

2015 QUEST Workshop

April 28th – 30th at the Statistics Finland in Helsinki, Finland

Session 1

Cognitive Interviewing as an Inductive Method to Construct Question Design Principles **Stephanie Willson, National Center for Health Statistics**

Designing survey questions that capture the constructs intended to be measured is not a straightforward process. Indeed, the act of writing survey questions has often been likened more to art than science – an odd premise, given that the task of survey data collection is located squarely in the realm of statistical measurement. This paper takes a scientific approach to question design by employing an inductive method to arrive at question design principles. Through an examination of findings from multiple cognitive interviewing studies, it will identify common interpretive patterns associated with disparate survey questions on health topics, summarize the connective theme among the patterns, and posit a question design principle based on these findings. A primary advantage of an inductive approach to question design is that the conclusions are grounded in empirical evidence. Data are examined from reports in Q-Bank, an on-line repository of cognitive interviewing final reports. Survey topics include smoking behavior, cancer control, dental health, and alternative medicine. The author will demonstrate that different kinds of questions have the potential to (unintentionally) create dichotomous categorizations of self-concept, some of which do not exist a priori among respondents but, rather, emerge as an artifact of the question-response process.

Effects of dependent interviewing on respondents and interviewers **Sophia Nebel, Federal Statistical Office of Germany**

In the context of the reform of household statistics Germany plans to adopt an infra-annual rotation pattern for the Labour Force Survey (LFS), which will consequently reduce the distances between the interviews and increase the burden of households. Currently independent interviewing is predominantly used for conducting household surveys. Especially in panel studies respondents are confused by confronting them with recurring questions for several waves. Therefore the Federal Statistical Office of Germany (FSO) is planning to implement dependent interviewing as a data collection method for the German microcensus. Dependent interviewing is a questioning method, in which data of prior interviews is used to substantially shorten the interview and to design the current survey instrument. Through the use of existing data, the consistency between different survey waves can be increased. Furthermore the technique should help to simplify the response process for respondents and thus bring a temporal and cognitive relief. But at the same time the risk of an underestimation of change rises, because respondents may be tempted to report no change in order to avoid follow up questions. However, there is only little evidence in empirical literature about the effects of dependent interviewing on respondents and interviewers. To test the feasibility and practicability of the technique the FSO carried out a pretest, where the behavior of the interviewers, the reaction of the respondents and the application of the survey instrument was reviewed using cognitive interviews and behavior coding. The presentation focuses on results and challenges of the pretest.

Questionnaire design in Mixed-mode - the use of open or closed questions. **Fredrik Scheffer & Pia Hartwig, Statistics Sweden**

During the year of 2014 the Unit for Cognitive methods at Statistics Sweden took part in developing a web option to a telephone survey on party preferences. The web option was evaluated in an experiment. Among many challenges in the mixed-mode experiment, the biggest was likely how to design the main question of the survey, the party preference question, for the web. The question is open-ended in the telephone questionnaire but does not necessarily have to be open-ended on the web. However, open-ended and closed questions can lead to different cognitive processes and, in the end, different response distributions. Two web versions of the main question were designed. One with an open-ended format where suggestions appeared based on the two first typed letters.

The other version was a closed question with parties in a for Sweden traditional left wing – right wing order. Another challenge was how to design the “don’t know” options for comparison with telephone interviewing where this option is not visually available for the respondent. The main subjects (dependent variables) for question design evaluation were differences in item nonresponse and “don’t know”-answers between different modes as well as different question designs (open- or closed ended). This presentation will present the results from these evaluations.

On the relationship between response burden and response quality
Karianne Lund, Statistics Norway

The European Statistics Code of Practice (CoP) for the National and Community Statistical Authorities states that we, as the national statistical office, are obliged to keep the response burden "proportionate to the needs of the users and should not be excessive for respondents". An underlying premise for this principle is the assumption that there is a correlation between response burden and response quality. This paper summarizes findings from 3 separate but inter-connected studies exploring the correlation between response burden and response quality. Our first study explored the relationship between the response burden and the measurement error by comparing the level of perceived response burden and the rate of control violations. Conclusions from this study suggest no significant connection between perceived response burden and response quality. In our second study, we concluded that there was a low correlation between response burden and response quality using a slightly different approach. Here, the suggested relationship between the response burden and measurement error were explored by comparing both “perceived” and “actual” response burden with the level of manually edited key variables and the total number of reminders. In our third study, we tried to shed light on both previous studies by using new empirical data from Statistics Norway’s annual survey on depreciation of operational resources.

On questionnaire design and pre-testing in mixed mode
Marjaana Järvensivu, Statistics Finland

Over the past 4 years, Statistics Finland has invested considerably in introducing the web data collection for household surveys. During this development phase, we have pre-tested cawi questionnaires for the Labour Force Survey (LFS) and the Use of Information and Communications Technology by Individuals (ICT) survey. The pre-testing has mainly concentrated on usability issues of the cawi questionnaires. However, at the same time, pre-testing reveals problems in comprehension and interpretation of questions. The challenge has been how to ensure that question wording and format used in the cawi mode meet the survey specific requirements of measuring (i.e. data equivalent to and combinable to cati data). At the same time, very important viewpoint to keep in mind in questionnaire design is to support respondent’s fluent answering process when filling in the cawi questionnaire without the help of the interviewer. Therefore, we have confronted the need to modify question wordings and formats either in the cawi questionnaire or in the cati questionnaire, and sometimes in both modes. We are currently reflecting, how to achieve the balance between the basic presumption that survey questions should be as identical as possible between modes (unimode approach) and, at the same time, take into account that mode effects might be reduced by optimizing each questionnaire for corresponding mode (mode-specific approach). In my presentation, I will present some pre-testing results from both the LFS and the ICT survey. I will also discuss how the chosen approach (unimode or mode-specific) in the questionnaire design affects what kind pre-testing might be needed and if the pre-testing result concerning one mode are exploitable to another mode.

Up, Down, and Sideways: What Happened When We Presented Response Options in Double Columns
Carol Cosenza, Center for Survey Research - University of Massachusetts Boston

The “business” of survey research has us all trying to find the most cost-effective way to collect the best data we can. Every day we ask questions like: Can we ask fewer questions without compromising the integrity of the data? What if we did it all online? How about just texting responses? Does a paper instrument really need a cover page? Or a full page of instructions? And just how important is spreading out the questions and having a lot of “white space”? We decided to look at those last issues more closely. As part of a field test of a health survey conducted in a university-based health system, we undertook several small methodological experiments with alterna-

tive formatting of the questionnaire – specifically focusing on the presentation of the response options. We created three different versions. In all of them we compressed the original (standard) 12 page survey into 4 pages. Version 1 presented the response options in one vertical column following the question. Version 2 presented the response options in two vertical columns. Version 3 presented the responses in 2 horizontal rows. We're looking at how respondents answered using the alternative formats, how they compared to the standard form, and why there might be differences.

Session 2

Pretesting survey questions via web probing – Does it produce similar results as f2f cognitive interviewing?

Timo Lenzner, GESIS

Asking probing questions in online surveys (web probing) has recently been advocated as a promising method for evaluating survey questions. For example, Behr et al. (2012) have demonstrated that web respondents give meaningful answers to open-ended probing questions and that this information can be used for the post-survey assessment of item validity. The present study examines whether web probing can also be used to evaluate questions prior to their use in a survey. In particular, we examine the following three research questions: 1. Does web probing produce similar results as f2f cognitive interviewing? 2. How many probes per question and how many probes per page should be used? 3. Is it advisable to implement non-response probes if respondents provide unsatisfactory (e.g., very short) answers to probing questions? 508 respondents drawn from a non-probability online panel were randomly assigned to one of three experimental conditions that varied the number of probing questions asked, the number of non-response probes asked, and the number of probing questions presented per survey page. The web survey included four questions from the ISSP 2013/2014 that had been tested previously via f2f cognitive interviewing. In our qualitative analyses, we examine whether web probing and cognitive interviewing identify similar problems in these four questions. In addition, we examine quantitative indicators such as drop-outs and survey completion times between the three conditions. Practical implications of the findings and directions for future research on web probing are discussed.

Owning a bicycle - A Case Study and Evaluation of pre-test interviews executed by regular interviewers Rachel Vis-Visschers, Statistics Netherlands

At Quest I would like to present a case study. In January I conducted a pre-test on a really Dutch subject: owning a bicycle. In our regular Mobility Survey we have a time series of how many bicycles are in a household and how many persons own a bicycle. This time series was very steady, until 2012. In 2013 questions about electrical bicycles (e-bikes) were introduced. The introduction of the new questions caused a mode effect. In Web it appeared that 20% less people owned a bike, while in CAPI and CATI the trend was unchanged. To test what happened and how the questions could be improved, in order to reestablish the time series, we executed a pre-test. For this test we asked the assistance of five regular Stat Neth telephone interviewers. In the presentation I will address, why the new questions on e-bikes could cause such a mode effect. I will discuss the method of the pre-test, and our recommendations. And finally I will evaluate the test method. Was the assistance of the regular interviewers efficient?

The value of eye tracking in testing web and paper documents Debbie Collins & Jo D'Ardenne, NatCen Social Research

The value of eye tracking in testing web and paper documents: reflections and tips In recent years eye tracking technology has become more widely available and more affordable, allowing researchers to move beyond people's verbal reports on what they are doing to also collect objective evidence of what people look at. Eye tracking is commonly used in market research but its use in social research has been patchy. In this paper we reflect on our experience of using eye tracking combined with cognitive interviewing to test web and paper instruments. We will draw on our experience from three studies: testing a web questionnaire, an event history calendar and an

electoral ballot paper and consider whether the eye tracking element identified any issues or problems that would have otherwise gone undetected. We will also discuss practical issues to consider when using eye tracking as a methodology and make some suggestions regarding its wider use.

The Use of Common Office Software for Wireframing and Early Stage Questionnaire Testing
Brandon Kopp, Bureau of Labor Statistics

I have often encountered restrictive timelines for usability testing and cognitive interviews that require testing before a functioning web or computer-assisted instrument can be developed. At first this meant using paper tools, but paper tools do not provide the tester with an authentic experience. I then discovered a powerful set of tools for prototyping/wireframing that are available on just about every office computer. For the last three years, I have been using “Developer Tools” in Microsoft Office to not only draw the basic outlines of wireframes but also build functioning prototypes that simulate many of the features survey stakeholders want in the final product. Developer Tools allow you to add working data entry fields to your design, including open-ended text boxes, radio buttons, check boxes, and drop-down lists. You can also simulate the proposed functions of your prototype; for example, presenting specific error messages. In this presentation, I will describe how I used Microsoft PowerPoint and Excel in conjunction with Visual Basic for Applications (VBA) programming to design a wireframe version of a mobile-optimized data entry website for use in testing. Several other examples will be available upon request.

Evaluating the Impact of Pretesting and Redesign with Voluntary Feedback Questions
Jussi Rouhunkoski, Statistics Finland

Cognitive interviewing is arguably the most popular of questionnaire evaluation methods due to its inexpensiveness and the unique insights it provides into response process. Yet, the method has also received criticism, from the way the testing process is conducted to the results that it produces. A problem, which we have encountered is that the impact of pretesting is rarely followed up after the survey data is collected with the redesigned questionnaire. Sometimes the pretesting is repeated before the next round of data collection but it will only produce more qualitative data from a small sample and all too often, we do not have enough information to know if the redesign made the questionnaire better or worse – e.g. was the data quality improved or the response burden reduced?

One way to address this problem would be to utilize controlled experiments but they are expensive and difficult to implement in a proper manner. When looking for a more frugal and agile solution we were inspired by the Statistics Denmark that has included voluntary questions on burden and usability in their business survey questionnaires. We developed and included our feedback questions in redesigned web form of the survey on transportation of goods by road to evaluate the success of our solutions. As we did not have baseline for the comparison we utilized the results of the more comprehensive response burden measurement conducted earlier in the Statistics Finland. In our presentation we will discuss findings and experiences from this development project.

Session 3

Challenges in using cognitive interviewing with people with intellectual disabilities
Elena De Palma, Istat

The World Disability Report has stressed the importance not only to use the International Classification of Functioning, Disability and Health (ICF) as conceptual framework to improve quality data at the national and international level, but also to focus on subgroups of persons with disabilities. In its definition of persons with disabilities the U.N. Convention on the Rights of People with Disabilities has clearly included, among others, persons with mental and intellectual impairments that --in interaction with various barriers-- may hinder their full and effective participation in society on an equal basis with others. People with disabilities often have unique insights about their disability and their situation, but some types of disabilities may represent a challenge for researchers in developing survey questionnaires and conducting interviews. What about persons with intellectual/learning disabilities? Istat is working to develop questions to estimate the number of persons with mental and intellectual

disability in Italy. In this context, cognitive tests may be performed with persons with intellectual/learning disabilities. A questionnaire developed for the general population is not always appropriate for persons intellectually challenged. The literature underlines problems that may be faced by these individuals in participating in research activities. The presentation will address the main challenges related to conducting cognitive interviews with this target population, the main issues that should be taken into consideration and will try to suggest possible ways to overcome potential problems.

Multilingual, Multicultural, Multi-mode Testing of Questions for U.S. Census
Alisú Schoua-Glusberg, Research Support Services

As part of the research program for the population census in 2020, the U.S. Census Bureau is currently conducting a set of multilingual studies to refine and improve the census form in Spanish, Arabic, Chinese, Vietnamese, Korean, and Russian. A multi-mode approach to pretesting is being implemented. Expert review panels were assembled to weigh in on translation and cultural issues in the form and ancillary materials. Following their recommendations, revisions were made to the form in each language. Cognitive testing of the revised forms in each language is being carried out to identify patterns of interpretation of questions and terminology, as well as uncover cultural issues that may be impairing question equivalence across languages and target populations. Cognitive testing is being conducted in different administration modes: interviewer administered and self-administered, on paper and electronic devices. Focus groups will also be convened this summer to test acceptance and adequacy of messages for each of the target populations. This presentation will discuss the types of findings from each study component completed (expert panels in all languages, Spanish cognitive testing on paper and devices), the kinds of issues uncovered, and the different possible strategies we are adopting to solve the translation and/or cultural issues found. Strategies for resolving translation issues in the questions and materials tested will be presented. Future analysis plans with the additional languages will be discussed.

What cognitive interviews can tell about bias in cross-cultural survey research?
Jose-Luis Padilla, University of Granada

There is a broad consensus among survey researchers on key concepts behind any cross-cultural or lingual survey: Equivalence and Bias. Both concepts determine to what extent valid comparisons across different cultural and linguistic groups can be made. The aim of this paper is to show how cognitive interview findings can be used for identifying sources of construct, method and item bias, and how to link them to quantitative analysis within a mixed method research framework. We illustrate our proposal by presenting a CI study carried out within a mixed methods design aimed at assessing bias in Quality-of-Life (QoL) scales, used in large-scale quantitative studies. A total of 50 participants, 25 participants from Spain and 25 from the Netherlands responded to QoL scales, and then took part in the cognitive interviewing study. Evidence of construct, method and item bias associated with linguistic and contextual factors was obtained. We will also discuss suggestions on how to design and conduct cognitive interviewing within a mixed research framework in order to obtain a comprehensive view of bias in cross-cultural survey research.

Development and Resulting Data of a Sexual Identity Measure for the National Health Interview Survey
Kristen Miller, National Center for Health Statistics

For several decades, the lesbian, gay, bisexual, and transgender health movement has pushed for the collection of nationally representative health data of the LGBT population. Such data are required to achieve health equity, eliminate disparities, and improve the health of LGBT people. To this end, multiple health surveys have written their own ‘sexual orientation’ questions. The resulting data, however, has generated non-comparable estimates across surveys as well as high rates of missing data, particularly among low SES and Spanish speaking respondents. In reality, sexuality is a complex phenomenon that incorporates numerous, even contradictory, meanings, attitudes, and types of experiences which create a major challenge in developing a single measure that is both meaningful and accurate. The production of flawed data can lead to a mischaracterization of sexual minorities and, in the end, promote inadequate policy and health programs. It is critical, therefore, that a single quality measure be developed that most accurately collects data pertaining to this population. In 2013, the National Center for Health Statistics worked to develop and test a sexual identity question for the National Health Interview

Survey (NHIS) that would account for the problematic complexities surrounding a single sexual identity question. Both qualitative and quantitative methods were used to develop and test a final measure. This paper will describe the development process for the measure. Additionally, 2013 NHIS data will be presented, illustrating how estimates are improved by the new design.

Session 4

Where To? Identifying the Next Frontiers in Cognitive Interviewing Methodology **Paul Beatty, U.S. Census Bureau**

During its first few decades of use, methodologists made significant strides in establishing key parameters of cognitive interviewing, particularly with regard to data collection methods. More recently, methodologists have developed increasingly sophisticated frameworks for the analysis and interpretation of cognitive interviewing data. Given considerable advances in these areas, what are the key remaining challenges for methodologists working in this field? In this presentation, I suggest that we can maximize the effectiveness of cognitive interviewing by devoting energy to three key areas. The first is exploring possibilities for accumulating findings across individual projects. An early and oft-repeated criticism of the method is that findings regarding one questionnaire rarely inform subsequent evaluations. What are the prospects for deriving new or improved questionnaire design principles from multiple studies, and to what extent are we successfully utilizing results of earlier studies? The second area involves expressing the consequences of cognitive problems in more statistical terms. For example, to what extent does ignoring various cognitive problems decrease the precision or introduce bias into key estimates, and how much do these problems actually contribute to total survey error—and what sort of studies would help us to learn that? The third area involves increased focus on which aspects of our work are most (and least) valuable to our sponsors and collaborators. Some cognitive interview investigations focus on fixing “problems,” while others are primarily conducted to understand respondent interpretations. How do sponsors and other collaborators actually use these results, and are there ways to improve the utility of findings?

Evaluation and testing questionnaires in NSI of Spain **Víctor Balsa Criado, National Statistics Institute of Spain**

In year 2013 we launched a small team to carry out some little projects in the way of Evaluation and testing questionnaires. In year 2014, we did our first big project Pilot studies and quality improvement of the LFS and its modules, granted by EUROSTAT. We did 50 interviews and 3 focus group, specifically led to unemployed people, entrepreneurs and producers of goods. Despite de large number of interviews conducted (50), we found problems due to that was a multi-focused project. Some target question were specific for target groups that were extremely difficult to recruit. Upcoming studies:

1. Pretesting of the Labour Force Survey WEB forms. The questionnaire has been developed in a self software property of the NSI of Spain. This new software is not as functional as it is desirable, so it has several lack of flexibilities that avoid to have the optimal forms. We are interested in dealing with: - Usability of the web questionnaire - Rewording of original CATi CAPI questions - Self classification (NACE, ISCED, etc.) - The WEB channel it would be reserved for subsequent interviews, so what responses of the former interview must be pre-filled? , such as: - Family members information (sex, date of birth, etc.) - Date of the last contract, - Occupation, Establishment Activity (text), - Upper educational degree obtained (text)
2. EUROSTAT project. Study the effect in the LFS of reordering same key questions of the questionnaire.

Developing ONS Suite of Financial Surveys **Ruth James & Genevieve Groom, Office for National Statistics**

The aim of this session is to present the successes and challenges of ONS’ recent developments to its suite of questionnaires sent to the financial sector. Areas covered include ONS steps taken to:

- Develop an initial proving survey to establish a suitable foundation in terms of coverage, and make initial contact with specialist populations who have not been approached previously.

- Create and implement an agile approach to questionnaire development and testing so that complex changes can be delivered quickly.
- Directly involve customers and topic experts in the development process.
- Adopt an holistic approach, including developing a set of harmonised questions to enhance coherence between the questionnaires and to maximise efficiency within the development process.

Some suggested areas for discussion include:

- How best to engage with customers and topic experts, and harness their specialist knowledge in a way that enhances the testing process for complex questionnaires.
- The opportunities and risks offered by agile working practices.

Harmonising questions between surveys, whilst still tailoring developments to the specific needs of specialist industries.

Identifying Methods to Preserve and Utilize Methodological Knowledge within Organizations

Dawn V. Nelson, U.S. Census Bureau

For more than thirty years, survey methodologists at the Census Bureau have evaluated questions from a myriad of federal surveys. Although many final reports are available on the Census website, many more have been lost in filing cabinets and on computer directories. Our 'knowledge management' system needs an overhaul. And we suspect we are not alone. In this presentation I will discuss our current efforts to capture and organize results from our earlier question evaluation projects. I will also provide an overview of our plans for utilizing this knowledge in future question evaluation projects. I will end with questions for the audience about their organization's knowledge management systems for preserving survey question pretesting findings.

Electronic Questionnaires at Statistics Canada

Paul Kelly, Statistics Canada

Over the past 3 years, Statistics Canada has introduced electronic questionnaires as the primary mode of collection for many of the agency's surveys. Plans are in place to have most of the agency's surveys offering an electronic questionnaire within the next 3 years. As the focal point for questionnaire design and evaluation at Statistics Canada, the Questionnaire Design Resource Centre (QDRC) is responsible for planning and implementing cognitive and usability testing strategies for these new electronic questionnaires. This paper will present some of our recent experiences with electronic questionnaires as well as discuss some key findings and challenges.