



Questionnaire Development, Evaluation, and Testing Methods (QDET) Conference

Good Discussions, Debates, and Dining in Charleston

La conférence -- Développement, évaluation et méthodes de test pour les questionnaires (QDET) -- Bonnes discussions, débats et dîners à Charleston

Jennifer M. Rothgeb



Electronic version

URL: <http://journals.openedition.org/bms/1425>

ISSN: 2070-2779

Publisher

Association internationale de méthodologie sociologique

Printed version

Date of publication: 1 January 2003

Number of pages: 35-61

ISSN: 0759-1063

Electronic reference

Jennifer M. Rothgeb, « Questionnaire Development, Evaluation, and Testing Methods (QDET) Conference », *Bulletin de méthodologie sociologique* [Online], 77 | 2003, Online since 01 January 2006, connection on 10 December 2020. URL : <http://journals.openedition.org/bms/1425>

This text was automatically generated on 10 December 2020.

© BMS

Questionnaire Development, Evaluation, and Testing Methods (QDET) Conference

Good Discussions, Debates, and Dining in Charleston

La conférence -- Développement, évaluation et méthodes de test pour les questionnaires (QDET) -- Bonnes discussions, débats et dîners à Charleston

Jennifer M. Rothgeb

- 1 The International Conference on Questionnaire Development, Evaluation, and Testing Methods (QDET) was held November 14-17, 2002, in Charleston, South Carolina at the Embassy Suites Convention Center Hotel. The 338 conference attendees represented 23 countries and six continents. There were 32 sessions with 22 invited papers and 54 contributed papers. In addition, there were 15 poster presentations. The four short courses (listed below) were well attended and well received by participants.
- 2 Norman Bradburn was the keynote speaker and his talk focused on "The Future of Questionnaire Research." He emphasized the need for the field to keep abreast of how to maximize the utility of technological advances, the increasing challenges the field faces with multicultural, multilanguage issues, and the benefits which may be gained through the incorporation of more sociolinguistics in survey research.
- 3 One of QDET's first sessions provided a lively, provocative debate about the quantitative versus qualitative aspects of pretesting. The resulting floor discussion was still going on weeks after the conference! Other sessions of noted interest for participants were those related to different methods of cognitive interviewing, split-sample experiments, testing surveys with children, multilanguage and multicultural issues, and sessions related to web surveys and usability research.
- 4 The sessions certainly served their intended purpose -- to stimulate discussion and motivate researchers to develop alternative research ideas. From the presentations and resulting discussions, it is obvious that there is a great deal more to be learned about questionnaire evaluation and testing. Conference participants had opportunities to share

ideas each night, during the hotel-sponsored evening receptions, before heading down to historic Charleston for low-country dining.

- 5 QDET was fortunate to have funds available to provide conference support for several graduate students and persons from countries typically underrepresented at conferences such as QDET. SRMS funds supported 12 Conference Fellowships. Fellowships were granted to persons from South Africa, Kenya, the Philippines, Slovenia, Italy and Korea, as well as to foreign graduate students in the U.S. A grant received from the National Science Foundation also allowed QDET to offer an additional ten Conference Fellowships to U.S. citizens, most of whom were graduate students.
- 6 QDET invited papers will be included in a monograph which should be published in December 2003. The *Journal of Official Statistics* (JOS) has offered to produce a volume consisting of selected contributed papers. Tentatively, it is planned that the selected QDET papers will be in the Winter 2003 volume. Contributed papers are posted on the conference web site (www.jpsm.umd.edu/qdet) and will remain there for three months. Conference photos are also available on the conference web site.
- 7 From the completed conference evaluation forms, it is apparent that QDET attendees are interested in future conferences focusing on topics such as questionnaire design, web surveys, and multicultural issues related to survey research. SRMS eagerly awaits conference proposals on such topics !

QDE Final Conference Program

Short Courses

- 8 I. Methods for Questionnaire Appraisal and Expert Review ; Barbara Forsyth <barbaraforsyth@westat.com>, Westat, Inc., USA ; Gordon Willis <willisg@mail.nih.gov>, National Cancer Institute, National Institutes of Health, USA
- 9 II. Cognitive Interviewing ; Eleanor Gerber, U.S. Census Bureau
- 10 III. Question Testing for Establishment Surveys ; Kristin Stettler, U.S. Census Bureau ; Fran Featherston, National Science Foundation, NSF
- 11 IV. Behavior Coding : Tool for Questionnaire Evaluation ; Nancy Mathiowetz, University of Maryland / University of Michigan, USA

Sessions

- 12 1. Invited Session : Design Considerations for Pretesting ; Chair, Judy Lessler, RTI International, USA ; Design Considerations for Pretesting, Roger Tourangeau <rtourangeau@survey.umd.edu>, JPSM, University of Maryland, USA ; Discussant, Jim Chromy, RTI International, USA ; Discussant, Eleanor Gerber, U.S. Census Bureau
- 13 2. Contributed Session : Cognitive Interviewing for Establishment Surveys ; Chair, Trish Gallagher <csr@umb.edu>, Center for Survey Research, University of Massachusetts, USA ; Cognitive Testing of Mail Surveys at Statistics Sweden, Gunilla Davidsson <gunilla.davidsson@scb.se>, Statistics Sweden ; Establishments as Respondents : Is Conventional Cognitive Interviewing Enough ?, Robert Rutchik <robert.rutchik@eia.doe.gov> and Stanley Freedman, U.S. Energy Information Administration ; A Comparison of Appraisal and Cognitive Interview Methods for Testing

- Organizational Survey Questionnaires, Barbara Forsyth <forsybt1@westat.com>, Westat, Inc., USA, Elisa Weiss, New York Academy of Medicine, USA, Rebecca Miller Anderson, Mount Sinai School of Medicine ; Discussant, Don Dillman <dillman@wsu.edu>, Washington State University, USA
- 14 3. Contributed Session : Usability - Noninternet ; Chair, Janice Ballou, Mathematica Policy Research, USA ; Approaches for Incorporating User-Centered Design into CAI Development, Bill Mockovak <mockovak_w@bls.gov> and Jean Fox, U.S. Bureau of Labor Statistics ; Pretesting an Interactive Voice Response Survey, Tracey Haggerty Heller <traceyhaggertyheller@westat.com>, Sid Schneider, and David Cantor, Westat, USA ; Analyzing Audit Trails in the National Survey on Drug Use and Health: Means for Maintaining and Improving Data Quality, Michael Penne <penne@rti.org>, Jeanne Snodgrass and Peggy Barker <ppb@rti.org>, RTI International ; Discussant, Mick Couper, University of Michigan, USA
- 15 4. Invited Session : Cognitive Interviewing - II ; Chair, Nancy Bates, U.S. Census Bureau ; Aspects of Data Quality in Cognitive interviews. The Case of Verbal Reports, Fred Conrad <fconrad@isr.umich.edu>, University of Michigan, USA, Johnny Blair, Abt Associates, USA ; Cognitive Interviews : Do Different Methods Produce Different Results ?, Terry DeMaio <theresa.j.demaio@census.gov> and Ashley Landreth, U.S. Census Bureau ; Discussant, David Cantor, Westat, USA
- 16 5. Contributed Session : Internet - I ; Chair, Susan Ciochetto, U.S. Census Bureau ; Identifying and Reducing the Response Burden in Internet Business Surveys, Gustav Haraldsen, Dag Roll-Hansen and Tore Notnes, Statistics Norway ; Evaluation of Users' Experience of the Census 2000 Internet Form, Elizabeth Murphy <elizabeth.d.murphy@census.gov> and Courtney Stapleton, U.S. Census Bureau ; Usability Testing of Web Data Collection Instruments, Elizabeth Wiebe <efw@rti.org>, Lisa Thalji, and Robert Wagers, RTI International, USA ; Discussant, Bill Mockovak, U.S. Bureau of Labor Statistics
- 17 6. Contributed Session : Mathematical Methods for Questionnaire Evaluation ; Chair, Carolee Bush, U.S. Census Bureau ; Survey Design and Validation Using the Rasch Model, Kathy Green <kgreen@du.edu> and Cathy Frantom, University of Denver, USA ; Principle Component Factor Analysis: An Analytic Strategy to Increase Content Validity of Questionnaire Factors, Manouchehr Afshinnia <afshinnia@hotmail.com> and Farsad Afshinnia, Isfahan University, Iran ; Applications of Item Response Theory (IRT) in Questionnaire Evaluation, Bryce Reeve <reeveb@mail.nih.gov> and Louise Masse, National Cancer Institute, National Institute of Health, USA ; Discussant, Joop Hox <j.hox@fss.uu.nl>, University of Utrecht, Netherlands
- 18 Keynote Speaker : The Future of Questionnaire Research, Norman Bradburn, National Science Foundation, USA
- 19 7. Invited Session : Cognitive Interviewing - II ; Chair, Barbara O'Hare, Arbitron Company, USA ; Cognitive Interviewing Revisited : A Useful Technique, in Theory ?, Gordon Willis, National Cancer Institute, National Institutes of Health, USA ; The Dynamics of Cognitive Interviewing, Paul Beatty <pbeatty@cdc.gov>, National Center for Health Statistics, USA ; Discussant, Roger Tourangeau, University of Michigan and University of Maryland, USA
- 20 8. Contributed Session : Pretesting and Quality Assessment ; Chair, Carol Cosenza <carol.cosenza@umb.edu>, University of Massachusetts, USA ; Iterative, Multiple-Method Questionnaire Evaluation Research : A Case Study, James Esposito <esposito_j@bls.gov>,

- U.S. Bureau of Labor Statistics ; Sun Exposure Recall : Instrument Development and Evaluation, Diane Nishri <diane.nishri@cancercare.on.ca>, Beth Theis, Jennifer Frood, Division of Preventive Oncology, Cancer Care, Canada, Fred Ashbury, PICEPS Consultants Inc., Canada, David Northrup, York University, Canada, Loraine Marrett, University of Toronto, Canada ; Using Reinterview Methods to Design and Evaluate Survey Questions, Jeremy Morton <jmorton@rti.org>, RTI International, Paul Mullin, IBM Business Consulting Services, Paul Biemer, RTI International ; Discussant, Judy Lessler
- 21 9. Contributed Session : Pretesting Challenges among Various Ethnic and Cultural Groups Within the U.S. ; Chair, Kristen Miller <ksmiller@cdc.gov>, U.S. National Center for Health Statistics ; Mexican Immigrants and the Use of Cognitive Methods in Questionnaire Development, Robert Agans <agans@unc.edu> and Natalia Deeb-Sossa, University of North Carolina at Chapel Hill, USA ; Cross Site Tool Development Focusing on Co-Occurring Populations, Pamela Clark <clark_pamela@yahoo.com>, Acrobat Research, USA, Robert Walker, University of Kentucky, USA, Tom Doub, Dual Diagnosis Management, USA ; Discussant, Stanley Presser, University of Maryland, USA
- 22 10. Invited Session : Using Information from Respondents to Improve Surveys ; Chair, Dawn Nelson, U.S. Census Bureau, Response Latencies and Perceived Question Difficulty as Indicators for Response Error, Stasja Draisma <ar.draisma@scw.vu.nl> and Wil Dijkstra <w.dijkstra@scw.vu.nl>, Vrije University, Amsterdam ; Using Vignettes and Respondent Debriefings for Questionnaire Design and Evaluation, Elizabeth Martin <emartin@census.gov>, U.S. Census Bureau ; Discussant, Bob Belli <bbelli2@unl.edu>, University of Nebraska, USA
- 23 11. Invited Session : Testing Computerized Instruments ; Chair, Birgit Henningsson <birgit.henningsson@scb.se>, Statistics Sweden, Sweden ; Usability Testing as a Means of Evaluating Computer-Assisted Survey Instruments, Sue Ellen Hansen <sehansen@umich.edu> and Mick Couper, University of Michigan, USA ; Methods for Testing and Evaluating CATI Questionnaires, John Tarnai <tarnai@wsu.edu> and Danna Moore, Washington State University, USA ; Discussant, Brad Edwards, Westat, USA
- 24 12. Contributed Session : Novel Approaches to Questionnaire Development and Evaluation ; Chair, Johnny Blair, Abt Associate, USA ; Evaluation of Respondent and Interviewer Debriefing Techniques on Questionnaire Development Methods for Health Provider-Based Surveys, Catharine Burt <cburt@cdc.gov> and Susan Schappert, U.S. National Center for Health Statistics ; Using Vignettes in Cognitive Research on Establishment Surveys, Rebecca Morrison <rebecca.l.morrison@census.gov>, Kristin Stettler, and Amy Anderson, U.S. Census Bureau ; Strategies for Subject Matter Expert Review in Questionnaire Design, Carl Ramirez <ramirezcgao@gao.gov>, U.S. General Accounting Office ; Discussant, Jack Fowler, University of Massachusetts, Center for Survey Research, USA
- 25 13. Contributed Session : Design and Evaluation of Questionnaires ; Chair, Theresa Leslie ; Complex Questionnaire Procedures for Sensitive Topics : Developing Best Practice Procedures in a Realistic Survey Setting, Gerty Lensvelt-Mulders, Joop Hox and Hennie Boeijs, Utrecht University, Netherlands ; Contextuality of Survey Responses as a Challenge to the Development of Questionnaire Testing Methods, Anja Ahola <anja.ahola@stat.fi> and Marjaana Lehtinen, Statistics Finland ; Discussant, Jean Martin, Office for National Statistics, UK

- 26 14. Invited Session : Beyond Cognitive Interviewing : Split-Sample Comparisons ; Chair, Mick Couper, University of Michigan, USA ; Getting Beyond Pretests and Cognitive Interviewing : The Case for More Split-ballot Pilot Studies, Jack Fowler, University of Massachusetts, Center for Survey Research, USA ; The SIPP Methods Panel Project : Using Field Experiments to Improve Instrument Design, Jeff Moore <jeffrey.c.moore@census.gov>, Joanne Pascale <joanne.pascale@census.gov>, Julia Klein Griffiths, Anna Chan and Pat Doyle, U.S. Census Bureau ; Discussant, Dan Kasprzyk, Mathematica Policy Research, USA
- 27 15. Invited Session : Questionnaire Development in Specific Domains ; Chair, Diane O'Rourke, Consultant, USA ; Development and Testing of Internet Questionnaires, Reginald Baker <reg_baker@marketstrategies.com> and Scott Crawford, MS Interactive ; Children as Respondents : Developing, Evaluating, and Testing Questionnaires for Children, Edith de Leeuw <edithl@xs4all.nl> and Natacha Borgers, Utrecht University, Netherlands, Astrid Strijbos-Smits <estics@wxs.nl>, Statistics Netherlands ; Discussant, Sandra Berry, Rand Corporation, USA
- 28 16. Contributed Session : Establishment Survey QDET Methods : Issues and Current Practices Around the World - I ; Chair, Fran Featherston, National Science Foundation, USA ; Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies ; Karen Goldenberg <goldenberg_k@bls.gov>, U.S. Bureau of Labor Statistics, Amy Anderson and Diane Willimack <diane.k.willimack@census.gov>, U.S. Census Bureau, Stanley Freedman and Robert Rutchik, U.S. Energy Information Administration, Luann Moy, U.S. General Accounting Office ; Establishment Survey Instrument Development and Testing in an Integrated Survey Environment at Statistics Canada, Frances Laffey <laffey@statcan.ca>, Statistics Canada ; Discussant, Clyde Tucker, U.S. Bureau of Labor Statistics
- 29 17. Poster Session
- A. Evaluating Questionnaire Revisions in a Survey of Immunization Providers, Katherine Ballard-LeFauve, Lee Giesbrecht and Elizabeth Anderson, Abt Associates, USA
- B. Questionnaire Development by Using Semi-Structured and Cognitive Interviews, Mary Boynton (SCB)
- C. National Foreign Language Assessment and Linguistic Diversity : The FL-NAEP Language Survey and Background Questionnaire and the Case of "Heritage" Spanish Language Students in the U.S., Katherine Richardson Bruna, Michael Fast and Nina VanDyke, American Institutes for Research, USA
- D. Reformatting a Self-Administered Questionnaire Based on Item Nonresponse, Fran Chevarley (AHRQ)
- E. Do Cognitive Interviews Improve Education Surveys, Young Chun and Kevin Carter, American Institutes for Research, USA
- F. Web Survey Comments : Does Length Impact Quality ?, Carrie Christianson DeMay, Jonathan Kurlander, Kristofer Fenlason, Data Recognition Corporation, USA
- G. Using Cognitive Follow-up Interviews to Develop the 2002 Census of Agriculture Report Form, Nancy Dickey and Zulma Riberas, U.S. National Agriculture Statistical Service
- H. Developing Measures of a Complex Theory of Brand Loyalty for Use on the Internet, Jean Durall and Melinda Smith de Borrero, Knowledge Networks, USA
- I. An Application of the Three Step Test-Interview (TSTI) : A Validation Study of the Dutch and Norwegian Versions of the Illegal Aliens Scale, Tony Hak, Kees van der Veer <vdveer@staff.scw.vu.nl> and Reidar Ommundsen

- J. The Use of Constrictive Questions : Effects and Solutions Bregie Holleman and Huub van den Bergh
- K. Quest, A Generic XML-and Metadata-Based Questionnaire Management System, Wolfgang Koller and Guenther Zettl (Statistik)
- L. Developing Questions to Measure the Victimization of Developmentally-Disabled Respondents in the National Crime Victimization Survey, Denise Lewis and Kathleen Creighton, U.S. Census Bureau
- M. Development of a Multitrait Scaling Macro : A Construct Validity Tool, Louise Hiller and Roger Holder, University of Birmingham, UK
- N. Measuring Barriers to Employment : Development and Evaluation Questionnaire, Daniel McMillin, Edwin Sasaki, Laura Hecht and Kenneth Nyberg (Csubak)
- O. A Cognitive Analysis of Passionate Love, Victor de Munck (Bestweb)
- P. Improving Mail Surveys of Establishments : Testing the Effect of Incentives on Questionnaire Completion and Data Quality, Danna Moore and John Tarnai, Washington State University, USA
- Q. Use of Focus Groups in Questionnaire Design, Melvin Prince and Mark Davies
- R. Cognitive Laboratory Experiences : On Pre-Testing Computerized Questionnaires and Data Quality, Ger Snijders <gsks@cbs.nl>, Statistics Netherlands
- S. Tradeoffs and Rating Scales : How They Compare in Measuring the Priorities of Diverse Populations, Nathaniel Stone (Communication.gc)
- T. How Good is Good ? Comparing Numerical Ratings of Response Options for Two Versions of the Self-Assessed Health Statistics Question, Barbara Foley Wilson, Barbara Altman and Karen Whitaker, U.S. National Center for Health Statistics
- U. Comparing Open-Ended and Closed-Ended Questions, Merlijn Wouters
- 30 18. Invited Session : Comparisons of Question Evaluation Methods ; Chair, Jennifer Rothgeb, U.S. Census Bureau, USA ; Does Question Pretesting Make a Difference ? An Empirical Test Using a Field Survey Experiment ?, Barbara Forsyth, Westat, Inc. USA, Jennifer Rothgeb, U.S. Census Bureau, Gordon Willis, National Cancer Institute, National Institutes of Health, USA ; Procedures for Testing Self-Administered Questionnaires : Cognitive Interview and Field Test Comparisons, Don Dillman, Washington State University, USA, Cleo Redline, National Science Foundation, USA ; Discussant, Pamela Campanelli, Survey Methods Consultant, U.K.
- 31 19. Contributed Session : Assisting the Respondent in the Recall Task : Methods and Results ; Chair, Stanley Presser, University of Maryland, USA ; Calendar Survey Methods : Association Between Verbal Behaviors and Data Quality, Robert Belli, University of Nebraska, USA ; The Time-Line : The Effects of an Experimental-Aided Recall Technique in a Real-Life Survey, Wander van der Vaart <w.vandervaat@cable.a2000.nl>, Vrije University, Amsterdam ; Relating Questionnaire Design to Survey Accuracy and Response Rate with RGI, S. James Press <jpress@ucr.ac1.ucr.edu>, University of California at Riverside, USA, Judith Tanur, State University of New York at Stony Brook, USA ; Discussant, Paul Biemer, RTI International, USA
- 32 20. Contributed Session : Establishment Survey QDET Methods : Issues and Current Practices Around the World - II ; Chair, Karen Goldenberg, U.S. Bureau of Labor Statistics ; Expert Review Followed by Interviews with Editing Staff -- Effective First Steps in the Testing Process for Business Surveys, Olwen Rowlands <olwen.rowlands@ons.gov.uk>, Jack Eldridge and Sarah Williams, Office for National Statistics, U.K. ; The Review of the French Survey on R&D in Business Enterprises, Dominique Francoz

- <dominique.francoz@education.gouv.fr>, Ministry of Youth, Education, Research and New Technologies, France ; Coherence Analysis as a Tool for Questionnaire Evaluation in Enterprise Statistics, Johan Erikson <johan.erikson@scb.se>, Statistics Sweden ; Discussant, Cynthia Clark, U.S. Census Bureau
- 33 21. Contributed Session : Special Populations ; Chair, Margaret Africa, Statistics South Africa ; What Types of Survey Items Can Elicit Valid Responses from Fourth and Eighth Grade Students ?, Mette Huberman <mhuberman@air.org> and Roger Levine ; Minimizing Item Non Response in Telephone Surveys of People with Disabilities, Karen CyBulski <kcybulski@mathematica-mpr.com>, Julie Fishtein, and Anne Ciemnecki, Mathematica Policy Research, USA ; Discussant, Edith de Leeuw, Utrecht University, Netherlands
- 34 22. Invited Session : Case Studies Using Multiple Methods for Questionnaire Evaluation ; Chair, Karen Bogen, Johns Hopkins University, USA ; Improving the Clarity of Closely Related Concepts, Nora Cate Schaeffer <schaeffe@ssc.wisc.edu> and Jennifer Dykema, University of Wisconsin-Madison, USA ; Design, Testing and Evaluation of Stated Preference Questionnaires, Michael Kaplowitz <kaplowit@msu.edu>, Frank Lupi, and John Hoehn, Michigan State University, USA ; Discussant, Rachel Casper, RTI International, USA
- 35 23. Contributed Session : Methods and Issues in Developing Multilanguage, Multicultural Questionnaires ; Chair, Elizabeth Martin, U.S. Census Bureau ; Developing Bilingual Questionnaires for Use in an Indigenous Population : Experiences from New Zealand in the Development of the 2001 Maori Language Survey, Lyn Potaka <potaka@stats.govt.nz> and Suzanne Cochrane, Statistics New Zealand ; Issues in Translating Surveys : Methods and Approaches, Sylvia Kay Fisher <fisher_s@bls.gov>, U.S. Bureau of Labor Statistics, Eleanor Gerber, U.S. Census Bureau Questionnaire Translation and Questionnaire Design, Janet A. Harkness <harkness@zuma-mannheim.de>, ZUMA, Germany, Alisú Schoua-Glusberg, Beth-Ellen Pennell, University of Michigan, USA ; Discussant, Manuel de la Puente, U.S. Census Bureau
- 36 24. Contributed Session : Cognitive Interviewing Methods - I ; Chair, Kristen Hughes, U.S. Census Bureau Not Your Grandparent's Cognitive Testing : Exploring Innovative Methods in Making Cognitive Evaluation of Questions More Efficient and Easier to Use, Carol Cosenza, University of Massachusetts-Boston, USA ; Cognitive Laboratory Methods : Current Best Practices, Gers Snijders, Statistics Netherlands ; The Three-Step Test-Interview (TSTI), Kees van der Veer, Vrije University, Amsterdam, Netherlands, Tony Hak, U.S. Census Bureau, Harrie Jansen, Addiction Research Institute, Netherlands ; Discussant, Cleo Redline, National Science Foundation, USA
- 37 25. Contributed Session : Internet - II ; Chair, Fred Conrad, University of Michigan, USA ; Satisfaction Scales in a CAWI Survey on University Teaching Evaluation, Luigi Fabbris, University of Padua, Italy ; Using the Multimedia Capabilities of Web-Enabled Probability-Based Survey Methodology to Gather Vaccination Information for the National Immunization Program, William McCready <bmccready@knowledge networks.com>, Michael Dennis and Lisa Thalji, Knowledge Networks, USA ; Web as a Questionnaire Designing Tool : Is There a Limit ?, Katja Lozar Manfreda <katja.lozar@uni-lj.si>, University of Ljubljana, Valentina Hlebec and Vasja Vehovar ; Discussant, Dianne Anderson, Iowa State University, USA
- 38 26. Invited Session : International Issues in Questionnaire Development ; Chair, Lilli Japac ; Evolution And Adaptation of Questionnaire Development, Evaluation And Testing

Methods in Establishment Surveys, Diane Willimack, U.S. Census Bureau, Lars Lyberg <lars.lyberg@scb.se>, Statistics Sweden, Jean Martin, Office for National Statistics, U.K., Lilli Japac, Statistics Sweden, Patricia Whitridge; Developing Cross-National Survey Instruments, Tom Smith <smitht@norcmail.uchicago.edu>, NORC, University of Chicago, USA; Discussant, Gustav Harlidsen <gha@ssb.no>, Statistics Norway

- 39 27. Contributed Session : Cognitive Interviewing Methods - II ; Chair, Ashley Landreth ; A Comparison of Focus Group and One-on-One Cognitive Interviewing for Questionnaire Evaluation, Kristen Miller, U.S. National Center for Health Statistics ; Interviewer Debriefing by E-mail, Birgit Henningsson, Statistics Sweden ; Pretesting the American Time Use Survey, Lisa Lee <leelisa@norcmail.uchicago.edu> and Catherine Haggerty, NORC, USA, Diane Herz and Lisa Schwartz, U.S. Bureau of Labor Statistics ; Discussant, Ger Snijkers, Statistics Netherlands
- 40 28. Invited/Contributed Session : Interaction Analysis, Question Sequencing, and Behavior Coding to Improve Questionnaire Performance ; Chair, Jim Esposito ; The Diagnostic Approach : Evaluating Survey Questions by Analyzing Patterns of Behavior Codes and Transcripts of Question-Answer Sequences, Johannes van der Zouwen <zouwen@scw.vu.nl> and Johannes H. Smit <jh.smit.emgo@med.vu.nl>, Vrije Universiteit, Amsterdam ; Evaluating Questionnaires by Analyzing Question-Answer Sequences, Wil Dijkstra and Yfke Ongena, Vrije Universiteit, Amsterdam ; What's the Interviewer Have to Do with It ? : Interviewer Behavior and Response Rates, Kathryn Downey-Sargent <kathryn.downey-sargent@arbitron.com>, Elisha Smith, Barbara O'Hare, Arbitron Company, USA ; Discussant, Nora Cate Schaeffer, University of Wisconsin
- 41 29. Contributed Session : Technical Documentation ; Chair, Debbie Collins, U.K. ; On the Documentation and Analysis of Electronic Questionnaires, Jelke Bethlehem <jbthm@cbs.nl> and Anco Hundepool, Statistics Netherlands ; New Tools for the Documentation of Questionnaire Development, Frauke Kreuter <frauke.kreuter@uni-konstanz.de>, UCLA, USA, Rainer Schnell, University of Konstanz, Germany ; Discussant, Patricia Doyle, U.S. Census Bureau
- 42 30. Invited Session : Statistical Methods for Developing and Evaluating Questionnaires ; Chair, Jana Asher, Carnegie Mellon University, USA ; A Scientific Approach to Questionnaire Development, William Saris <wsaris@planet.nl>, William van der Veld, Irmtraud Gallhofer and Irmgard Corten, University of Amsterdam, Netherlands ; The Use of Latent Class Analysis for Identifying Flawed Questions, Paul Biemer, RTI International, USA ; Discussant, Colm O'Muircheartaigh, NORC, University of Chicago, USA
- 43 31. Contributed Session : Innovative Methods from the Netherlands and Israel ; Chair, Nancy Mathiowetz, JPSM, University of Maryland ; How to Investigate Interaction Patterns Between Partners in a Large-Scale Survey : The Development of the Dutch Family Monitor, Dirkjan Beukenhorst <dbkt@cbs.nl> and Deirdre Giesen, Statistics Netherlands ; Informed Consent : The Assessment of Competency in Elderly Persons with Cognitive Impairments, Johannes Smit, Astrid Vellinga, Evert van Leeuwen, Willem van Tilburg, and Cees Jonker, Vrije University, Amsterdam, Netherlands ; Interrater Reliability in an Imperfect Field Setting, Dror Walk <drorw@jdc.org.il>, Rachel Fleishman, Gad Mizrahi, and Miriam Bar-Giora, JDC-Brookdale Institute of Gerontology and Human Development, Israel ; Discussant, Peter Mohler, ZUMA, Germany
- 44 32. Contributed Session : Dependent Interviewing ; Chair, Johannes van der Zouwen, Free University, Netherlands ; Pre-Printing Effects in Official Statistics, an Experimental Study,

Anders Holmberg <anders.holmberg@scb.se>, Statistics Sweden ; Alternative Methods for Exploring Confidentiality Issues Related to Dependent Interviewing, Joanne Pascale, Thomas Mayer and Julia Klein Griffiths, U.S. Census Bureau ; Survey Design Techniques for Web Interviews, Adriaan Hoogendoorn, Free University, Netherlands ; Discussant, Eleanor Singer, University of Michigan, USA

Abstracts of Invited Papers

Design Considerations for Pretesting, Roger Tourangeau

- 45 This paper examines the statistical and design issues involved in conducting pretests for questionnaire development. Researchers often view pretesting as a qualitative activity that doesn't require much attention to statistical issues (such as power) or to traditional experimental design issues (e.g., confounding). Even when a pretest compares two or more versions of a question experimentally, the pretest may be regarded as exploratory so that statistical and experimental design issues can be disregarded. This paper argues for the opposite view. It discusses some of the key design issues raised by experiments that compare different versions of a questionnaire. These include a) the relative advantages of the laboratory and field as settings for such experiments, b) the pros and cons of factorial designs versus designs that compare questionnaires that differ along multiple dimensions simultaneously, and c) options for assigning the different versions of the questionnaires to different sample units. The paper also discusses such statistical issues as power, selecting an appropriate alpha level for significance tests, and practical versus statistical significance. Attention to classical experimental design concerns is likely to yield clearer results from questionnaire design experiments, as well as findings that are more likely to cumulate across studies. The paper also argues that greater attention to some of these same considerations can help yield more conclusive findings from pretests based on genuinely qualitative techniques, such as cognitive interviewing and behavior coding.

New Perspectives on Cognitive Interviewing : What Does it Mean to Say We're "Lab Testing" a Questionnaire ?, Gordon Willis

- 46 The general use of cognitive interviewing in order to detect problems in survey questions has become widespread. However, there is currently no definitive evidence that this technique is effective, or agreement concerning appropriate standards of evidence that demonstrate effectiveness. Further, although it is clear that cognitive interviews are carried out very differently across practitioners, it is not necessary clear what underlies such variation, or which variants are best for particular purposes. I argue that in order to evaluate technique efficacy, we first need to determine more precisely why we are conducting these interviews, what disciplinary backgrounds may be influencing our practices, and what practitioners expect cognitive interviews to reveal. In this paper I examine several perspectives which give rise to variations in practice, and suggest how each of these leads to different criteria for use in evaluation studies, or variation in the manner in which these techniques should be practiced or further developed.
- 47 In particular, two theoretically-oriented perspectives are reviewed: the Cognitive perspective, and the Anthropological/Ethnographic. The paper discusses the manner in

which a focus on each implies modifications in the manner in which we develop further theory or alter practice in terms of subject recruitment, the nature of interviewer-based probing, or analysis of qualitative results. Further, a discussion is presented of several perspectives on how interviewer probing should be accomplished; specifically, whether it should be more oriented toward problem verification or discovery, and whether it should involve probing that is oriented toward the tested question, as opposed to probing of the answer to that question.

The Dynamics of Cognitive Interviewing, Paul Beatty

- 48 The growth of cognitive interviewing has had a profound impact on questionnaire development over the last fifteen years. Useful overviews of cognitive interviewing methodology have been written that discuss the general varieties of interviewing behavior (e.g., encouraging participants to “think out loud”, concurrent probing, retrospective probing, and so on), and provide examples of each. However, many parameters of cognitive interviewing have not been clearly established. Although cognitive interviewers are often given suggested probes, they also maintain a great deal of freedom regarding what they may say during the interview. One reason to look at actual cognitive interviewing practice in more depth is to foster discussions about “best practices.” Another is to foster continuing methodological research on pretesting methods.
- 49 The primary goal of this chapter is to explore the dynamics of the cognitive interview, or how interviewer behavior shapes what participants say— and possibly what is concluded as well. This investigation opens several other important questions: how can we determine that cognitive interview findings are “real” survey problems? More generally, what specific probing strategies do cognitive interviewers employ— and do they do anything other than probe per se? The investigations reported in this chapter are limited to one cognitive laboratory (and are primarily based on one cognitive interviewing project). Rather, than generalizing to all cognitive interviews conducted anywhere, the objective is to explore the dynamics of interviewing within this particular project in depth, determining what was actually done and how that might affect what is concluded.

Assessing Data Quality in Cognitive Interviews, Frederick G. Conrad and Johnny Blair

- 50 Cognitive interviewing is well into its second decade of use for pretesting surveys yet there is no single accepted definition of the method. Specific cognitive interview techniques are constructed from a menu of laboratory procedures producing many disparate techniques. It seems reasonable to think that these techniques may differ in their data quality. Although these data are routinely used to determine whether questions for major surveys are in need of repair, very little research on data quality has been done.
- 51 In the first part, we propose a methodology for assessing cognitive interview data quality. We propose that problem identification is the fundamental purpose of verbal reports. The methodology addresses the interpretation of verbal reports, the coding of question problems, and the analysis of these data. Two kinds of reliability and three kinds of validity are discussed.

- 52 The second part illustrates the methodology by describing a study that compares data quality -- in particular two kinds of reliability -- of two cognitive interview techniques. One technique represents the practices of experienced cognitive interviewers. The other technique closely follows procedures used in psychology for eliciting verbal reports and constrains interviewer probing to explicit indications of problems in respondents' verbal reports. The results suggest that verbal reports about answering survey questions are difficult to interpret consistently leading to concerns about the quality of problem detection in cognitive interviewing. They further suggest that constraining interviewers' probes leads to fewer but more reliably identified problems.

Comparing the Effectiveness of Alternative Methods of Cognitive Interviewing, Theresa J. DeMaio and Ashley Landreth

- 53 In recent years, cognitive interviews have become widely used for pretesting questionnaires in the Federal government and survey organizations, and have become accepted as a survey methodological tool. There is no standardized definition of what a cognitive interview is, however, and variations exist in the way cognitive interviews are conducted. Different survey research organizations conduct cognitive interviews in different ways, and these differences may have implications for the ultimate objective of pretesting -- identifying problems in the questionnaire and making recommendations for changes to improve the accuracy of the data collected.
- 54 This paper presents the results of a split-panel experiment to evaluate alternative methods of conducting cognitive interviews, in an effort to address some of the gaps in our knowledge of the effectiveness of cognitive interviewing. The experiment includes three panels, which reflect actual differences in interviewing practice among survey research organizations. The three methodological approaches include the following elements that might be expected to impact the cognitive interviewing results : 1) the type and training of personnel who conduct interviews ; 2) the role of survey researchers in the overall process ; 3) the types of data collected and the degree to which they are reviewed ; and 5) facets of the interview protocol and the degree to which interviewers treat it as flexible or rigid.
- 55 The results of the experiment include : 1) an assessment of the number and types of problems identified by the methods ; 2) a comparison of the problems identified by the methods with an independent "standard;" and 3) an examination of the recommendations for questionnaire revision resulting from the three methods.

Using Interaction Analysis for the Identification and Explanation of Inadequate Parts of a Questionnaire, Johannes van der Zouwen and Johannes H. Smit

- 56 The analysis of the interaction between interviewer and respondent in survey interviews (sequence analysis for short) usually focuses at the description of general mechanisms like the effect of question format or interviewer competence on the interaction. Only rarely sequence analysis has been used as a 'diagnostic' instrument for the evaluation of (parts of) questionnaires, probably because it evokes two methodological challenges.

- 57 Firstly, only non-paradigmatic sequences provide information about interactional and cognitive processes. As these sequences form a minority in a survey, a procedure for their selection and retrieval is required. Secondly, a sequence can be viewed as the 'product' of the competence of the interviewer, of the cognitive capacities of the respondent, and of the quality and difficulty of the question. Because we are interested in the assessment of the quality of the question, we have to apply a method for controlling the effects of interviewers' competence and respondents' capacities.
- 58 The method developed for dealing with these challenges is illustrated by data from the third survey (1999) of the Longitudinal Aging Study Amsterdam (N=1771). LASA studies autonomy and well-being of elderly persons (60-90) and includes a large variety of instruments measuring social and health related topics.
- 59 The LASA data are quite appropriate for the illustration of the 'diagnostic' approach. Firstly, all interviews are audio recorded, which enables the selection of particular non-paradigmatic sequences : for example those with non-substantive responses (like 'don't know') or (combinations of) responses that are very unlikely. The analysis of the selected sequences aims at identifying those characteristics of the question that have caused the problems of cognitive or interactional nature that eventually led to these problematic data.
- 60 Secondly, we are able to check the robustness of the outcomes of the analysis because a number of interviewers had conducted more than 100 interviews and the data were re-analyzed for each of these interviewers separately. Finally, in LASA the cognitive capacities of the respondent are assessed by different cognitive tests, which makes it possible to classify respondents according to their cognitive capacities and see whether the outcomes of the sequence analysis differ between these groups.
- 61 The 'diagnostic' approach is illustrated by applying it to a part of the questionnaire consisting of eight questions : about the income of the respondent, retrospective questions about the occurrence of a decrease of income, and about the satisfaction with, and expectations about, this income. These questions belong to the standard instruments of survey research. Nevertheless, the analysis of non-paradigmatic sequences showed all kinds of misunderstandings of core concepts (like net income) ; and a mix up of a nominal and 'real' increase of future income ; of expectations about, versus hope for, improvements, etc. Also an unexpected question order effect was detected.
- 62 The analysis also shows that the occurrence of (specific types of) non-paradigmatic sequences is related to the cognitive capacities of the respondent and competence of the interviewer, and - most important - that these sequences can be linked to particular question characteristics. The diagnostic sequence analysis clearly shows how particular questions lead to problematic data.

Response Latencies and Perceived Question Difficulty as Indicators for Response Error, Stasja Draisma and Wil Dijkstra

- 63 Measurement of response latencies in survey research has often been used to trace several kinds of problems with different types of survey questions and question formulations.
- 64 For example, Bassili and Scott (1996) found that problematic question formulations resulted in longer response latencies. Several studies have been performed which

demonstrate the relation between attitude strength, accessibility, and intensity on the one hand and response latencies on the other. Moreover, Basilli and Fletcher (1991) demonstrate that discernible types of questions obtain different response latencies : easy factual questions obtain shorter reaction times than supposedly more difficult ones.

- 65 In some of our earlier research (Dijkstra, Draisma and Van der Zouwen 1995 ; Draisma, 2000) relationships between question characteristics that affect cognitive processing and response errors were investigated. It was found that the difficulty of the questions was related to the probability of giving a correct answer and non-substantive (“don’t know”) responses.
- 66 Response latencies may thus be indications for the difficulty of questions and for the quality of the data obtained and be useful for the evaluation of survey questions. In the paper we will address the following research questions : (1) Which procedures to assess response latency (RL) in answering survey questions are possible and valid, considering different operationalizations and measurement procedures ? (2) Are these RL’s related to response errors ? (3) Is the perceived difficulty of questions related to response errors ? (4) Are RL’s related to the perceived difficulty of questions, as judged by the respondents themselves and by an expert jury ? (5) Are RL’s related to other paralinguistic and linguistic indicators of uncertainty ?
- 67 Data came from a telephone survey among approximately 300 members of a large Dutch environmental organization. Of the survey questions to be used, the individual ‘true scores’ or correct answers could be determined by information from the records of the organization (for instance about membership duration and the size of the periodical contribution paid to the organization). All interviews were tape recorded, so that response latencies could be exactly determined. After the interview, respondents were asked to evaluate some previously posed questions according to the perceived difficulty.
- 68 The consequences for the usefulness of measuring reaction times for the practice of survey research, especially as an indicator for response error, will be discussed. It is argued that response latencies are a clear indicator of problems in answering survey questions. As a measure of information processing in survey interviews, we suggest to obtain response latencies during the interview, which can be implemented easily into CATI and CAPI techniques.

Pretesting Strategies to Improve Respondent Comprehension and Recall in Factual Surveys, Elizabeth Martin

- 69 Over the past two decades, there have been substantial theoretical and empirical advances in the understanding of cognitive sources of response error which have led to richer understandings of errors introduced by failures of comprehension or retrieval. Two field-based methods which may be applied to investigate these sources of error are respondent debriefing and the use of vignettes to identify comprehension and retrieval problems in a survey.
- 70 This paper describes how vignettes and respondent debriefing may be applied to identify measurement problems and to craft and test questionnaire designs to address the problems. By conducting an iterative program of design and pretesting, it is possible to gain much richer knowledge both about the performance of questions and the nature of the errors affecting measurement of a phenomenon. To illustrate and illuminate

problems and strategies for addressing them, the paper draws upon research (much of it hitherto unpublished) conducted for the redesign of several Census Bureau surveys, including the National Crime Survey and the Current Population Survey. Four types of applications of vignettes to questionnaire design are described and illustrated, primarily using research conducted to address problems of interpretation and comprehension in the Current Population Survey. The application of respondent debriefing questions is illustrated using questionnaire design research conducted to address recall and retrieval problems in the National Crime Victimization Survey, as well as other surveys. The advantages and disadvantages of the two methods are compared and contrasted, and evidence pertaining to their validity and consistency with other measures is summarized.

Getting Beyond Pretests and Cognitive Interviewing : The Case for More Split-Ballot Pilot Studies, Floyd Jackson Fowler, Jr

- 71 The past decade or so has seen three important evolutions in the routine procedures that are used for pre-survey evaluation of questions. First, some kind of cognitive testing is frequently used to help evaluate how questions are understood and what answers mean. Second, field pretests are often now augmented with behavior coding, which makes the evaluation of the question and answer process more systematic. Third, there has been some advance (though perhaps less developed and widely accepted) in systematic pre-survey evaluation of questions using fixed standards, such as those developed by Lessler.
- 72 There is no doubt that these techniques have improved the ability of researchers to identify problem questions. On the other hand, based on these techniques alone, researchers do not have information about how “problems” that are identified, or the “fixes” that are made, will actually affect the resulting data. This is not a trivial problem for at least three reasons.
- 73 1. Sometimes fixing a problem, for example defining a term or concept that is misunderstood by some respondents, makes a question worse from other perspectives. For example, the new added definition may make the question harder for an interviewer to read as worded. In that context, knowing how much difference the fix makes in data quality is important.
- 74 2. In a similar way, one of the most important conservative forces pushing for not fixing bad questions is the desire to use items from previous surveys to maintain comparability. When “problems” are found, how much the problems affect data quality, and how improved versions of the questions will affect mean estimates, are important considerations.
- 75 3. Finally, even if the above two issues are not relevant, when a researcher “fixes a problem”, it often is important to validate that in fact the changed question produces data that are likely to be of better quality, whatever that means.
- 76 Split-ballot experiments, in which parallel questions are asked of comparable samples, provide the potential to address these questions. Without them, the significance of question problems identified in pre-survey evaluations and how changes in questions will affect data and data quality are likely to remain unknown.
- 77 This paper presents data from a number of question evaluation studies that illustrate how split-ballot pilot studies can help evaluate proposed question changes that emerge from pre-survey question evaluation, such as cognitive testing. The results demonstrate how

split-ballot studies contribute to realizing the full potential of presurvey question evaluation protocols for improving data quality.

The SIPP Methods Panel Project : Employing Multiple Research Tools to Improve Instrument Design, Jeff Moore, Anna Chan, Julia Klein Griffiths, Joanne Pascale, Pat Doyle

- 78 The Census Bureau established the Methods Panel project to evaluate and redesign the questionnaires for the Survey of Income and Program Participation (SIPP). The overall objective of the project is to produce improved core questionnaires for use in the 2004 SIPP Panel. Specific improvement goals include reducing burden and improving efficiency (and thereby reducing nonresponse and attrition), revising question wording to improve “flow” and naturalness, and improving data quality.
- 79 The Methods Panel employs multiple research techniques, but its three field experiments -- designed to allow three iterations of testing and refining the Wave 1 core instrument and two iterations for the follow-on Wave 2+ instrument -- are the focus of this paper. Each split-sample test includes about 2,000 interviewed Wave 1 households, with 1,000 randomly assigned to each of the control (standard SIPP) and experimental (new and improved SIPP) instrument treatments.
- 80 Field experiments bring both strengths and weaknesses to the task of improved questionnaire design, the particulars of which vary according to the improvement goal. This paper examines the ability of the Methods Panel’s field tests to yield strong evidence about the effectiveness of various attempts at questionnaire improvement. We find split-panel field tests fairly well-suited for evaluating some types of improvement attempts — attempts to reduce nonresponse (including item nonresponse) and attrition, for example, and efforts aimed at improving the performance of the survey instrument in the field (efficiency, CAI de-bugging, interviewer evaluations). For other types of improvements (e.g., improved data quality) such tests generally supply much less definitive evidence.

Modeling Measurement Error to Identify Flawed Questions, Paul Biemer

- 81 This paper proposes a general strategy for investigating flawed survey questions consisting of four steps : (a) conduct preliminary data analysis to identify potentially flawed questions (i.e., questions with poor reliability or high levels of classification error), (b) conduct further data analysis to elucidate the probable sources of error for the questions identified, (c) verify the sources and identify the root causes of the problem through the collection of additional data, and (d) develop and implement appropriate solutions to eliminate the problem. The focus of the paper then turns to steps (a) and (b). Several general methods will be described and illustrated using real data and the strengths and weaknesses of each method will be discussed. The emphasis of the paper is on actual applications of the methodology rather than statistical theory.
- 82 Three general evaluation designs will be discussed in some detail. One design considers the case where two locally independent measurements are available for a random sample from the population. These measurements maybe parallel (i.e., have identical error properties) or non-parallel. For example, the remeasurement might be considered to be a

gold-standard measurement or a measurement with unknown statistical properties. The second design extends the discussion to three measurements which be locally dependent as well as non-parallel. For example, all three measurements may come from the same interview using different questions, or combinations of questions, to measure the same characteristic. Finally, the three measurement case is extended to remeasurements which correspond to different points in time as in a panel survey situation. In this design, no remeasurements are available other than the panel survey measurements. The usefulness of this technique for data mining will be exploited in an example from the U.S. Current Population Survey.

A Scientific Approach to Questionnaire Development, W. E. Saris, W. van der Veld, I. N. Gallhofer, A. Scherpenzeel

- 83 In an ongoing project an inventory has been made of all the choices which have to be made in the development of items for survey questionnaires. It has been shown by Scherpenzeel (1995) that different choices may change the correlations between the variables considerably. Therefore, the effects of these choices on the reliability, validity and method effects of survey items have been studied in different countries. Andrews (1984) and Rogers, Andrews and Herzog (1992) made a study in the US, Költringer (1995) did a study in Austria on German surveys, Scherpenzeel and Saris (1997) did a study on Dutch surveys. In a recent project a database with all these experiments is built. The data base contains at this moment 1067 measurement instruments based on 87 experiments done on random samples from at least regional but most of the time national samples of 300 till 2000 respondents. The purpose of this study is to generate cross national generalizations of the findings which have been published so far based on national studies. This analysis provides a quantitative estimate of the effects of the different choices on the reliability validity and the method effects.
- 84 The result of that study makes it possible to implement these estimates in a computer program to predict the quality of a survey item before data are collected. For Dutch questionnaires a prototype of a full automatic "Survey Quality Prediction" program called "SQP" has been made. SQP reads survey items, codes them on the characteristics which have effects on the reliability, validity and method effects and makes predictions of the quality of these items before the data is collected on the basis of all so far collected information about effects of design factors. For the moment the program SQP works only for Dutch requests (Van der Veld, Saris and Gallhofer, 2000) but the idea is to develop a program, first of all, for English, German and Dutch and extend it later to other languages. For the moment the quality prediction for English and German questionnaires is done with a non automatic program where the users have to answer a series of questions which code the item on the necessary characteristics. After that the program predicts the quality of the question. The quality estimates generated by SQP can warn researcher for low quality survey items before their data is collected and provide suggestions for quality improvement. In this way we hope that questionnaire development becomes more a scientific activity and will not be an art anymore.

Concepts and Procedures for Testing Paper Self-Administered Questionnaires : Cognitive Interview and Field Test Comparisons, Don A. Dillman and Cleo D. Redline

- 85 In practice the primary use of cognitive interviewing methods has been to identify wording problems in interview-administered questionnaires. In this paper we provide a conceptualization of ways in which the objectives and procedures for testing self-administered questionnaires may differ from those used for interview instruments. In addition we report three case studies in which cognitive interviews and field experiments were conducted simultaneously so that the results from each could be compared. Each case study focuses on a different issue. The first case study reports evaluations of alternative Census questionnaire mailing packages in which likely response rates from alternative designs was the issue of primary interest. The second case study reports evaluations of alternative formats for providing branching instructions to respondents, where the issue of interest was to identify the capabilities of these designs for lowering the number of branching errors made by respondents. The third case study measured item-nonresponse to a particular question that exhibited high rates of nonresponse in nonexperimental data collections. In each case conclusions reached from the cognitive interviews are contrasted with those reached from the field experiments.

Usability Testing as a Means of Evaluating Computer-Assisted Survey Instruments, Sue Ellen Hansen and Mick P. Couper

- 86 Computer assisted interviewing (CAI) affects how interviews are conducted, in that it segments the questionnaire, presenting one screen at a time, and the computer controls the flow of the interview. Evaluation of CAI survey instruments therefore should extend beyond traditional techniques, to the evaluation of their usability, which focuses on the impact of instrument design on users (interviewers or respondents) of computers. Usability focuses on the degree to which the computer makes it possible for the user to complete tasks easily and correctly. In contrast to other questionnaire design problems, problems that are strictly usability problems focus on screen design issues, that is, placement of information on the screen, the way screen elements are formatted and made distinct from other elements, the consistency of design across computer screens, and the impact of any aspect of questionnaire design on user-computer interaction.
- 87 The primary methods of instrument usability evaluation are : (1) usability inspection methods, that is, evaluation or review by one or more experts ; (2) evaluation of automatically generated performance data, such as counts of functions invoked, (3) usability testing, or laboratory-based observation of computer assisted interviews. This paper focuses on the latter. Six sections (1) review prior CAI research, (2) present a conceptual model of the computer assisted interview, (3) present guidelines for CAI design based on principles of human-computer-interaction (HCI), (4) provide an overview of usability evaluation methods, (5) describe in more detail usability testing as an evaluation method, and (7) presents findings from laboratory-based usability tests.

Methods for Testing and Evaluating CATI Questionnaires, John Tarnai and Danna Moore

- 88 A particularly difficult task in computer assisted interviewing (CAI) is adequate testing and evaluation of CAI questionnaires, especially since this is an easily neglected task. Over half of survey research centers nationwide report that they have written procedures for testing and debugging CAI questionnaires, and yet over 65 % of them also report having had to recontact survey respondents because of errors found in a CAI survey. In this monograph we summarize the literature on testing and evaluating CAI questionnaires, and report the results of comparing common methods of testing and debugging CAI questionnaires. The focus here is not on evaluating questionnaire wording, but instead on testing and evaluating how well the questionnaire has been programmed. Included among the CAI testing methods that we evaluate are scenario testing, computer simulation, and other commonly used methods. The monograph concludes with recommendations based on the results of these comparisons and discusses the effectiveness of using computer simulations versus other methods to test CAI questionnaires. We believe that computer simulation of CAI interviewing offers much potential to survey researchers in improving the accuracy of questionnaires. This paper suggests ways of improving the testing and evaluation process for survey researchers using CAI.

Testing Web Questionnaires, Reg Baker and Scott Crawford

- 89 Questionnaire development and testing for traditional paper and pencil questionnaires have tended to focus on the key elements of question type selection, wording, and order. Where self-administration is used, graphic design is also important. The transition to computer-assisted methods has added a new set of concerns, namely technical correctness (e.g., branching, calculations, text fills, etc.) and, more recently, usability.
- 90 The advent of web-based interviewing has further complicated the questionnaire development and testing phases of survey design and implementation. We now face new concerns about web-based communication styles, unique question types and formats, screen layout and use of color, added complexity to tests of technical correctness, and technical performance (e.g., speed of page delivery, appearance under different browsers, etc.).
- 91 This paper describes a series of testing protocols for Web questionnaires. It conceives of a Web questionnaire as having six main components: presentation, instructions and questions, logic and functionality, respondent environment, application software, and hosting platform. Effective testing requires that each of these components be evaluated and tested individually, as well as a thorough test of the assembled whole. The type of testing done and its point in the life cycle of survey design and implementation will vary by component. In developing these protocols the authors draw on the survey methods literature, the literature on software testing, and their own experience both testing computer-based questionnaires and deploying hundreds of Web surveys.

Does Pretesting Improve the Quality of Survey Questions ? An Empirical Test within a Field Survey Environment ?, Jennifer Rothgeb, Barbara Forsyth and Gordon Willis

- 92 Questionnaire pretesting using cognitive methods is standard practice for U.S. statistical agencies and organizations that design or conduct national surveys. Some common methods include expert review, cognitive interviewing, and behavior coding. Informed decisions about pretest standards and practices are enhanced by pretest methods research. This paper presents research designed to determine (1) whether cognitive pretesting predicts actual problems encountered in survey administration and (2) whether survey administration and data quality improve with revisions based on pretest results.
- 93 The research was conducted in two phases. In phase 1, researchers at three organizations applied a set of pretest methods to a collection of questionnaire items. We used a classification scheme to code questionnaire problems identified in pretesting (see Rothgeb *et al.*, 2001). In phase 2, we developed revised question wordings based on pretest results. The original and revised questions were embedded in an RDD telephone survey conducted by the Census Bureau. We gathered three measures of data quality : behavior coding data, item nonresponse rates and interviewer ratings.
- 94 Analyses address three research questions : 1. Do pretest results predict problems in survey administration or data quality in the field experiment ? 2. Do questionnaire revisions made based on pretest findings produce improved survey administration or data quality in the field experiment ? 3. Do pretest results predict the types of improvements observed in the field experiment ? The paper discusses costs and benefits from using accepted cognitive methods to identify survey questionnaire revisions, along with suggestions for how the testing and revision process might be improved.

Improving the Clarity of Closely Related Concepts, Nora Cate Schaeffer and Jennifer L. Dykema

- 95 We report on a development and testing effort that combined multiple methods to attempt to improve the measurement of joint legal custody. In the U.S., when parents live apart, a court may grant parents joint legal custody, so that both parents have the authority to make decisions about the children ; this concept is difficult to measure, however, because the common language used to refer to it is ambiguous and because it can be easily confused with physical custody. Our study began with a series of eight focus groups, followed by four rounds of cognitive interviewing. The final survey, the Parent Survey 3 (PS3) included two versions of the series of questions targeted at joint legal custody. Responses in the PS3 can be compared with the legal record abstracted in the Court Record Database (CRD). We have several ways to evaluate our efforts : the accuracy of the response, how sure the respondent was about her or his answer, the relationship between how sure the respondent was and her or his accuracy, whether any differences between the two forms of the instrument are reflected in the interaction codes, and whether the interaction between the interviewer and respondent can inform us about the accuracy of the respondent's answer. We also compare our results to those of an earlier survey to further assess the results of our development efforts. Overall, it appears that

the target question has a low proportion of negatives that are false, but a substantially higher proportion of positives that are false.

Design, Testing and Evaluation of Stated Preference Questionnaires for Environmental Valuation, Michael D. Kaplowitz, Frank Lupi, John P. Hoehn

- 96 The reported research illustrates an iterative, multiple method approach for designing and evaluating a self-administered stated-preference questionnaire for environmental valuation. Stated preference questionnaires, common in marketing research and increasingly used to estimate economic values for environmental quality, describe the attributes of goods and services and ask respondents to make choices. Crafting stated preference questionnaires for environmental valuation is inherently difficult because environmental resources are intrinsically complex and not widely understood. In the reported case, an adaptive questionnaire design approach was used for questionnaire development. The design phase of the research began with a series of focus groups and a structured group interview with subject matter experts. Based on the revealed information, two alternative prototype questionnaires were developed and later evaluated using focus groups and semi-structured individual cognitive interviews. The result of this evaluation was the selection and redrafting of the preferred questionnaire. The draft questionnaire was then tested and evaluated using a series of individual pretest cognitive interviews. Three sets of individual pretests and debriefings were conducted with randomly recruited members of the general public (80 total). Each cognitive interview began in a flexible manner with an open format, and gradually shifted to evolved into more structured interviews with specific questions. The reported research used mixed methods because different methods were expected to, and did, reveal different types of information. The case study shows the usefulness of an iterative, multiple methods approach to designing and evaluating questionnaires.

Current Practices in Questionnaire Development, Evaluation and Testing for Establishment Surveys : An International Overview, Diane K. Willimack, Jean Martin, Patricia Whitridge, Lilli Japac and Lars Lyberg

- 97 Establishment surveys are essential to the national accounts and other short-term economic indicators for monitoring and managing the economy. In addition, they provide data supporting decision-making processes in fields such as education, transportation, health care, and agriculture. Characteristics of establishment surveys that differ from household surveys, along with a more elaborate survey response process, have resulted in an alternative culture for development and testing data collection instruments for establishment surveys. This paper documents methods used for establishment survey questionnaire development, evaluation and testing (QDET). Based on a review of publicly available literature, supplemented by our own international survey of government statistical agencies, national statistical institutes and other survey organizations, we will describe how establishment survey QDET methods have evolved to accommodate the special needs and circumstances of establishments. Where appropriate, our discussion will also note similarities and differences between establishment methods and household

methods in terms of how they are conceived and implemented. In addition, since establishment surveys have been at the forefront in developing electronic data reporting, we will also present an overview of methodologies used to test and evaluate electronic instruments. We will conclude by identifying gaps in the QDET process for establishment surveys, and suggest areas for future research and development.

Children as Respondents : Developing, Evaluating and Testing Questionnaires for Children, Natacha Borgers & Edith de Leeuw

- 98 Children are no longer neglected as respondents in surveys. They participate more and more in surveys. However, systematic methodological knowledge on survey techniques and questionnaire development for children is scarce, and researchers have to rely on ad-hoc knowledge from diverse fields as child psychiatry and educational testing or on methodological knowledge on how to survey adults.
- 99 The purpose of this paper is (1) to integrate the current theoretical and empirical knowledge regarding questionnaire research with children as respondents, and (2) to present instruments and strategies for the evaluation and testing of questionnaires for children.
- 100 There are special problems to be solved when developing and testing questionnaires for children. Children still develop the cognitive and social skills, which are necessary to answer questions. Although this is a continuous developmental process, it is useful to distinguish successive stages of development, each of which presents their own difficulties to survey research with children. We present a theoretical frame for the different stages of children's development and its consequences for questionnaire development. We also summarize the existing empirical knowledge on children as respondents, focusing on what is known for different stages of development. We present a checklist for children's questionnaires based on this empirical knowledge. This checklist can be used both as a guideline when constructing a questionnaire, and as a coding scheme for the evaluation of questionnaires (e.g., in expert-evaluation).
- 101 Besides expert evaluation, other frequently used methods for questionnaire testing are focus groups, cognitive (in depth) interviews, and observations, such as the monitoring of standardized interviews and self-administered questionnaire sessions. We systematically discuss these methods for children's questionnaires. In this we reflect on what the development of children and their cognitive and social abilities in different stages mean for how researchers may test and evaluate questions. Thus providing guidelines and prerequisites for the optimization of questionnaire testing methods for different age groups.

Developing Cross-National Survey Instruments, Tom W. Smith

- 102 As challenging as developing questions, scales, and entire questionnaires within a monocultural and monolingual context is, the task becomes considerably more difficult when done in a multi-cultural and multi-lingual setting. Overlaying the standard need to create reliable and valid measures are the complications inherent in cross-cultural and cross-national differences in language, customs, and structure. Only by dealing with these challenges on top of the usual instrument design issues can scientifically credible cross-national survey instruments emerge.

- 103 Considering the value of cross-national research, the importance of obtaining comparable measurements, and the frequent failure to take measurement seriously, there is an obvious need for general improvement. This chapter contributes towards that goal by discussing 1) the development of equivalent questions in surveys, focusing on a) the question-asking and b) answer-recording parts, 2) response effects that contribute to measurement error in general and variable error structures across nations, considering in particular social desirability, acquiescence bias, extreme response styles, Don't Knows (DKs) and non-attitudes, neutral and middle options, response order, question order, and mode of administration, and 3) steps to enhance validity and comparability in cross-national surveys, including the form of source questions, translation procedures, and item development and pretesting.

Past Conferences and Related Monographs

- 104 This conference is the latest in a series of international gatherings of researchers in the survey field :

- International Symposium on Panel Surveys, 1986, Washington, D.C., USA
- International Conference on Telephone Survey Methodology, 1987, Charlotte, NC, USA
- International Conference on Measurement Errors in Surveys, 1990, Tucson, AZ, USA
- International Conference on Establishment Surveys : Survey Methods for Businesses, Farms, and Institutions, 1994, Buffalo, NY, USA
- International Conference on Survey Measurement and Process Quality, 1995, Bristol, England, UK
- International Conference on Computer Assisted Survey Information Collection, 1996, San Antonio, TX USA
- International Conference on Survey Nonresponse, 1999, Portland, OR USA
- International Conference on Establishment Surveys-II, 2000, Buffalo, NY USA

- 105 Each of these conferences has brought together active researchers in the particular sub-field. All of them led to interchanges of perspectives and comparisons of findings that stimulated further advances in the fields. Collectively, the papers presented in the conferences form a solid base of the survey statistical and methodological literature over the past ten years. In addition to the excitement and intellectual stimulation the conferences offered, each has produced a product of more lasting and broader value -- monographs summarizing the state of the art of the field :

- *Telephone Survey Methodology*, Groves, Biemer, Lyberg, Massey, Nicholls, and Waksberg, (Wiley, 1988)
- *Panel Surveys*, Kasprzyk, Duncan, Kalton, and Singh, (Wiley, 1989)
- *Measurement Errors in Surveys*, Biemer, Groves, Lyberg, Mathiowetz, and Sudman, (Wiley, 1991)
- *Business Survey Methods*, Cox, Binder, Chinnappa, Christianson, College, and Kott, (Wiley, 1995)
- *Survey Measurement and Process Quality*, Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin, (Wiley, 1997)
- *Computer Assisted Survey Information Collection*, Couper, Baker, Bethlehem, Clark, Martin, Nicholls, and O'Reilly, (Wiley, 1998)
- (All of these monographs can be obtained from www.wiley.com)

ABSTRACTS

The International Conference on Questionnaire Development, Evaluation, and Testing Methods (QDET) took place in Charleston, South Carolina, on November 14-17, 2002. There were 32 sessions with 22 invited papers and 54 contributed papers. In addition, there were 15 poster presentations. The author describes the meeting, presents the entire program (sessions / posters, authors, titles) and provides the abstracts for the invited papers.

Cette conférence a eu lieu du 14 au 17 novembre 2002 à Charleston dans la Caroline du Sud. Il y a eu 32 sessions comprenant 22 présentations invitées et 54 autres présentations. De plus, 15 présentations sur affiche ont été faites. L'auteur décrit ici cette conférence, présente le programme complet (sessions / affiches, auteurs, titres) et fournit les résumés des papiers invités.

INDEX

Mots-clés: Enquêtes, Questionnaires, Méthodes de test, Méthodes de développement

Keywords: Surveys, Questionnaires, Testing Methodes, Development Methods

AUTHOR

JENNIFER M. ROTHGEB

U.S. Census Bureau ; jennifer.m.rothgeb@census.gov