



QUEST Workshop 2009

Bergen, Norway

18th to 20th of May 2009

Abstracts

Questionnaire Testing for Business and Establishment Surveys at Statistics Canada

Dave Lawrence, Statistics Canada

Since the early days of questionnaire testing, there has been much emphasis placed on evaluating social and household surveys. Over the past 20 years, Statistics Canada has tested a wide variety of business and establishment questionnaires. What have we learned from our experiences? How has this helped us establish sound testing practices? In which areas do we need to improve or gain more experience?

This paper provides a brief synopsis of the history of business and establishment survey questionnaire testing at Statistics Canada, looking specifically at practices and protocols over time. The paper also addresses the question of what does the future hold for researchers with respect to issues such as electronic data reporting, usability testing and international comparability of questions.

Relations between Cognitive Problems and Statistical Bias

Gustav Haraldsen, Statistics Norway

In this paper I will raise two questions. First I will ask to what extent qualitative methods like cognitive interviewing disclose questionnaire problems that actually have a significant effect on the survey results. Secondly I would like to turn this question the other way around, and ask to what extent the statistical bias recognised in surveys capture systematic errors detected in qualitative tests. The topic will be discussed with reference to examples from some qualitative and quantitative tests that were carried out in a Norwegian survey about working conditions. There both different terms and different response scales used in the questions were first addressed qualitatively during the questionnaire development and subsequently tested in quantitative experiments and re interviews during the field period.

Benefits of Concurrent Cognitive and Usability Testing

By Jennifer Childs, Jennifer Romano, Erica Olmsted-Hawala, and Elizabeth Murphy,
U.S. Census Bureau

Typically survey pretesting involves separate timelines and research staffs for cognitive and usability testing. In this paper, we make the case that a more comprehensive and less labor-intensive approach to pretesting is to conduct both cognitive and usability testing concurrently with usability and cognitive testing experts working together. By testing the same questionnaire concurrently with respondents and interviewers (the users in this case), potentially problematic question wording and instrument design can be more efficiently

identified in a way that can be used to improve the questionnaire for both the respondent and the interviewer.

Prior to the 2006 Census Test, the U.S. Census Bureau conducted separate rounds of cognitive and usability testing on an interviewer-administered non-response follow-up questionnaire in preparation for the 2010 Census. The usability testing was conducted in the Census Bureau's Usability Lab with an early version of the instrument. Later, the Census Bureau's Cognitive Lab conducted cognitive testing of the instrument. In doing the testing separately, we learned that in addition to usability issues, usability testing also identifies cognitive question wording issues. Similarly, while examining question wording, cognitive testing also identifies poor usability features. Though many techniques used in these two types of studies are similar, usability staff are not typically trained specifically in survey methodology, and survey methodologists are not typically trained in human-computer interaction. Usability staff are perfectly capable of recommending changes in wording, but their recommendations may not be informed by research known to the survey methodologists who conduct cognitive testing. Similarly, survey methodologists may make recommendations for improving usability that are not backed up by research in human-computer interaction. While well intentioned, such uninformed recommendations could do more harm than good.

Based on this observation, in 2008, the Cognitive and Usability Labs at the Census Bureau conducted 40 cognitive and 20 usability interviews concurrently and in conjunction to test the questionnaire and presented results and recommendations from both types of testing together. When testing is conducted concurrently, staff from both labs, representing both specialties, can be at the table at once, creating a more efficient methodology. By examining these two case studies, this paper will discuss what can be gained by conducting these studies in concert above and beyond conducting them independently. Examples of the kinds of findings that are possible through this joint research and the synergy from having both research teams involved will be described.

Key words: Cognitive Testing, Usability Testing, Survey Pretesting

Laying the foundation for good survey questions

Bente Hole, Statistics Norway

This paper focuses on the very first stages of a questionnaire design- and evaluation process, i.e. on what Esposito calls the observation and conceptualisation phases (Esposito 2002). These phases involve forming concepts, defining important sub-domains of the concept's meaning and finding empirical indicators for each concept or sub-domain (Hox, 1997). The paper points to some of the central literature on the subject, presents a few examples on how this part of the process has been implemented in practice and discusses advantages and drawbacks related to the different approaches. It seems clear that an iterative process is needed and that proper assessment in connection with every iteration is crucial in order to achieve good results. Further more it is considered whether techniques related to focus groups and exploratory interviews can be of value and whether a combination of a bottom up and a top down research strategy is profitable in order to fill in the gap between theory and measurement.

Testing of concepts - trust as an example.

Merja Kallio-Peltoniemi, Statistics Finland

My presentation concerns the testing of concepts. The testing of the theoretical concept is important when formating an operational definition of a phenomenon and before formulating a survey question. A concept can, however, also be tested afterwards with data driven approaches. We can test a concept from a "what we actually measured" point of view and ask if the operational definition is the same as the definitions used by the respondents describing the phenomenon. My focus is on such data driven approaches.

The Cognitive Laboratory of Statistics Finland is about to test the concept of trust in summer 2009 with cognitive and thematic interviews. Trust (in other people, confidence in institutions) is an important indicator of social capital. In this presentation I will introduce the test protocol we are going to use and discuss how to use the results from the concept testing as a tool when interpreting the statistical measures of trust.

keywords: testing of concepts, trust, definitions by respondents

Testing of questions about working hour patterns.

Ellen-Merete Duvold, Statistics Norway

The Norwegian Living Conditions Survey contains questions about different working hour patterns. Our challenge has been to formulate questions adequate to grasp the employees' actual working hour patterns. The aim of this paper is to test the questions (and their wordings) and important concepts with particular emphasis on response categories for shift and rotation work.

Cognitive interviews were undertaken with four respondents currently engaged in jobs with shift- or rotation working hours. How do they experience the categories of shift and rotation? What are their perceived definitions of weekend, night and day? According to their own understanding, do the working hours actually rotate or shift? How do they define terms like night, morning, early afternoon and evening? When does the weekend start and when is it finished? Are there any differences between objectively defined timings of the clock and subjectively defined timings of the mind? If so, what do these differences consist of? How can we explain these differences?

The current questionnaire was tested cognitively, both by means of regular telephone interviews and by means of "think aloud" sessions, supplied with use of cards. Findings and recommendations for the questionnaire are discussed.

Key words: Patterns of working hours, Cognitive testing

Using split ballot testing to evaluate findings of how to best measure health knowledge questions

Carol Cosenza, University of Massachusetts Boston

There are many formats a researcher can use to measure health knowledge. Offering respondents answer choices in a closed ended format has potential pitfalls. A respondent could simply guess and have a relatively good chance of choosing the correct answer. Whatever options the researcher includes provides the respondent with some boundaries and the potential to rule out answers that are wrong. For example, if the answer options to a question are “1-5, 6-10, 11-20, 20 or more”, the respondent knows that the answer is probably a relatively small number and has at least a 25% chance of just guessing the correct answer. However, respondents who are asked the question in an open-ended format are given no boundaries and therefore have the ability to answer whatever they want.

This paper will examine data collected from a series of split ballot experiments with questions that measure respondents' perception of health risk (such as being diagnosed with breast cancer or dying from colorectal cancer). We will compare and contrast two formats for asking knowledge questions. Half of the respondents will be asked a question in an open-ended format (for example: Out of every 100 people, about how many will get colon cancer?). The other half of the respondents will be asked the question in a closed ended format (Out of every 100 people, about how many will get colon cancer? Please choose the number that you think is closest to the correct answer: 2, 6, 14, 24, or 43).

Using multiple methods to understand ethnicity response Lyn Kaye, Statistics New Zealand.

"In recent years there has been a growing number of the New Zealand population who describe themselves as 'New Zealanders' in response to the ethnicity question in the country's Census of Population and Dwellings. Respondents are opting to write in this description under the category of 'other', rather than select from the more specific ethnic groups named in the response list. This paper discusses the research that has been undertaken at Statistics New Zealand to attempt to understand this trend. In particular, it explores the results of several qualitative research projects undertaken to date, including cognitive testing and focus group testing. It further outlines future quantitative projects expected to expand the findings from this early work. Finally, it draws some conclusions about the ways in which multiple research methods can inform and enrich our understanding of respondent behaviour."

Standardization at Statistics Sweden - experiences and challenges

Helena Bäckström, Gunilla Davidsson and Andreas Persson,
Statistics Sweden

In 2007 Statistics Sweden was re-organized from an organization based on products to a process-oriented organization. All methodologists were centralized to a process department. For each process involved in production of statistics, a group of process owners are supposed to supply a set of standardized methods and tools. For us working as questionnaire designers in the measurement lab, this meant that we had to standardize our methods. During 2007 two standardization projects were conducted to develop standard procedures, guidelines and checklists for the following methods: focus groups, in-depth interviews, expert reviews, cognitive interviews, usability testing and editing staff debriefing. These guidelines and checklists are now included in a process-support tool, that includes all processes and can be found by all employees at our intranet (our local network).

During 2009, we will conduct another project to try to establish guidelines concerning the choice of pre-test methods. The goal is to develop some sort of checklist to help survey managers in deciding which pre-test strategy would be most useful in testing their survey.

Thus, Statistics Sweden has many new standards concerning pre-testing. However, many challenges remain in putting them into practice.

Developing Standard Questions for Surveys – Do Standard Questions Work?

Paul Kelly, Statistics Canada

In 2006 Statistics Canada started an initiative for greater standardization of questionnaire modules across its household survey program. At that time very few standard approaches were in place for collecting socio-demographic and economic data in the various household surveys administered by the agency.

The benefits of standardized questionnaire modules across social surveys can be significant in terms of cost and time savings. There may also be a reduction in respondent and interviewer burden with the use of common questions. Data coherence across surveys and the ability to combine data from different sources would certainly improve with an initiative such as this.

However, what are the challenges and what are the risks? This paper will describe the ongoing Statistics Canada project to standardize social survey questions. It will explore the benefits and challenges of implementing such an initiative.

Health Insurance Measurement: A Synthesis of Cognitive Testing Results

Joanne Pascale, U.S. Census Bureau

Over the past two decades, health researchers have been grappling with the issue of how to best measure health insurance. These efforts have included literature reviews, comparative studies, taxonomies of differing methodologies and estimates across surveys, cognitive testing, and split-ballot field experiments. This paper focuses on one aspect of these efforts – cognitive testing – whose aim is to better understand how the questions are being understood and answered from the respondent’s perspective, and to identify features of the questions and questionnaire as a whole that may be associated with misreporting. Individual reports on cognitive testing, while informative, lack a certain power due to the subjective nature of this type of testing and the small sample sizes. To address these shortcomings, an effort was made to assemble all known reports on cognitive testing of health insurance questionnaires and to synthesize the findings across those reports in order to identify common, persistent problems. This approach reduces subjectivity (given the increase in the number of contributing researchers), increases the sample size of respondents, and increases the chances that the final set of synthesized findings is comprehensive (i.e.: that no important findings are overlooked). Altogether ten reports from various federal agencies and private research firms were identified, and they assessed nine different questionnaire designs, including several major national surveys (e.g.: the Current Population Survey, the American Community Survey and the National Health Interview Survey), as well as some experimental redesigns. Though each questionnaire was unique, many contained items with very similar wording, and several questionnaires had common features, such as a household-level design. This paper synthesizes results across the ten reports and takes some first steps toward recommendations for a research agenda and best practices for questionnaire design.

Key words: cognitive testing, measurement error

The design and testing of questions for mixed-mode surveys: Lessons learnt and considerations for the future.

Michelle Gray, National Centre for Social Research

As noted by de Leeuw (2005), Mixed-mode surveys have been in use for a long time and have become the norm in some countries as survey managers seek to use collection procedures that produce the best possible data within existing constraints of time and budget. Supporting this, and given the fact response rates are on the decline, offering the choice of mode of data collection has been shown to be an effective way of improving response rates (Shettle and Mooney, 1999). We also know that some respondents have preferences for the mode in which they receive the questionnaire and therefore mixed mode designs allow for tailoring of mode for specific populations.

If there is to be a move towards more mixed mode designs, a concern for survey researchers is whether people who respond by one mode would have provided the same answers had they responded in another mode. Experiments have shown, for example, that in the presence of an interviewer (i.e. face-to-face and possibly even telephone), respondents are more likely to

experience social desirability effects compared to the self-administered format. When a stimulus is completely visual, in self-administered formats, the respondent has complete control over how fast the question and answer categories are read and whether they are even read in their entirety, leading to possible primacy or recency effects depending on what types of short cuts are taken in reading. Similar effects may be found when using showcards, while for telephone interviews with read out answer categories, recency effects are most likely.

When mixing modes, the goal to achieve equivalence is to adhere to a Unimodal design, whereby questions are designed to provide the same stimulus in all survey modes to reduce differences in the way respondents respond to the survey questions in different modes.

This paper will discuss the use of mixed modes: some of the advantages and disadvantages, some practical considerations and some lessons learned, with particular emphasis on designing and pre-testing unimodal questions. Specific examples from a recent project, which involved unimodal questionnaire design and testing to feed into a business survey toolkit, will be used to aid the discussion.

What do respondents want and expect from electronic self-completion surveys? A discussion.

Lucy Tinkler, Office for National Statistics

The use of electronic modes as a means of collecting self-completion survey data has become increasingly popular. Surveys that use electronic modes of data collection are most often thought of as those which are completed directly on the internet, and those that are administered via email. The Office for National Statistics (ONS) has recently completed a literature review into research investigating respondent perspectives on electronic modes of data collection, the focus of which was respondent preferences. It was found that there has been little research carried out into whether respondents prefer to complete questionnaires using electronic modes, what respondents understand an electronic 'questionnaire' to be, and, the level of burden incurred compared with paper self completion. ONS proposes to investigate this topic further focusing on business surveys. In this session Lucy Tinkler will briefly outline the research that has been carried out into this topic, this will be followed by a discussion exploring the type of research and methods ONS data collection methodology branch could employ to carry out this research.

How to present “don’t know” in web surveys.

Rachel Vis-Visschers, Statistics Netherlands

Statistics Netherlands (StatNeth) investigates the possibility to execute its social surveys in a mixed-mode design, since primary data collection is expensive. Web interviewing will be an increasingly important mode. In 2007 StatNeth’s Questionnaire Laboratory conducted several pre-tests to investigate question formats that work for multiple-mode surveys.

In this paper we will discuss the results of a cognitive laboratory test in which four different ways of presenting “don’t know” (DK) in a web survey were tested:

- Option 1: DK is always presented together with the other answer categories.
- Option 2: DK is never offered. Respondents are forced to choose one of the offered answer option.
- Option 3: DK is not explicitly presented on screen, but when a respondent tries to skip a question a warning appears, that the question has to be answered. The answer option DK has then appeared on the screen as well. This procedure is explained at the beginning of the questionnaire.
- Option 4: DK is presented as a button at the bottom left side of the computer screen.

According to our expectations, we found that DK was answered most often when the answer category is always presented on screen (option 1). We also found that respondents got frustrated when they did not know the answer and it was not possible to answer DK (option 2).

Contrary to our expectations options 3 and 4 appeared not successful. For both options we found that half of the respondents knew how to find DK and half of the respondents did not.

Keywords: cognitive test, don’t know, web survey, mixed mode research.

Testing questions on sexual identity: experiences from optional data collection modes.

Elisabeth Gulløy, Gustav Haraldsen, Aina Holmøy and Marit Wilhelmsen,
Statistics Norway

For the first time in 2008/2009, the Norwegian living conditions survey includes questions on sexual identity. The initiative came from external users of statistics; government, research and interest organisations. To ask questions on sexual identity is traditionally seen as problematic in social surveys. Sensitive issues are a possible threat to response rates. The relationship between sensitivity and non-response is complicated. From previous studies we know that four groups of respondents are expected to be particularly sceptical to questions on sexual identity:

- The elderly, since they seldom speak about these issues, or are not sexually active
- Persons from immigrant societies with strong taboos on sexuality
- Persons being uncertain about their own sexuality, and therefore uncertain about what to answer
- Persons with an established sexual identity unknown to their families, friends and colleagues

Statistics Norway decided to develop a draft set of questions for a pilot to be included in the living conditions survey focusing on health 2008-2009. The project included both literature studies and careful testing of questions. This included both formerly used questions picked up from other researchers or statistical agencies, and our own newly developed questions. The process can be characterised as an iterative process of testing – development – new testing – revision etc. Cognitive and focus group interviews included both potential respondents and

Statistics Norway's own interviewers. The development and testing gave the following conclusions:

1. we must secure an understanding among the respondents about sexual identity as a relevant dimension of living conditions and health, and that the questions are an integral part of a living conditions survey framework
2. sexual attraction, orientation and identity are different concepts and the survey questionnaire has to reflect this
3. sensitivity is a larger problem for interviewers than for most of the respondents

We ended up with four questions, ideally to be asked in a personal interview/telephone interview, but with an option for including them in a paper questionnaire distributed by mail after the interview. The questions aim to measure what sexuality means for quality of life, sexual attraction to same/opposite sex, sexual identity, and finally whether sexual orientation has ever caused problems in relation to family/friends/colleagues. Heterosexuals will only be asked the first two questions. The final question is only asked to homosexuals, lesbians, bisexuals, or those who claim they are none of these alternatives. The paper will present preliminary results regarding data collection from the field work. The following questions will be discussed:

- Preparation of interviewers ahead of data collection
- How many respondents did abrupt the interview when we reached the questions on sexual identity?
- How many respondents preferred to receive questions by mail instead of giving their response directly to an interviewer?
- What kind of feedback from respondents did the interviewers report?
- Was sensitivity actually a problem for most of the interviewers?
- Optional mixed mode: how did it work out for the organisation?
- Conclusions on the development of new survey questions on sexual identity: did it work out as expected?

What Kinds of Problems Does Cross-Cultural Pretesting Reveal?

Gordon Willis, National Cancer Institute, National Institutes of Health

Attempts to attain cross-cultural equivalence of survey questions have increasingly relied on the adaptation of cognitive interviewing and behavior coding. The National Cancer Institute has increasingly applied these methods across a range of survey questionnaires, in multiple modes, and across multiple language groups. The presentation will summarize the procedures used for these studies, and in particular will focus on key findings from four studies: (a) Korean versus non-Koreans, for the telephone-administered California Health Interview Survey; (b) Multiple Asian languages and Spanish, for an interviewer-administered U.S. national tobacco surveillance survey; (c) English, Spanish, Chinese, and Korean for a self-administered questionnaire on cancer risk factors; and (d) Highly acculturated versus low-aculturated Hispanics, for interviewer-administered items on physical activity, and on level of acculturation to U.S. society. Across these investigations, the results fell into three basic categories: (a) Translation Problems; (b) Problems of Cultural Adaptation; and (c) Generic problems of Questionnaire Design. Somewhat surprisingly, the latter category appeared to

dominate. I will discuss these results and their implications for cross-cultural and multi-national questionnaire development and pretesting.

Challenges in designing and testing questionnaires to be administered in multiple languages on sensitive topics.

Margaret Blake, National Centre for Social Research

In the United Kingdom and many other countries large scale surveys covering the general population include respondents from a wide range of ethnic, cultural and linguistic backgrounds. Where the topic of the survey is particularly relevant for ethnic minorities or the sample includes a boost over-sampling of certain ethnic groups, interviews may be offered in multiple languages to enable inclusion of respondents who are not fluent in the main language(s) of the country.

Offering interviews in multiple languages is challenging for a number of reasons:

- The ethnic minority population in the UK and other countries is diverse and the list of potential languages which could be offered is long.
- The topics of surveys focussing on ethnic minority groups often include questions which may be sensitive for some groups (e.g. violent extremism, migration, alcohol consumption).
- Lack of trained (cognitive) interviewers who read, write and speak necessary languages.
- Timetables and budgets for projects which often do not allow for fully testing translated questions.

New questions on surveys are often cognitively tested and then piloted prior to the main fieldwork. Where translations are needed, these are often carried out after cognitive testing in the source language only and prior to piloting in multiple languages. This approach raises a number of questions which will be addressed in the presentation:

- Is cognitive testing of translated questions needed or is good cognitive testing of source language enough?
- If cognitive testing of translated questions is needed how do we deal with the fact that cognitive testing in all the languages to be offered by the survey is impractical?
- What other approaches could be used instead of cognitive testing? For example, round table discussions of questions with language experts prior to translation, qualitative interviews in translation prior to development of questions, respondent debriefing.
- How do you build a team of cognitive interviewers with the language skills needed?
- When the researchers on the project do not speak the languages in question, what is the best way to approach the testing in terms of translation of probes, writing up of notes etc and how can the quality of these processes be controlled?
- Should changes be made to the source language questionnaire if problems identified by cognitive testing relate to cultural rather than linguistic issues?

The paper will outline the challenges facing researchers in developing and testing translated questionnaires and discuss possible solutions drawn from the experiences of other organisations, as reported in the literature. I hope that researchers present at Quest also will be able to share and learn from their experiences in this area.

Experiences from testing public user surveys "adapted" to the immigrant population in Norway

Elisabeth Gulløy, Statistics Norway

Do respondents from different cultures in Norway understand questions on their opinions regarding public sector services in a similar way? What kind of challenges do we meet when we try to adapt a standardised public user satisfaction survey to a diverse immigrant population? These questions will be discussed when we present the test results from a development project commissioned by the Directorate of Integration and Diversity (IMDI) and performed by Statistics Norway in cooperation with Institute for Labour and Social Research (FAFO).

The project's aim is to develop guidelines for standardised public user satisfaction surveys specially adapted to measure the experiences and opinions of an increasingly diversified minority population in Norway. Such surveys are already established for monitoring public user satisfaction in the population as such. In spring 2009, we will develop a questionnaire more or less adapted, and more or less similar to, the ordinary version. This questionnaire will be tested in a row of cognitive interviews. Results from the testing will be important in developing guidelines for future surveys in this field.

Important dimensions in the testing will be

- interpretation of key constructs: public service, public servant, availability, service information, service quality, service satisfaction
- comparing key constructs in origin culture and Norwegian culture
- interpretations and cognition of common user satisfaction survey questions
- meaning of responses
- meaning of scales
- social meaning of response process

Dual-Development of the 2009 Rehearsal England and Wales Population Census Questionnaires

Ruth Wallis, Office for National Statistics

A Welsh language Household questionnaire has been used in Wales to collect population statistics in the Census since at least 1841, but, in 1993, the Welsh Language Act made the use of a Welsh language questionnaire a legal requirement. Until now, the process used for developing the Welsh language Census questionnaires has been to translate the English questionnaire into Welsh, at the end of the English questionnaire development process.

For the 2011 Census, a new method has been implemented to develop equivalent Welsh and English questionnaires. This method builds on that used for the bilingual New Zealand Population Census, using a dual-development approach to designing questions in Welsh and English. Core to this methodology is the parallel development of questions in both languages. By using this process, issues unique to each language can be given consideration at every stage of the development cycle. Both languages are given equal status throughout the development cycle, so that compromises are not made at the expense of either language, and ultimately both versions of the questionnaire should meet an equivalent quality standard. Testing so far shows that the resulting questionnaires will be easier to complete for the Welsh-speaking population, and will collect data of a higher quality than previous Welsh Census questionnaires. This paper outlines the dual-development process.

Data Quality in Cognitive Interviewing: A new perspective on standardizing probes in multi-language and multi-cultural projects

Stephanie Willson, National Center for Health Statistics

The strength of cognitive interviewing as a method of question evaluation lies in the ability of the analyst to examine the construct validity of survey items. This is accomplished using in-depth interviews whereby interviewers explore how respondents interpret and answer survey questions. The advantage of the method is the flexibility that allows interviewers to explore issues as they arise during data collection. However, data quality as well as the ability to compare and summarize findings across interviews conducted by multiple interviewers in different languages among culturally distinct groups requires some degree of standardization. The question is how to achieve standardization without losing the benefits of the qualitative method.

Most efforts at achieving standardization have focused on how to word probes (i.e., follow-up questions). Informed by quantitative methods, this approach suggests that in order to obtain conceptually specific and comparable data from respondents, all must receive exactly the same follow-up questions. It's believed that imposing this level of control over the interview process improves cognitive interview data quality by minimizing the effects of interviewer skill level and maximizing the chances that specific topics will be covered.

This paper is an examination and evaluation of probes, with special emphasis on how they may be used to collect standardized and comparable data in cross-national or cross-cultural projects. Using data from cognitive interview evaluation studies conducted by the Questionnaire Design Research Laboratory at the National Center for Health Statistics, this paper will show that there are two different styles of probing techniques, each motivated by competing assumptions about what respondents know and how they are able to report it. The first approach assumes that respondents use logic and reason to understand survey questions. The second assumes that respondents base their understandings of questions on personal experiences. This paper will argue that the key to standardization lies in implementing a consistent approach to probes, shifting attention away from the specific wording of probes to acknowledging the underlying assumptions that probes make about how respondents are able to answer questions. Furthermore, the paper will demonstrate that the quality of interview

data obtained by probes that assume an experiential epistemology is higher than the quality of data obtained by probes assuming a logic-based epistemology.

A proposal of best practices for conducting cognitive interviewing in cross-cultural/national surveys

José L. Padilla, University of Granada

Due to the growing number of surveys in which different cultural and linguistic groups are involved, several organizations and research groups have recently tried to provide information on best practices or guidelines, across the multiple phases of cross-cultural/national survey. Among the most significant achievements, it is worth mentioning the cooperative Survey Design and Implementation (CSDI) Guideline. The CSDI Guidelines take care of the various organizational and operational aspects that should be considered in the structural design of a cross-national project. Focusing on the cognitive interviewing, the purpose of this paper is to propose detailed best practices for conducting cognitive interviewing in cross-cultural/national surveys. Together with a review of the “state-of-art” in the field, the proposal is mainly based on the “lessons learned” from the multi-national testing project conducted by the Comparative Cognitive Testing Workgroup. The project was coordinated by Kristen Miller, representing the Budapest Initiative, and Rory Fitzgerald from the European Social Survey. The paper will intend to stimulate discussions among QUEST meeting participants on the aims of the best practices (to provide criteria for evaluating cognitive interviewing or to guide cognitive interviewing designs); how to organize best practices in meaningful categories; how to describe best practices, etc. Finally, the design of a survey on Internet intended to get expert comments on the strengths and weaknesses of the proposal will be presented.

Cognitive interviewing: How to manage qualitative data.

Karen Blanke and Sabine Sattelberger, Federal Statistical Office, Germany

For several years, the Federal Statistical Office has been working on the systematic implementation of questionnaire testing. Marking an important step within this development, our pretest laboratory was established by the end of 2007. Questionnaires of paper-and-pencil as well as online surveys are now increasingly evaluated by qualitative testing methods. In the long run, we aim for reducing the burden for respondents and for increasing data quality for official statistics. However, it is not only a challenge to do pretesting, but also to organize the overall evaluation process as systematically as possible. Our intention is to develop efficient standards for questionnaire testing and to distribute tasks among several colleagues.

In the majority of pretests, a three step approach is applied: Firstly, we observe subjects, while filling in a questionnaire on their own (“reality without words”, e.g. shaking the head, moving backward and forward through the forms as indicators for possible problems). Secondly, we conduct interviews, using cognitive testing protocols, to show what respondents have

understood in theory. Both steps are videotaped. Thirdly, we evaluate the self-completed questionnaire, whether it was completed correctly, due to our knowledge after the interview. Combining these three sources of information (observation, cognitive interview and questionnaire) provides a huge amount of qualitative data that has to be analyzed and stored.

The analysis of audio and video recordings is highly time-consuming. The overall transcription of cognitive interviews is – due to lack of time and human resources – impossible. However, in order to avoid the risk to draw final conclusions from first general impressions, we use a computer based tool, often applied in social sciences and psychology. Apart from organizing team work, the underlying idea is to structure and to store data, to accelerate analysis and to simplify interpretation. On the whole, we set great store that our results follow objective criteria and can be verified by colleagues.

This presentation aims to give an insight into work in progress. We put our methods up to discussion and initiate exchange of experience with regard to benefits and constraints of how to manage qualitative data electronically.

Q-Notes: Development and use of analysis software for cognitive interviews to examine survey question comparability

By Kristen Miller, National Center for Health Statistics

Recent work has shown that cognitive interviewing studies can provide essential information regarding the comparability of survey questions, specifically, how respondents interpret and process questions and whether particular sub-populations or groups may process questions differently from others. To achieve this goal, however, studies must be based on empirical evidence and systematically analyzed across interviews and sub-populations—a process which can yield a massive amount of qualitative data across numerous countries and in multiple languages. To be sure, one of the biggest challenges for comparative, multinational cognitive testing is data management, that is, the organization and reduction of cognitive interview data such that it can be analyzed systematically. This paper will describe software that was specifically developed to analyze cognitive interviews in this capacity. To illustrate the software's use, the paper will draw heavily from the Washington Group's testing project to evaluate an extended set of disability questions in a global context.
