic, survey errors in each of the groups may invalidate that comparison. The existence of this possibility was called strong incomparability. The possibility that the effect of the total survey errors is strong enough to invalidate the particular comparison made was termed weak incomparability.

Incomparability may stem from the "representation" side of total survey error, or from the "measurement" side. Traditionally, the literature has focused on measurement issues, i.e., the fields of IRT, DIF, invariance testing, MTMM modeling, response style modeling, and anchoring vignettes.

These approaches do not usually consider the possibility that nonresponse, frame, or other selection biases may also invalidate comparisons. It was shown here that recent developments in the "representation" side, in particular the R-indicator, may in some cases provide useful evidence as to the comparability of survey statistics with respect to such representation errors. Future work in this field may extend the notion of comparability to the representation side.

The methods and models discussed in this chapter may be employed to prevent, detect, and correct for incomparability. The purpose of this chapter was to give the reader an impression and demonstration of the issues and possible solutions, but it was not possible to provide an exhaustive overview here. For further information the reader is referred to the specialized literature cited at the end of this chapter.

References

Alwin, D. F. (2007). *Margins of error: A study of reliability in survey measurement* (Vol. 547). New York: Wiley.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(1), 143–156.

Bethlehem, J. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics, 4*(3), 251–260.

Billiet, J. B., & McClendon, M. K. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608–628.

Boateng, S. K. (2009). Significant country differences in adult learning. *Population and social conditions*, Eurostat statistics in focus, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-09-044/EN/KS-SF-09-044-EN.PDF .

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456.

Christian, L. M., & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly, 68*(1), 57.

Cohany, S. R., Polivka, A. E., & Rothgeb, J. M. (1994). Revisions in the current population survey effective January 1994. *Employment and Earnings, 41*, 13.

Davidov, E., Schmidt, P., & Billiet, J. (2010). *Cross-cultural analysis: Methods and applications*. New York: Taylor and Francis.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2008). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York: Wiley.

Donsbach, W., & Traugott, M. W. (2008). *The SAGE handbook of public opinion research*. London: Sage Publications Ltd.

Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.

Ganzeboom, H. B. G., & Schröder, H. (2009). *Measuring level of education in the European social survey. Keynot speech.* Presented at the European survey research association (ESRA), Warsaw.

Goudy, W. J. (1976). Nonresponse effects on relationships between variables. *Public Opinion Quarterly, 40*(3), 360.

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). New York: Wiley.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly, 72*(2), 167.

Harkness, J. A. (2003). Questionnaire translation. *Cross-Cultural Survey Methods, 325*, 35.

Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., et al. (2010). *Survey Methods in Multicultural, Multinational, and Multiregional Contexts* (Vol. 552). New York: Wiley.

Harkness, J. A., Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire design in comparative research. *Cross-Cultural Survey Methods*, *325*, 19–34.

Harzing, A. W. (2006). Response styles in cross-national survey research. *International Journal of Cross Cultural Management, 6*(2), 243.

Hoffmeyer-Zlotnik, J. H. P., & Harkness, J. A. (2005). *Methodological aspects in cross-national research*. Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA).

Holland, P. W. (1982). *Test equating*. New York: Academic Press.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology, 16*(2), 131–152.

Jowell, R. (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. London: Sage Publications Ltd.

Kankaraš, M., Moors, G., & Vermunt, J. K. (2010). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: methods and applications*. New York: Taylor and Francis.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1), 191–207.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology, 5*(3), 213–236.

Lessler, J. T., & Forsyth, B. H. (1996). A coding system for appraising questionnaires. In N. Schwarz & S. Sudman (Eds.), *Answering questions* (pp. 259–292). San Francisco: Jossey-Bass.

Malhotra, N., & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. *Political Analysis, 15*(3), 286–323.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*(2), 127–143.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin, 115*(2), 300.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Millsap, R. E., Meredith, W., Cudeck, R., & MacCallum, R. (2007). Factorial invariance: Historical perspectives and new problems. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: historical developments and future directions* (p. 131). Mahwah, NJ: Lawrence Erlbaum Associates.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*(3), 479–515.

Oberski, D. (2011). *Measurement error in comparative surveys*. Tilburg: Tilburg University.

Oberski, D., Gruner, T., & Saris, W. E. (2011). *SQP 2*. Retrieved from http://www.sqp.nl/.

Oberski, D., Saris, W. E., & Hagenaars, J. (2007). Why are there differences in measurement quality across countries. In: G. Loosveldt, D. Swyngedouw & B. Cambré (Eds.), *Measuring meaningful data in social research*. Leuven: Acco.

Oberski D., Saris W.E. & Hagenaars J. (2008). Categorization errors and differences in the quality of questions across countries. In: T. D. Johnson & M. Braun (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC)*. New York: Wiley and Sons, Ltd.

Petty, R. E., & Krosnick, J. A. (1995). *Attitude strength: Antecedents and consequences.* Mahwah: Lawrence Erlbaum Associates, Inc.

Reeskens, T., & Hooghe, M. (2008). Cross-cultural measurement equivalence of generalized trust. evidence from the european social survey (2002 and 2004). *Social Indicators Research, 85*(3), 515–532.

Saris, W. E. (1988). *Variation in response functions: A source of measurement error in attitude research* (Vol. 3). Amsterdam: Sociometric Research Foundation.

Saris, W. E., & Andrews, F. M. (2011). Evaluation of Measurement Instruments Using a Structural Modeling Approach. In: P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Thiowetz & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 575–597). New York: John Wiley & Sons.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research* (Vol. 548). New York: Wiley.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology, 35*(1), 101–113.

Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J., & Zhang, L. (2008). *Statistical Properties of R-indicators. RISQ Work Package*, *3*. Retrieved from http://www.risq-project.eu/papers/RISQ-Deliverable-2-1-V2.pdf.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, *25*, 78–90.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. Piscataway: Transaction Publishers.

Stoop, I., Billiet, J., & Koch, A. (2010). *Improving survey response: Lessons learned from the European Social Survey*. New York: Wiley.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research* (Vol. 1). London: Sage Publications, Inc.

von dem Knesebeck, O., Verde, P. E., & Dragano, N. (2006). Education and health in 22 European

countries. *Social Science and Medicine, 63*(5), 1344–1351. doi:10.1016/j.socscimed.2006.03.043.

Voogt, R. J. J. (2004). *I'm not interested: nonresponse bias, response bias and stimulus effects in election research*. Amsterdam: University of Amsterdam.

Wand, J., King, G., & Lau, O. (2007). Anchors: Software for anchoring vignette data. *Journal of Statistical Software, 42*, 1–25.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. :London: Sage Publications, Inc.

Zavala, D. (2011, February 7). Deviations found through SQP coding in the ESS Round 5 questionnaires. Report given to the European Social Survey's national coordinator's meeting.

Employee Surveys as Catalysts for Change: Turning Data into Action

28

Patrick Hyland and Orly Dotan-Eliaz

28.1 Introduction

Employee surveys are only useful if they lead to strategic organizational action. As survey research specialists at Sirota Survey Intelligence, this principle guides our work. In fact, we tell our clients that unless their organization is committed to turning their data into action, they should not conduct a survey. As one insightful executive observed over 50 years ago, "An attitude survey is like a hand grenade—once you pull the pin you have to do something with it. Otherwise it may hurt you rather than help you" (Vitales 1953, p. 394). Without action, employee surveys waste time and money, raise false hopes, and undermine workforce motivation.

Launching a survey without planning for action is a common mistake. Many researchers assume that if they prepare a sound instrument, develop effective communications, design an efficient administration process, and deliver feedback reports, then leaders and managers will review their results and take appropriate actions. Based on our observations, this is not always the case. Even when leaders and managers have the best intentions, a number of factors can prevent organizations from tapping the full potential of their data.

Over the past 40 years, our firm has developed a deep understanding of what it takes to turn survey data—particularly employee engagement data—into action. We have found that organizations are more likely to be successful at actioning when eight core elements are in place:

- **C**ommitment
- **A**ccountability
- **T**raining
- **A**lignment
- **L**earning Practices
- **Y**ardsticks
- **S**upport Systems
- **T**ools and Technology

Considered together, these eight elements comprise our CATALYST model for survey actioning. In this chapter, we will explore each element and explain why it is necessary for survey success. We will highlight roadblocks and blind spots that can stymie action efforts. And we will provide a series of recommendations designed to ensure your employee surveys promote strategic action, positive change, and better organizational performance.

28.2 Actioning: The Weak Link in Most Employee Survey Research Programs

We have entered the age of human capital (Fitz-enz 2000; Burke and Cooper 2008). Based on mounting evidence showing that employees

P. Hyland (✉)
Director of Research & Development,
Sirota Survey Intelligence,
One Manhattanville Road, Purchase,
NY 10577-2128, USA
e-mail: phyland@sirota.com

O. Dotan-Eliaz
69-39 Yellowstone Blvd. Apt 304, Forest Hills,
NY 11375, USA
e-mail: orlydotan@gmail.com

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_28, © Springer Science+Business Media New York 2012

represent one of the few unique competitive advantages that companies have (Pfeffer 1994, 1998; Ulrich and Brockbank 2005; Ulrich and Smallwood 2003), more organizations are seeking ways to attract and retain the best and brightest. Increasingly, business leaders are realizing that to win the war for talent, they need to offer their employees compensation that is competitive, a job experience that is fulfilling, and a work environment that is motivating.

This increased focus on human capital may explain the widespread utilization of employee surveys. According to various sources, many Fortune 1,000 firms now conduct a regular survey of their employees (Kraut 2006). At Sirota, we have seen a notable increase in client demand over the past decade, with a growing number of clients conducting on-boarding, engagement, exit, and life cycle surveys to develop a deeper and broader understanding of their employees.

Considering the widespread utilization of employee attitude surveys, one might assume that most organizations are adept at using their surveys to drive positive change. But data suggest otherwise. For example, in our normative database—which consists of data from over one million employees working in over 40 companies located around the globe—we find that just 48% of employees believe the results of their last organizational survey were used constructively by management. Out of 127 survey items in our 2010 normative database, these results rank 123rd, falling 22% points below the average benchmark item score of 70% favorable. In a recent study, Kraut found similar trends: survey practitioners identified lack of survey action as the greatest problem associated with employees surveys (Kraut 2006). Based on these results, Kraut concluded that "lack of action is the Achilles heel of the survey process" (Kraut 2006, p. 4).

28.3 Preparing Organizations for Action

So why is it so hard to turn survey data into strategic action? Many survey practitioners, leaders, and managers are well aware of the four basic steps of action research: diagnose issues based on the data; plan for change; take action, and evaluate progress (Coghlan and Brannick 2005).

But knowing what to do and how to do it are two different things (Pfeffer and Sutton 2000). Survey actioning is a complex task, requiring careful coordination between people, resources, and performance goals. Based on our experience, survey actioning campaigns only succeed when the entire organization is prepared for change. This requires that survey administrators think of themselves as more than just researchers facilitating a survey program; instead, they need to think of themselves as change agents implementing a total systems intervention (Church and Oliver 2006). It also requires that leaders, managers, employees, and critical departments—like Human Resources (HR)—are aware of their roles and responsibilities in the actioning process.

Fortunately, small changes can make big differences. In the following section, we present an eight-step process designed to create *readiness for action* in your organization. By following these steps, you can ensure that your survey becomes a catalyst for positive change.

28.4 Step 1: Build Commitment for Action

In *Leading Change* (1996), John Kotter argues that the first step in leading any organizational change effort is to establish a burning platform—a compelling reason why change is needed. The same is true when it comes to survey actioning. The first step in any successful employee survey effort is building a commitment for action.

The best way to build commitment for action is to ensure that leaders and managers understand two important points about employee surveys. First and foremost, inaction can erode employee morale. This is because surveys heighten awareness, generate curiosity, and create expectations for change. As Nadler states, "the mere act of data collection generates

energy around those activities that are being measured" and "creates expectations that the data collection will have consequences" (1996, p. 180). This means that if your survey asks employees to evaluate their job, their boss, and their benefits, they are probably going to be more aware of these topics in the days and weeks following the survey. They also may wonder how their colleagues feel about these topics. And they may expect that their job, their boss, and their benefits will change, particularly if there is widespread dissatisfaction with each.

Leaders and managers need to realize that every aspect of the survey process conveys a message to employees (Kraut 1996). When leaders and managers spend time evaluating survey feedback and designing appropriate response strategies, they send a strong message that employees are valued stakeholders. But when action is not taken, employees can become frustrated and despondent (Burke 2002).

We see evidence of this in the data we collect. As one employee recently stated in her written comments, "I believe this survey is a good idea, but resent the fact that when we did this survey 2 years ago none of our concerns were addressed. We had one local follow up meeting, then nothing again." Consistent with this comment, we have also found that when employees do not believe that their managers will take action based on survey feedback, they are significantly less engaged than employees who do believe their managers will take action.

The second reason leaders and managers should focus on actioning their data is because it makes good business sense. There is robust evidence that employee attitudes impact organizational performance and business results. For example, we have found that companies with highly engaged employees outperform their peers by 20% in the stock market (Sirota et al. 2005). We have also found that:

- Call center employees who are satisfied with their training offer significantly higher levels of service to their customers;
- Managers who are satisfied with their work-life balance are more likely to create a work environment that is collaborative, creative, and stimulating;
- Teams with collaborative work environments tend to have fewer accidents and injuries.

These findings are consistent with a vast body of literature showing that employee attitudes—particularly employee engagement—are related to everything from employee turnover, customer satisfaction, productivity, and profit (see Harter et al. 2002; Judge et al. 2001; Kreitner and Kinicki 2010). Jac Fitz-enz (2000) neatly sums up decades of research by stating "there is clear and abundant evidence that an organization that makes work as fulfilling as possible will develop and retain the most productive workers and enjoy the most loyal customers" (p. xiii).

Some leaders and managers have a tendency to dismiss survey results based on the assumption that employee attitude data is soft, nonstrategic, and unrelated to organizational performance and business success. That is not the case. Employee attitudes can serve as key performance indicators, highlighting organizational strengths and opportunities for improvement. When leaders and managers take their survey data seriously and take action based on their results, they are promoting business success. When they do not, they are missing an opportunity for both personal and organizational development.

Considering the importance of action, survey practitioners should build commitment for change early, at the very start of the survey project. Otherwise, leaders and managers may not realize the critical role they need to play once survey results are delivered. These six recommendations can be used to build a commitment for action in your organization:

1. *Design a survey that is aligned with your strategic goals*. The best way to ensure that survey results serve as key performance indicators is to design a survey that is aligned with your most important business objectives. Unlike traditional employee surveys that focus exclusively on employee satisfaction, strategic surveys assess the people issues that make the greatest difference in business performance (Schiemann and Morgan 2006).

For example, if customer service is a core business goal, strategically designed surveys will include a number of items that measure the extent to which your employees feel they can meet the needs of their customers. Because of the business relevance of strategic surveys, they tend to generate a high level of interest and buy-in from leaders.

2. *Ensure everyone understands the motivational and business imperative for action*. Surveys create expectations for change. When actions are not taken, employee motivation often decreases, and that can impact organizational effectiveness and business performance. To prevent this from happening, ensure that your leaders, managers, and employees understand the importance of taking action based on the survey results.
3. *Create commitment from the top down*. Senior executive buy-in is critical to the success of any organizational initiative, particularly change efforts (O'Toole 1995). When senior leaders communicate that they regard the employee survey as a strategic business tool and that they expect survey results to lead to action, it sends a strong message throughout the organization (Burke et al. 1996).
4. *Contract for Action*. Once senior leadership buy-in has been attained, it is important to continue driving commitment throughout the organization. One way to do this is to contract for action with business unit leaders, middle managers, and frontline supervisors. Ask them to make a public statement to their bosses, their peers, and their employees that they will take action based on survey results. Such public statements set clear expectations, create a healthy level of peer pressure, and prime leaders, managers, and employees for post-survey action.
5. *Plan to conduct a follow-up survey*. In organizations, people pay attention to what gets measured. Nothing builds commitment to action like the promise of a follow-up survey. We recommend conducting an annual engagement survey at least once a year. This helps leaders, managers, and employees monitor progress and evaluate action efforts. Less frequent surveying (e.g., once every 24 or 36 months) can convey a lack of real commitment and lead to a loss of momentum.
6. *Include Employee Survey Results in Manager Performance Reviews*. One powerful way to communicate that the employee survey is important is to hold leaders and managers accountable for results. Influenced by Kaplan and Norton's (1992) balanced scorecard approach, some of our clients include employee survey results in their performance management process. This sends a clear message that employees—and their opinions—are valued.

28.5 Step 2: Clarify Who is Accountable for What

Once a clear commitment for action has been established, leaders and managers may start raising a number of questions. What am I responsible for doing with my survey results? What kind of improvement strategies can I pursue on my own, and which ones will require approval? What if my employees are concerned about issues that I have no control over? The best way to address these questions is to develop an action accountability plan.

Like any organizational change effort, strategic survey actioning is a complex task, requiring input, analysis, feedback, and effort from various stakeholders (Nadler 1977). By establishing an action accountability plan early in the process, survey practitioners can ensure that leaders, managers, employees, and critical decision-makers from different departments and functions are prepared to work together. Otherwise, a number of process problems could emerge. Leaders and managers may struggle with role ambiguity or role overload, resulting in increased stress and decreased personal effectiveness (Kreitner and Kinicki 2010). They may also solve their problems in silos, failing to communicate and coordinate their actions into an effective organization-wide campaign. Or they may do nothing at all, assuming that

their individual efforts will not be able to effect meaningful organizational change.

Effective accountability plans can prevent these types of problems from emerging by specifying what needs to be done, by whom, and by when. Before launching an employee survey, we recommend developing an accountability plan that clarifies four things:

1. *Define post-survey responsibilities*: In any change effort, scripting the critical steps can help overcome resistance and anxiety (Heath and Heath 2010). By specifying what you expect leaders and managers to do once they receive their feedback reports, you can ensure that the action planning process will be executed in a uniform and consistent way across the organization. At a minimum, we recommend that leaders and managers do at least these five things:
   - Review and analyze survey results
   - Conduct feedback meeting with direct reports to identify action priorities
   - Discuss action priorities with their bosses
   - Develop and record action plans
   - Execute action plans
2. *Clarify Spheres of Influence*: Employee surveys often focus on a broad array of topics, including items that ask about everything from leader behavior to pay satisfaction. While such surveys are effective when it comes to understanding employee perceptions about a wide range of important organizational topics, they can create confusion when it comes to actioning. As leaders and managers review their results, they may wonder: how am I going to take action on topics—such as pay and benefits—that I have no control over? To prevent these types of concerns from derailing action, it is important to help managers and leaders understand which survey items and topics are within their sphere of influence, and which require actioning support from other leaders or departments. One of the best ways to do this is to preview the survey with leaders and managers—before results are in—and ask them to consider five questions:
   - Which survey items ask about local topics (e.g., my team, my personal leadership style) that are under my control and I can take action on independently?
   - Which survey items ask about departmental or divisional topics (e.g., cross-functional teamwork) that will require coordinated actioning efforts with other leaders in my department?
   - Which survey items ask about general organizational issues (e.g., tools and resources, pay and benefits) that will require support from staff functions such as Information Technology or Human Resources?
   - Which survey items ask about strategic organizational topics (e.g., strategy, culture, vision) that will require enterprise-wide senior leadership support?
   - Which survey items fall under multiple spheres of influence and will require coordinated actioning efforts from multiple leaders?

   This exercise can help leaders and managers prepare for action and establish a clear sense of collective accountability. Responses can be discussed, debated, and charted, creating a clear roadmap for responsibility and action once results are delivered.

Survey items	Sphere of influence			
	Local team	Department/ division	Line function	Enterprise wide
1. My manager recognizes me when I do a good job	X			
2. We receive a high level of service from other teams in our division		X		
3. Overall, I am satisfied with my pay			X	
4. Employees at this company are treated fairly	X	X	X	X

3. *Ensure everyone knows they have a role to play*: Successful action campaigns require coordinated effort from employees throughout the organization. When defining roles and responsibilities for senior leaders, department heads, and immediate managers, it is important to consider their level of authority and their sphere of influence. We generally recommend that senior leaders serve as the survey champions, responsible for articulating the need for action and developing any organization-scale actions that involve the company's vision, strategy, structure, systems, and culture. Department, business unit, and functional leaders are often best suited to serve as survey action facilitators, focused on tackling any division-specific operational or performance issues while making sure that managers and supervisors throughout their divisions have the support they need to follow up on results. HR leaders and managers typically serve as survey action coaches and consultants, providing guidance, support, and structure throughout the actioning process. This leaves immediate managers and supervisors with the most important role: local change agent. In this role, immediate managers are responsible for developing local action plans, making sure they are executed (along with any organization-wide or division-wide initiatives), and improving the team environment on a day-to-day basis. This is often a tall task, so it is important to make sure that employees realize they have a critical role to play as well—providing feedback, ideas, and support for action plans. For any organizational change effort to be successful, participation from all key stakeholders—including employees—is critical (Burke 1982).
4. *Set clear expectations about the size and scope of actioning efforts*: It is important to set clear expectations about the number of action plans you expect leaders and managers to pursue; otherwise they may fall into the trap of trying to do too much or too little. We have found that leaders and managers are more likely to succeed when they focus on a limited number of improvement opportunities. For most of our clients, we recommend that leaders and managers focus on developing action plans for no more than three critical priorities. As we'll discuss below, these action plans often integrate local, divisional, and organizational priorities.

In *Confidence* (2004), Rosabeth Moss Kanter's exploration of why some companies flourish and others fail, she writes: "Accountability is the first cornerstone of confidence, a pillar of winning streaks. When accountability crumbles—when troubles provoke denial, or people cover up their own mistakes or find an enemy to blame—winning streaks end" (p. 182). The same is true when it comes to survey actioning. With accountability in place, leaders and managers know what needs to be done, by whom, and by when. Without it, actioning efforts can become disorganized and chaotic, and surveys run the risk of doing more harm than good.

28.6 Step 3: Train Leaders and Managers How to Turn Data into Action

Once a foundation of commitment and accountability has been established, it is important to make sure that leaders and managers have a clear understanding of the specific skills, steps, and activities required to action their data. Training can help.

Surveys, like all assessment processes, have the potential to generate anxiety in organizations (Burke et al. 1996). Left to their own devices, anxious leaders and managers may fall into any one of a number of fear-based traps when they receive their feedback report. Some managers may over-personalize their data, assuming results are a referendum on their personal leadership capabilities while failing to consider contextual factors that may be impacting employee attitudes. Others may under-personalize their feedback, incorrectly determine that root causes are beyond their control, and conclude there is nothing they can do to improve their organization. Some may rush to solutions, setting a course of action before getting feedback and input from direct reports, colleagues, and bosses. Others may suffer from

analysis paralysis, wasting weeks or months over-analyzing their data.

One of the best ways to help leaders and managers overcome their anxieties and avoid these mistakes is to break the actioning process into discrete steps. In our training program, we focus on ten key concepts.

1. *Understanding the Feedback Reports*: We start with the basics, making sure leaders and managers understand what their feedback reports contain and how to read graphs and charts. We also explain the difference between dimension and item level results, and help leaders understand how to use both levels of data to develop an understanding of the big picture and specific details.
2. *Interpreting the Numbers*: Once leaders and managers understand the content of their reports, we focus on their actual results. Many come to training with a grade school mentality about assessment scores, assuming that if they received anything less than 90% favorable on any item they are not doing well. We counter this mindset by reminding participants how hard it is to attain consensus among adults: getting three out of four people to feel favorable about a topic is often a success. We provide leaders and managers with various rules of thumb for evaluating their quantitative results, asking them to consider their results from a climate strength perspective (i.e., what items and dimensions have high vs. low consensus); an internal and external benchmark perspective (i.e., how do my results compare to internal and external norms); and from a historical perspective (i.e., are my results improving, declining, or staying stable).
3. *Processing Initial Reactions*: After participants have had a chance to digest their results, we help them process their emotional and intellectual reactions to the data. We start by reminding leaders and managers that it is normal and healthy to experience a wide range of emotions when receiving survey feedback. We tell them that some results may seem confusing, surprising, or alarming at first, and that is okay. We encourage participants not to rush for answers, explanations, and solutions at this point—that those will come later, based on discussions with direct reports. Instead, we encourage leaders and managers to be mindful of these surprises and turn them into discussion questions. We also encourage leaders and managers to explore their written comments to gain more insight.
4. *Considering the Context*: To help leaders and managers start to make sense of their data, we encourage them to consider contextual factors that could be affecting employee attitudes. Through a series of questions, we ask participants to reflect on a wide range of internal and external factors that employees may be reacting to. In doing so, we help leaders start to develop working hypotheses about the positive and negative dynamics within their work environment.
5. *Understanding Employee Engagement*: While most leaders and managers have heard about the concept of employee engagement, few have a deep understanding of what engagement is, why it is important, and what factors increase or decrease engagement levels. During our training, we explain that engagement represents the motivational capital that exists within an individual, a unit, or an organization. We demonstrate, based on empirical evidence, why it is a valuable resource that can boost company performance; and we provide an overview of our *Three Factor Theory of Human Motivation in the Workplace*, which identifies the key factors that drive employee engagement: achievement, camaraderie, and equity (Sirota et al. 2005). This helps leaders and managers understand what they can do to create a motivating environment for their employees.
6. *Identifying Action Opportunities*: At the core of our training is a fundamental point: surveys are only useful if they lead to strategic organizational action. With this in mind, we ask leaders and managers to

identify their biggest business priorities for the upcoming year. Then we ask them to evaluate their survey data vis-à-vis these business goals and determine if their workforce is attitudinally aligned with these priorities. For example, if a manager's top business goal is to increase innovation in her unit, we ask her to look at her survey results to determine whether she is creating an environment that fosters innovation. We also encourage leaders and managers to look at their results from an employee engagement perspective, identifying any factors that may be getting in the way of employee performance, commitment, and motivation. Using this process, we help managers and leaders start to identify their preliminary areas for actioning.

7. *Preparing to Tell the Story*: Effective surveys spark discussions among leaders, managers, and employees. These discussions start transforming a survey's potential energy into kinetic energy. While numbers, graphs, charts, and statistics are important elements of these discussions, we have found that survey results come to life when they are framed within a story. Stories are "important cognitive events, for they encapsulate, into one compact package, information, knowledge, context, and emotion" (Norman 1994, p. 146). During training, we help leaders and managers weave their results into a clear and concise narrative, integrating their initial reactions, surprises, concerns, contextual analyses, and business goals for the upcoming year. When leaders and managers can convey the story of their data in a brief elevator speech, we know they are ready to share their results with their boss, their peers, and their employees.
8. *Conducting the Feedback Meeting*: Once leaders and managers can convey the story of their data, we prepare them to conduct a feedback meeting with their employees. We emphasize that this meeting is different than most other meetings they usually conduct, and therefore it requires a different mindset. Rather than having all the answers and provide direction, leaders and managers should focus on asking good questions and generating rich discussion. We also train leaders and managers how to create a safe environment, focus on solutions, and listen for employee input and insights.
9. *Evaluating Actioning Opportunities*: Feedback meetings often produce a long list of possible actions. It is important to provide leaders and managers with different tools and frameworks they can use to determine which to pursue. We train leaders and managers to evaluate their actioning opportunities from different perspectives, using questions such as follows:
   - Which opportunities will have the biggest impact on my team from a business performance perspective?
   - Which opportunities will have the biggest impact on my team from an employee engagement perspective?
   - Which opportunities are most feasible? Which are least feasible?
   - Which opportunities are within my sphere of influence? Which ones will require more support from other leaders, managers, or functions?
   - If I were to pursue this opportunity, what resources will I need? How long will it take to complete?
   - How do these opportunities align with our broader organizational objectives and strategies?
10. *Developing and Executing Effective Action Plans*: As a final step, we ensure that leaders and managers know how to develop and execute effective action plans. We start by showing them how to create *SMART* action plans (Kinicki 2009), those that are specific, measurable, attainable, results oriented, and time bound. We provide them with a roadmap for executing their action plans (see Yardsticks below) and a process for keeping momentum going. We also ensure they know how to receive additional support if they feel they need it (see the Support section below).

28.7 Step 4: Align Actions for Maximum Impact

Before leaders and managers start pursuing their plans, it is important to develop a process for aligning their actions. Otherwise, they may start pulling the organization in a number of different directions all at once. As Kaplan and Norton point out, "alignment, much like the synchronism achieved by a high-performance rowing crew, produces dramatic results" (2006, p. 3).

Aligning post-survey actions is critical for three fundamental reasons. First, survey actions create systemic changes in organizations. By implementing actions in one part of the organization, other parts will eventually be affected (Burke 2002, 2006). If survey administrators do not plan for this, change efforts can backfire (Burke 1980). Aligning action plans can ensure that improvement efforts in one team or business unit will not work at cross-purposes with those being implemented in others. Second, post-survey actions often involve multi-level problems requiring multi-level solutions (Klein and Kozlowlski 2000). By aligning action plans, survey administrators can make sure that front-line managers, division heads, and senior leaders are working together to implement local, divisional, and enterprise-wide changes. Finally, aligning action plans can help promote organizational learning. By looking across plans, survey administrators can search for common themes and identify organizational effectiveness and leadership development opportunities.

Aligning action plans can be a challenging task, especially in large organizations. We have found that four coordinating practices can help in this process.

- *Clarify Organization-wide Strategic Actioning Priorities*: Nothing creates alignment and focuses attention like a clear sense of direction. Numerous organizational change scholars, including Kotter (1996) and Beckhard and Harris (1987), emphasize that in order for organizational change efforts to succeed, senior leaders need to articulate a change vision and strategy. When it comes to survey actioning, this means senior leaders should provide employees with an overview of their perceptions of the survey results, highlighting organizational strengths, weaknesses, opportunities, and threats. They should also define any actioning priorities that—because of their strategic importance—they want the organization to focus on. This could include areas like strategy execution, business performance, customer service, internal efficiency, cross-functional teamwork, communications, organizational culture, or leadership. Finally, they should set clear expectations about what they want leaders and managers to do to support the organization's strategic actioning priorities. In some cases, this could require leaders and managers to actively drive specific actions in their unit or department. In other cases, it may simply mean keeping employees aware of an organization-wide initiative that is being driven by a staff group like HR or IT while focusing on local actioning campaigns. In either case, alignment starts with senior leadership providing employees with a clear sense of the enterprise-wide actions the organization will take based on survey results. To ensure that strategic actioning priorities are understood and implemented, some organizations also assign a cross-functional, cross-level team to serve as a guiding coalition (Kotter 1996).
- *Conduct Alignment Discussions*: Once enterprise-wide priorities have been established, the next step is to conduct alignment discussions throughout the organization. In these discussions, leaders meet with the managers that report to them and review their survey results, their actioning priorities, and their action plans. During these discussions, leaders and managers work together to ensure that local actioning priorities are aligned "vertically" with overall enterprise-wide actioning priorities. Additionally, leaders and managers can ensure that action plans are aligned "horizontally" within departments and divisions (Labovitz and Rosansky 1996). With this in mind, we recommend holding alignment discussions in a large group setting, with

all leaders and managers within a business unit or division in attendance.

- *Utilize Alignment Tools*: In addition to alignment discussions, technology-based tools can be used to help leaders and managers share action plans, progress, and learning throughout the organization in a quick and efficient way. We utilize a web-based tool called Action Tracker (described below) to help leaders, managers, and HR professionals access best practices, monitor progress, and communicate virtually about successes and challenges. This can help organizations adjust their actioning efforts quickly.
- *Conduct Alignment Reviews*: Finally, it is important to conduct periodic reviews where leaders and managers meet and discuss their progress, share their learnings, and identify any issues that may be preventing action plans from working. During these meetings, success stories can be shared and short-term wins can be consolidated, adding momentum to the change process (Kotter 1996). We recommend conducting these meetings quarterly, after managers conduct action evaluations (see below) with their teams.

28.8 Step 5: Utilize Learning Practices

Various scholars have argued that in order to succeed in the twenty-first century, organizations need to master the practice of continual learning and the art of organizational change (Lawler and Worley 2006; Burke 2002; Marsick and Watkins 1999). Organizational psychologists have been using surveys to effect organizational learning and change for decades. Influenced by Von Bertalanffy's (1950) general systems theory and Kurt Lewin's model of change (1947), pioneers like Mann (1957) and Likert (1967) showed that survey feedback could be used to improve organizational performance. For decades, employee surveys were considered one of the primary ways to "unfreeze" an organization and create change.

Over the past 60 years, a number of other organizational learning and organization development interventions have been developed. When these interventions are utilized in conjunction with survey feedback, they can produce deeper learnings and more effective action plans for leaders and managers, teams, and organizations. The following techniques can be particularly effective.

- *Reflective Learning Reviews*: The latest research on leadership development (McCall 2010) shows that leaders develop through experience. Challenges and hardships have been shown to be particularly rich learning moments. With this in mind, we recommend that leaders and managers complete a brief post-survey reflection exercise that asks them to consider what they have experienced, how they have grown, and what they have learned since their last survey. This exercise can be conducted in various ways, including one-on-one coaching, check-in meetings with bosses, large group discussions, or personal journaling. By asking leaders and managers to review their survey results and reflect on their personal learnings, the survey can be utilized as a leadership development tool.
- *Appreciative Inquiry*: When teams and units are fixated on problems and unable to find solutions, Appreciative Inquiry (Cooperrider and Whitney 2005) could help. Unlike traditional deficit-based approaches to change—which focus on diagnosing the root cause of problems—Appreciative Inquiry assumes that people, teams, and organizations hold great potential and there is always something that works well within any social system. The promise of Appreciative Inquiry is simple but powerful: by discovering what is working well, dreaming about what could be, designing a plan to achieve these dreams, and delivering on that plan, teams can move beyond their problems to their potential.
- *Positive Deviance Analysis*: Another way to help struggling teams and units develop effective action plans is to conduct a positive deviance analysis. Like Appreciative Inquiry,

Positive Deviance Analysis is a strengths-based, problem-solving approach to creating change (Pascale et al. 2010). At the core of this technique is the observation that in every social system there are certain individuals or group who succeed where others fail. By finding these individuals and groups and by determining the unique behaviors and strategies they use to succeed, organizational members can learn from each other. Survey results can expedite this process, helping HR identify those leaders, units, and teams that excel in areas where others do not. Through interviews and observations, best practices can be identified and shared across the organization.

- *Work-Out*: If survey results indicate that widespread efficiency or cross-functional teamwork issues exist, then a large-scale intervention may be necessary. While a number of options exist, General Electric's Work-Out program (Ashkenas et al. 1999) can be particularly effective when it comes to improving internal process problems quickly. The program consists of three steps. First, an operational area in need of improvement is identified and target goals are set by senior leaders. Then, all employees involved in the process are brought together to identify improvement opportunities. Recommendations are proposed to senior leaders, who in turn approve or reject employee suggestions on the spot. Approved recommendations are then implemented, and improvement efforts are monitored and supported. Work-out recommendations can often be implemented in as little as 60–90 days. By getting employees involved from the start, this intervention can generate the commitment levels, support, and insights that companies need to succeed.
- *Action Learning*: Action learning provides another way for organizations to tackle systemic problems in an innovative way. In action learning programs, organizational members work together in ad hoc teams over a period of time (e.g., 6 months) to solve real organizations problems (Marsick et al. 2002). During the course of the program, coaches provide support, guidance, and feedback, creating an environment that promotes individual learning and group problem solving. Organizations are often looking for ways to develop their high potential employees and future leaders. By developing survey-based action learning programs, organizations can help develop their talent while solving real organizational problems.

By supplementing training with action learning interventions, organizations can ensure that their leaders and managers are developing deeper insights about themselves, their employees, and the organization.

28.9 Step 6: Use Yardsticks to Keep Actioning on Track

One of the best ways to ensure that organization-wide survey action campaigns stay on course is to provide leaders and managers with a series of yardsticks—critical dates, milestones, and progress guidelines—to guide their efforts. Researchers have found that deadlines are critical to team effectiveness, providing a temporal sense of structure that facilitates decision making and performance (Gersick 1988). When it comes to survey actioning, milestone and deadlines can ensure that leaders and managers roll out their actioning efforts in a timely fashion.

Effective yardsticks should guide leaders and managers through four distinct phases of the actioning process, defining the specific activities they need to take during each leg of the journey.

- *Action Planning*: During the first phase of the actioning process, leaders, managers, and employees need to understand, discuss, and digest their data. During this first phase, we recommend that leaders and managers review and analyze their survey results, conduct feedback meetings with employees to solicit their ideas and identify action priorities, and meet with their boss and other key stakeholders to ensure actioning efforts are aligned. The action planning phase is complete once leaders and managers have written and developed effective action plans and committed to pursuing them.
- *Action Taking*: When it comes to creating strategic organizational action, plans are necessary

but not sufficient. Real change requires leaders, managers, and employees to *behave* in new and different ways (Burke 2002). During the action taking phase, organizational members need to focus on translating their action plans into new individual behaviors, group norms, and organizational practices. Leaders, managers, and employees should consciously focus on practicing new behaviors, building new habits, and starting developing deliberate practice routines and rituals (Ericsson et al. 1993). While some new practices and behaviors will take longer to master than others, we recommend that leaders and managers focus on action taking for at least a month before moving to the next phase.

- *Action Evaluation*: Are things working? That is the key question to ask after actions have been put into practice. Two different techniques can help managers and team members determine if actions are working. For evaluations of specific new behaviors, practices, or norms, After-Action Reviews (Sullivan and Harper 1996) can help. In this process, key stakeholders discuss four questions:
  - What did we intend to happen?
  - What happened?
  - Why did it happen that way?
  - How can we improve what happened?

  For a broader evaluation of progress over time, four questions—inspired by Kirkpatrick's training evaluation model (1959)—can be utilized:
  - Do actions seem to be leading to positive change?
  - Are we learning anything new?
  - Have behaviors changed?
  - Are we producing the results (e.g., improved performance, efficiency, customer service) we expected?
- *Action Adjustment*: Based on action evaluation feedback, adjustments can be made to any new behaviors, norms, habits, or organizational practices. After organizational members have had a chance to practice these new behaviors, the process of evaluation and adjustment should follow, resulting in a continual cycle of individual and organizational learning and development.

Once action plans have been set during the initial phase, we recommend that leaders and managers work their way through the other three phases of action taking, evaluating, and adjusting at least once a quarter. This means that within the first 90 days following the release of survey results, action campaigns should be in full effect, and early signs of change should be noticeable. In the quarters that follow, action efforts should become increasingly effective, and signs of positive organizational change should become increasingly evident.

By committing to a quarterly review process, organizations can ensure that actioning efforts remain salient and lead to real change. Evaluations can be conducted formally through a brief follow-up survey, or informally during team meetings or individual discussions. Some organizations also choose to conduct a brief pulse surveys—usually 6 months following the last survey—to determine whether employee attitudes have improved.

Once yardsticks have been established, they can be turned into a checklist of activities to complete, along with a recommended timeline (see example below). This provides a clear

Week	First quarter activities for leaders and managers to complete
1	• Attend training
	• Review and analyze survey results
	• Prepare for feedback meeting
2	• Conduct feedback meeting with employees; solicit their input
3 and 4	• Review survey results, employee feedback, and action priorities with boss
	• Agree to areas for actioning
5	• Develop and record SMART action plans
	• Communicate action plans to employees

(continued)

(continued)

Week	First quarter activities for leaders and managers to complete
6–10	• Initiate action plans
	• Define new behaviors, norms, practices (what should we start, stop, and continuing doing)
	• Engage in deliberate practice of new behaviors, norms, and practices
11	• Conduct first evaluation
	–Do actions seem to be leading to positive change?
	–Are we learning anything new?
	–Have behaviors changed?
	–Are we producing the results (e.g., improved performance, efficiency, customer service) we expected?
	• Review feedback, make any necessary adjustments
12	• Provide status update to boss
	• Determine if additional support is needed to drive actions
	• Continue cycle of action taking, evaluation, and adjustment

roadmap for leaders and managers to follow and ensures a coordinated approach actioning.

28.10 Step 7: Create a Support System for Those Who Need Extra Help

For many leaders and managers, accountability planning, post-survey training, and well-positioned yardsticks will provide enough incentive, structure, and guidance to promote action taking. But others—particularly those facing challenging situations—may need more help from their boss or HR. With this in mind, it is important to develop a survey action support system to provide additional help to those who need it.

The first step in putting together a support system is to identify which leaders might be in need of additional help. Common candidates include:

- Leaders and managers with unfavorable survey results, particularly on items and dimensions that assess topics like leadership effectiveness, trust, and fairness;
- New leaders and managers who have never gone through the actioning process before, or who are taking over a team they did not manage when the survey was conducted;
- Leaders and managers who are trying to implement complex organizational change.

There are two good ways to determine who needs additional support. The first is to ask. At the end of training, we recommend asking leaders and managers if they feel clear, confident, and ready to take action based on their data. The second is to assess progress. Leaders and managers who are falling behind schedule—postponing feedback meetings, not submitting action plans on time, skipping evaluation updates with their boss—could be in need of additional support.

Once candidates for additional support are identified, the next step is to determine what is getting in the way of action. Simple questions, based on motivational interviewing techniques (Matulich 2010), can be used to identify barriers to action. For example:

- How important is it to focus on actioning your data? Where would you rank this in comparison to your other priorities?
- How confident do you feel about your ability to take action based on your results? On a scale of 0–10, how optimistic do you feel about creating change?

While leaders and managers may present a wide range of issues and concerns, we have found that the root causes of inaction often boil down to four fundamental dilemmas:

- "*I don't know what to do*." After reviewing results and talking to direct reports, some managers and leaders tend to feel overwhelmed by their data, their issues, and the number of actioning opportunities they feel they should pursue. Others may feel confused about the right step to talk. These managers and leaders may appear anxious, overwhelmed, and worried about taking the wrong step or pursuing the wrong opportunity.
- "*I don't know how to do this*." Some leaders and managers may lack confidence in their ability to make change happen. While they may have a clear sense of what they want to do, they may not understand the best way to reach their actioning goal. For example, a manager may know that she wants to create an environment that encourages innovation, creativity, and risk-taking, but she may not know how to bring this change about.
- "*My results are negative and my team seems upset with a number of things, including me. I don't know how to talk with them about these results*." Receiving unfavorable survey results can be embarrassing, disheartening, and even debilitating for leaders and managers, particularly if they are highly conscientious, sensitive, or insecure about their leadership abilities. When faced with poor results, some leaders and managers may not know the best way to approach their employees, find solutions, and move forward.
- "*I have more important things to do*." Some leaders and managers are unaware of the motivational and business imperatives for action. Others overestimate the amount of time and effort required to create positive change. As a result, they avoid action planning, assuming it will distract them from what they consider to be their real priorities.

Once root causes have been determined, appropriate support strategies can be developed. Various types of interventions, usually conducted by HR, can be helpful.

- *Individual Coaching* can be effective when leaders and managers need help prioritizing opportunities, identifying solutions, and developing action plans. They can also serve as action accountability partners, helping leaders and managers stick to the plans they develop. When coaching for action, we recommend utilizing motivational interviewing techniques (Matulich 2010) and brief coaching practices (Berg and Szabo 2005) because they focus on empowering leaders and managers to find their own solutions to problems (rather than becoming dependent on HR for making actioning happen).
- *Mentoring* is an excellent way to build confidence in leaders and managers who are not sure how to make a specific type of action plan take effect. Using survey results, highly effective leaders can be paired up with those in need of help. For example, a manager who knows how to create a climate for creativity can serve as a mentor for another manager who is trying to create this same environment. Through this process, best practices can be shared and leadership networks can be built within organizations.
- *Meeting Facilitation* may be necessary if a manager is concerned about negative results and uncomfortable conducting a feedback meeting with employees. In such instances, HR should meet with the manager to determine the best course of action. In particularly negative situations, it may be best for a HR representative to hold an initial meeting just with employees (without the manager present). This can ensure that employees feel free to be candid and express their concerns. Once employees have been given an opportunity to voice their concerns, the manager can be brought back into the discussion. Ultimately, HR needs to ensure that managers and employees are working together to discuss issues and find solutions.
- *Education* may be in order for leaders and managers who think they have more important things to do than focus on survey actioning. Ensuring they know the motivational and business imperative for action is a good place to start. We have found that peer pressure can be also effective: when leaders and managers realize that their colleagues and supervisors are serious about taking

action, they are often more likely to follow suit.

- *Actioning Best Practices and First Step Recommendations* can help leaders and managers get their action planning started. Beckhard and Harris (1987) have argued that in order to overcome resistance to change, it is important to define a practical first step that leaders and managers can take toward their new goal. Providing best practices and first step recommendations can help hesitant leaders and managers move from analysis-paralysis to action.

With well designed action support systems in place, even the most dire situations can be turned around. For example, one of our clients used their survey results to identify the 100 least engaged units and teams within the company—those with the lowest employee engagement scores. Through a specially designed program, the managers of these units received additional coaching, mentoring, and support in the months following the survey. A year later, when the annual survey was administered again, engagement scores had improved an average of 14% points in these units, versus just one point in all other units.

28.11 Step 8: Utilize Tools and Technology to Accelerate Actioning

Technology has transformed the employee survey process (Kraut 2006). Over the past 20 years, a number of organizations have used the Web to prepare for, conduct, and analyze employee surveys (Barbera and Young 2006). Technology has also had a positive impact on the actioning phase, providing a number of advantages:

- *Speed and accuracy*: Technology enables survey results to be analyzed and presented quickly and accurately, often in a matter of days. Such quick turnaround time helps senior leaders and managers get a "real time" gauge of employee attitudes. From the perspective of employees, they see that their feedback is attended to in a timely manner. This sense of immediacy increases the relevance of feedback and the likelihood that organizational stakeholders will commit to actioning.
- *Creating a flexible structure for the actioning process*: Having a clear structure can boost managers' confidence levels, which in turn makes them more likely to engage in actioning. Technology-based tools can help create a scalable, organization-wide, structure for the action planning process, creating templates and checklists that make the process clear and consistent. Technology also makes it possible to easily customize such templates.
- *Creating opportunities for learning and knowledge sharing*: Technology creates opportunities for virtual and distant learning opportunities. Within the context of survey actioning, managers can benefit from both online and offline learning modules about action planning (e.g., "how to" guides). Technology also creates virtual venues for sharing knowledge and best practices and presents unique opportunities for managers to connect with and support each other.

Building on these advantages, we have developed several technology-based tools that are designed to support managers and HR professionals throughout the actioning process.

- *Web-based Reports*: Survey results are presented along with relevant benchmark data in a user-friendly format. Using software, we customize reports based on a leader or manager's sphere of influence, providing senior leaders with a high-level overview of organization-wide results, while providing front-line supervisors with a much detailed view of their local work environment. Leaders and managers can also explore their data and search for various demographic, functional, and unit based trends using an online analysis portal. These reports, and the system that supports it, provide the basis for the actioning process, providing insight into employee attitudes and possible areas for action.
- *Action Tracker*^{TM}: This web-based system offers leaders and managers the opportunity to develop and record their action plans and to connect to our Best Practices Library. Using this tool, managers can develop SMART goals, specify how they intend to reach them, and

track progress toward the goals. Because all actions are organized within an online repository, managers can explore how others are addressing similar issues/survey categories and identify actions that have proven to be effective. Survey administrators and HR can identify common areas of opportunity within and across business units and support managers with relevant training opportunities or coaching.

- *Best Practices Library*: Managers are often unclear as to why and how survey topics should be addressed. To this end, we have developed a library of "best practices" mapped to specific survey items. These brief write-ups provide specific actions that managers can take to address employee concerns. This tool expands the way managers think about various management issues and helps them develop concrete action plans. Additionally, this library helps managers design the best course of action for their particular situation.
- *Mentor Match*: This system matches an organization's best-in-class managers with at-risk managers as indicated by survey results. Managers can also be paired based on other criteria, such as business type, geographic location, or tenure. The idea is to help managers connect with and gain learning opportunities and support from their peers.

When utilizing technology-based tools, it is important to keep a couple things in mind. First, actioning tools need to be user-friendly so that managers can easily integrate them into their everyday work routines. The tools themselves should not become burdensome. Second, tools are merely aids to facilitate managers' actioning efforts and should not impose a "cookie-cutter" approach to actioning. Tools need to be designed in ways that provide a clear and consistent template/structure while at the same time leaving ample room for managers to determine their own goals as well as determine the path by which they will accomplish them. Third, no tool or system should be a "stand-alone" offering. Tools will be most effective when they fit within the overall actioning strategy and are supported by additional resources, such as training and/or coaching.

28.12 Conclusion

At the core of this chapter is a fundamental point: if you are about to conduct an employee survey, you are in a unique position to be an agent of change in your organization. Employee surveys have the potential to improve organizational effectiveness, both from a business performance perspective and a human capital perspective. But surveys are only useful if they lead to action. Otherwise, they can cause more harm than good.

With that in mind, we encourage you to consider eight questions before administering your survey:

1. Have you created a clear and compelling *commitment for action* from the top down?
2. Have you clarified *accountability*, ensuring leaders, managers, staff heads, and employees know what their post-survey actioning roles and responsibilities are?
3. Have you developed a *training* program that will prepare leaders and managers to understand their data, conduct feedback meetings, and develop effective action plans?
4. Have you created a process for *aligning* actions across the organization?
5. What *learning practices* are you planning to utilize to ensure that your survey produces individual, team, and organizational level learning?
6. What are the *yardsticks* that you will utilize to keep leaders and managers on track during the actioning process?
7. What *support systems* do you have in place to help leaders and managers who require additional support?
8. What *tools and technology* will you use to facilitate the actioning process and help leaders and managers communicate, collaborate, share information, and learn from each other?

Survey actioning campaigns succeed when leaders, managers, and employees are ready, willing, and able to do the hard work of turning data into action. By utilizing the CATALYST model, you can prepare your organization for action and ensure your survey leads to positive organizational change.

References

Ashkenas, R., Jick, T., Ulrich, D., & Paul-Candhury, C. (1999). *The boundaryless organization field book: Practical tools for building the new organization.* San Francisco: Jossey-Bass.

Barbera, K. M., & Young, S. A. (2006). Online reporting: Real time, real impact, real opportunities. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications*. San Francisco: Jossey-Bass.

Beckhard, R., & Harris, R. (1987). *Organizational transitions: Managing complex change*. Reading, MA: Addison-Wesley Publishing Company.

Berg, I., & Szabo, P. (2005). *Brief coaching for lasting solutions*. New York: W.W. Norton.

Burke, W. W. (1980). System theory, gestalt therapy, and organization development. In T. G. Cummings (Ed.), *Systems theory for organizational development* (pp. 209–229). Sussex: Wiley.

Burke, W. W. (1982). *Organization development: Principles and practices*. New York: Little, Brown & Company.

Burke, W. W. (2002). *Organizational change: Theory and practice*. Thousand Oaks, CA: Sage Publications.

Burke, W. W. (2006). Organizational surveys as leverage for organization development and change. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications*. San Francisco: Jossey-Bass.

Burke, R. J., & Cooper, C. L. (2008). *Building more effective organizations*. Cambridge: Cambridge University Press.

Burke, W. W., Coruzzi, C. A., & Church, A. H. (1996). The organizational survey as an intervention for change. In A. I. Kraut (Ed.), *Organizational surveys: Tools for assessment and change* (pp. 41–66). San Francisco, CA: Jossey-Bass.

Church, A. H., & Oliver, D. H. (2006). The importance of taking action, not just sharing survey feedback. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications*. San Francisco: Jossey-Bass.

Coghlan, D., & Brannick, T. (2005). *Doing action research in your own organization*. London: Sage.

Cooperrider, D. L., & Whitney, D. (2005). *Appreciative inquiry: A positive revolution in change*. San Francisco: Berrett-Koehler Publishers.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Fitz-enz, J. (2000). *The ROI of human capital: Measuring the economic value of employee performance*. New York: American Management Association.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *The Academy of Management Journal, 31*(1), 9–41.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*(2), 268–279.

Heath, C., & Heath, D. (2010). *Switch: How to change things when change is hard.* New York: Broadway Books.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin,* 127(3), 376–407.

Kanter, R. M. (2004). *Confidence: How winning streaks and losing streaks begin and end.* New York: Crown Publishing.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard—Measures that drive performance. *Harvard Business Review, 70*(1), 71–79.

Kaplan, R. S., & Norton, D. P. (2006). *Alignment: Using the balanced scorecard to create corporate synergies*. Boston: Harvard Business Press.

Kinicki, A. J. (2009). *Performance management systems*. Superstition Mt., AZ: Kinicki & Associates.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of American Society for Training and Development, 13*, 11–12.

Klein, K. J., & Kozlowlski, S. W. (2000). *Multilevel theory, research, and methods in organizations*. San Francisco: Jossey-Bass.

Kotter, J. P. (1996). *Leading change*. Boston: Harvard Business School Press.

Kraut, A. I. (1996). An overview of organizational surveys. In A.I. Kraut (Ed.), Organizational Surveys: Tools for Assessment and Change (pp. 1–14). San Francisco: Jossey-Bass.

Kraut, A. I. (2006). Moving the needle: Getting action after a survey. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications*. San Francisco: Jossey-Bass.

Kreitner, R., & Kinicki, A. J. (2010). *Organizational behavior* (9th ed.). New York: McGraw Hill.

Labovitz, G., & Rosansky, V. (1996). *The power of alignment*. New York: John Wiley & Sons.

Lawler, E, I. I. I., & Worley, C. (2006). *Built to change: How to achieve sustained organizational effectiveness*. San Francisco: Jossey Bass.

Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; Social equilibria and social change. *Human Relations, 1*, 5–41.

Likert, R. (1967). *The human organization: Its management and value*. New York: McGraw Hill.

Mann, F. C. (1957). Studying and creating change: A means to understanding social organization. In C. M. Arensberg et al. (Eds.), *Research in industrial human relations* (pp. 146–157). Champaign, IL: Industrial Relations Research Association, Publication No. 17.

Marsick, B. J., & Watkins, K. E. (1999). *Facilitating learning organizations*. London: Gower Publishing.

Marsick, V., O'Neil, J., & Watkins, K. (2002). Action learning. In J. Waclawski & A. Church (Eds.), *Organization development: A data-driven approach to organizational change*. San Francisco: Jossey-Bass.

Matulich, B. (2010). *How to do motivational interviewing: A guidebook for beginners.* Amazon Digital Services.

McCall, M. W. (2010). Recasting leadership development. *Industrial and Organizational Psychology, 3*(1), 3–19.

Nadler, D. A. (1996). Setting expectations and reporting results: Conversations with top management. In A. I. Kraut (Ed.), *Organizational surveys: Tools for assessment and change*. San Francisco: Jossey-Bass.

Nadler, D. A. (1977). *Feedback and organization development: Using data-based methods*. London: Addison-Wesley.

Norman, D. A. (1994). *Things that make us smart: Defending human attributes in the age of the machine*. New York: Perseus Books.

O'Toole, J. (1995). *Leading change*. San Francisco: Jossey-Bass.

Pascale, R., Sternin, J., & Sternin, M. (2010). *The Power of positive deviance: How unlikely innovators solve the world's toughest problems*. Boston: Harvard Business Press.

Pfeffer, J. (1994). *Competitive advantage through people: Unleashing the power of the work force*. Boston: Harvard Business School Press.

Pfeffer, J., & Sutton, R. I. (2000). *The knowing-doing gap: How smart companies turn knowledge into action*. Boston: Harvard Business School Press.

Pfeffer, J. (1998). *The human equation: Building profits by putting people first*. Boston: Harvard Business School Press.

Schiemann, W. A., & Morgan, B. S. (2006). Strategic surveys: Linking people to business strategy. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications*. San Francisco: Jossey-Bass.

Sirota, D., Mischkind, L. A., & Meltzer, M. I. (2005). *The enthusiastic employee: How companies profit by giving workers what they want*. Indianapolis, IN: Wharton School Publishing.

Sullivan, G. R., & Harper, M. V. (1996). *Hope is not a method: What business leaders can learn from America's army*. New York: Broadway Books.

Ulrich, D., & Brockbank, W. (2005). *The HR value proposition*. Boston: Harvard Business Press.

Ulrich, D., & Smallwood, N. (2003). *Why the bottom line isn't!: How to build value through people and organization*. New York: Wiley.

Vitales, M. S. (1953). *Motivation and morale in industry*. New York: Norton.

Von Bertalanffy, L. (1950). An outline of general systems theory. *British Journal of the Philosophy of Science, 1*, 134–165.

Index

A
Absolute ranking, 100, 102
ACASI, 384
Action planning, 503
Albanian criminality, 413
Analysis, 463, 465, 466, 469, 471
Anchoring vignettes, 477, 482, 489, 494, 496
Anonymity, 27
Attrition bias, 73, 74

B
Behavioral consistency, 227
Belmont report, 24, 25, 29
Beneficence, 23–25, 29
Bias, 39, 40, 42–45
Biased questions, 94
Biased sample, 54, 60
Bogus pipeline
Budget, 463, 466, 469

C
45 Code of Federal Regulations 46, 24
Central limit theorem, 54
Close-ended/multiple choice questions, 94
Clustered sampling, 258
Cognitive interviewing, 227, 242
Comparability, 477–488, 492–496
Comparative surveys, 477, 479, 480, 482
Complex and simple concept questions, 64
Compliance with legitimate authority, 327
Concept, 93, 94, 102, 103
Confidence intervals, 69, 70
Confidentiality, 327, 340, 27, 28, 31–33
Construct, 92, 93, 104
Constructionist-collaborative paradigm, 227
Contact rate, 195, 197
Convenience sampling, 66
Conversational analysis, 227, 243
Conversational interview format, 227
Cost-benefit analysis, 135
Coverage, 37–39, 42, 43
Cross-cultural, 477, 481
Cross-cultural survey, 250
Cross-group, 494

D
Decision-making, 75, 76, 83, 85–87
Decision-stage model of response, 179
Delinquency, 118
Demographic characteristics, 195, 201
Differential item functioning (DIF), 477, 481, 489, 494, 496
Differential measurement error, 482, 492
Differential nonresponse, 479, 487
Difficult populations, 400
Digital signatures, 327
Domain names, 327
Double negative questions, 96
Double-barreled questions, 94
Dual frame, 286, 290, 292

E
Element, 53–56, 58–60, 66
E-mail, 313–324, 345–347, 350–358
E-mail headers, 327
E-mail methodology, 313
E-mail questionnaire, 313, 314
E-mail survey administration, 313
Email Surveys, 73
E-mail surveys time framework, 313
Equal probability of selection method, 55, 66
Equal probability of selection method or EPSEM, 55
Equivalence, 483, 489, 490, 492, 494
Ethics, 23, 33
Experiment, 463, 464

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2, © Springer Science+Business Media New York 2012

F
Face-to-face surveys, 73
Factorial, 463, 464, 468, 469, 471
Falsification, 31
Fear of crime, 443–452, 454–457
Financial costs, 365
Framing, 75, 76, 80–84, 86–89

G
Grounded theories, 109

H
HTML, 327, 335, 338
Human subjects, 23–26, 29, 31

I
Incentive amount, 209
Incentives, 29, 31
Increasing response, 179
Inequivalence, 491, 494, 495
Informants, 109, 110
Informed consent, 23, 24, 28, 29, 31
Institutional review board, 109, 111
Interview schedule, 109, 115
Interview techniques, 407
Interviewer training, 413
Interviews, 109
Invariance, 477, 481, 490–492, 494, 496
Invariance testing, 477, 481, 490, 494, 496
Item and topic sensitivity
Item bias, 481, 482, 490
Item nonresponse, 195

J
Justice, 23–25, 29

L
Leading and loaded questions, 43
Legitimacy, 327
Leverage saliency theory, 227, 238
Likert scale, 101
List serves, 327
Literature review, 95, 102
Lotteries versus guaranteed awards, 327

M
Mail survey, 297, 298, 303, 307, 310
Mail Surveys, 362
Managers, 504, 514
Material incentives versus nonmaterial incentives, 327
Maximum bias, 485, 486
Mean squarre error, 277
Measurement, 37–39, 45–47
Measurement error, 479, 481, 488, 492–494
Measuring response style, 160
Method bias, 481
Methodology, 114
Mind-blockage, 98
Mixed mode, 277
Mixed mode contacts, 327
Mixed mode surveys, 369
Mobile phone surveys, 277
Mobile-only, 277, 289
Mode of interview, 384
Modes of data collection, 179, 181, 186
Monetary incentives, 209–211, 213, 214, 216, 220
Motivation for survey response, 179
Multiple choice, 327, 109, 115, 116
Multiple choice questions, 109, 115, 116
Multiple contacts, 221
multitrait multimethod (MTMM), 149

N
Network sampling, 68
Non-probability, 73
Non-probability panel, 11
Nonprobability sampling, 421
Nonrandom sampling, 66, 67
Nonresponse, 195, 277, 280, 37–39, 43–45
Nonresponse bias, 195
Non-response bias assessment, 149
Nonresponse conversion, 195
Nonresponse error, 5
Non-sampling errors, 4, 5
Nuremberg code, 3

O
Online behavior, 313, 314
Online surveys, 313, 317, 323, 345, 346, 348, 351, 352, 358, 359
Open-ended questions, 88, 89, 109, 116
Optimisation, 277
Order and response scale effects, 172
Organizational surveys, 220
Organized crime, 23
Outcome rate, 195, 196

P
Panel recruitment, 346, 349, 350, 354–359
Panel surveys, 345, 346
Participation in surveys, 179
Phishing, 327
Phone coverage, 277
Phone surveys, 277
Pilot testing, 227
Population, 53–56, 58–70, 74
Prenotifications, 327
Prepaid incentives, 210–212, 221
Prepaid incentives versus postpaid incentives, 327

Privacy
Probability, 53, 54, 60, 61, 63–65, 68, 69, 71–73
Probes, 227, 239
Programming, 463, 466, 468–470
Progress indicators, 327, 340
Purposive sampling, 67

Q
Quality assessment on obtained and registered responses, 149
Question coding system, 483
Question order, 443, 450–457
Question wording, 443, 447–452, 456
Questionnaire design, 297
Quota sampling, 67, 74

R
Random digit dialing (RDD), 369
Random selection, 54, 62, 68, 69
Randomized response technique, 395
Ranking, 92, 100–103
Reciprocity, 327, 328
Referral sampling, 68
Relative ranking, 92, 101, 102
Reliability, 479, 482, 492, 493
Reminders, 327, 335, 336
Representative sample, 53, 60, 69
Respect for persons, 24
Response bias, 97
Response cooperation, 195
Response enhancement measures, 149, 152
Response errors, 5, 418, 437
Response function, 482, 489, 490, 494
Response quality, 179
Response quantity
Response rate, 195, 196, 204–206, 297, 300, 302, 303, 306–310
Response rates, 209–215, 217–221
R-indicator, 477, 485–487, 496

S
Salutations, 327, 333
Sample, 53–56, 58–74
Sample size, 53–56, 58, 59, 61, 62, 65, 68–72
Sample weights, 64, 65
Sampling, 37–43, 47, 48
Sampling bias, 364
Sampling distribution, 54, 58
Sampling error, 54, 55, 58, 62, 70
Sampling errors, 4–6
Sampling frame, 54, 56, 58–63, 68, 72
Sampling methods, 53–55, 60, 62, 63, 66, 68, 70, 74
SAS, 64
Scrolling versus paging, 327
Self-reported, 313, 314, 316
Self-report surveys, 417
Semi-standardized interview/semi-structured interview, 109
Sensitive topics, 313, 315, 323
Sensitive questions, 419
Sexual behavior, 378
Simple random sampling, 55, 58–62, 63, 67, 70, 72
Snowball sampling, 68
Social desirability, 227, 5, 109, 111, 112
Social exchange theory, 227, 238
Split ballot approach, 149, 160, 167
SPSS, 64
SQP, 483
Standardized interview format, 227
STATA, 64, 65
Stimulus-response/information-retrieval paradigm, 227
Stratification variable, 61, 62
Stratified sampling, 59, 61–63, 67, 69, 70
Structured questions, 109
Substance use/abuse, 379
Substantive consistency, 246
Survey, 55, 63, 65, 72, 74
Survey cooperation, 195
Survey costs, 277, 280, 284
Survey crafting, 463
Survey design, 443, 444, 446, 447, 455–457
Survey error, 277
Survey introduction, 351
Survey Invitations, 327
Survey length, 327, 340
Survey measurement, 75
Survey mode, 91
Survey participation, 209–211, 220
Survey Response, 179, 180, 277
Survey sampling, 277, 313
Survey utilization
Surveys, 195–198, 200–206, 109, 111, 113
Systematic sampling, 58–60, 63, 69, 70

T
Telephone surveys, 363
Theories of survey response, 179
Total design method (Dillman 1978), 362
Total survey error (TSE), 37, 48, 109
Translation, 478, 481–484, 488
Transmission Time, 327
Transparency, 33
Tuskegee Syphilis study, 23
Type I and Type II errors

U
Unmatched count technique
Unreliability, 479, 492
Unstructured questions, 109, 116–118
URL, 327, 335

V
Validity
Validity and reliability, 297
Variables, 463–466, 471
Variance, 39–45, 48
Victimization, 443–446, 450–457
Vignette, 463, 464, 466–471
Visibility, 5

W
Weak comparability, 477, 482, 485–487, 494
Web surveys, 346, 348
Web-based surveys, 367
Wording
wording, order and response scale effects, 149