

**2015 International Workshop on Comparative Survey Design and Implementation
City University London
Session Abstracts**

Techniques for Questionnaire Development

Joint Cognitive and Usability Testing of Translated Instruments: Identification of Translation Problems through Eye Tracking

Patricia Goerman, U.S. Census Bureau

Eye tracking is often used in usability testing to identify problematic areas of survey questions. Measures such as fixation count, fixation duration and visit count, where researchers keep track of the amounts of time respondents spend looking at particular survey items and terms are often used to assist in identifying problem areas in online surveys (Olmsted-Hawala, et al., 2014). Usability and cognitive testing of surveys are often carried out separately, but there is an increasing number of examples of joint cognitive and usability testing (Bergstrom, et al., 2013). There has been little (if any) research in which usability testing and in particular, eye tracking, has been used to identify problematic survey translations. This research aims to compare eye tracking findings across language versions of survey questionnaires. Through joint usability and cognitive testing of the source (English) and translated (Spanish) version of two online Decennial Census test survey instruments, we will seek to examine the extent to which differences in eye tracking data can help to identify translation problems. We will examine measures such as fixation count, fixation duration and visit count across languages to determine whether respondents looking at the survey questions in English and Spanish differ in where they focus on particular survey questions. In addition, we will examine whether findings from retrospective cognitive interview probes and the problematic wording identified in probing match up to locations where respondents spent more time gazing at particular terms, phrases and/or questions. While respondents are readily able to articulate some types of translation errors such as inappropriate word choices or sentence structures that are completely incomprehensible, certain types of errors may be less often mentioned by respondents, such as unexpected syntactic constructions or localization errors. This research will also investigate the types of translation errors that eye tracking data reveals and whether errors that respondents are unable to articulate during cognitive interviewing can be pinpointed via eye tracking during usability testing.

Analyzing Cognitive Interviews for Cross-National Studies

Jennifer Kelley, University of Michigan

Kristen Cibelli, University of Michigan

The expanding literature for cognitive interviewing illustrates the importance of conducting cognitive interviews to not only uncover various problems with survey questions (e.g. comprehension or retrieval errors), but to gain a deeper understanding of how survey questions function and if the questions or study constructs are understood as intended by different populations or population subgroups. For questionnaires that are developed for cross-national studies, understanding how the questions constructs perform in different cultural contexts

becomes even more crucial if one is aiming for comparability between the countries/cultures. Although there is an emerging literature on cognitive interviewing in the context of cross-cultural studies, the literature on how to analyze the data once collected is sparse. This presentation will review the current literature on cognitive interviewing analysis for cross-national studies, special considerations needed to analyze cross-national data, and a case study drawing on experience analyzing data from cognitive interviewing studies conducted in Nepal and China.

The Effect of Providing Think-Aloud Examples and Practice on Cognitive Interviewing in Nepal

Kristen Cibelli, University of Michigan
Jennifer Kelley, University of Michigan

Cognitive interviewing is most effective when respondents verbalize their thought process in response to the survey questions and interviewer probes. This process may be unfamiliar and difficult for some respondents, particularly in contexts outside of the United States and Europe where there is little survey tradition. Little guidance exists in the literature on how to apply cognitive interviewing in such contexts. This paper examines the effect of providing respondents with examples and practice think-alouds before the main cognitive interview.

The cognitive interviewing study was carried out in Nepal where experience with surveys and cognitive interviewing is rare. The study tested questions for an international household survey on disability. The Institute for Social and Environmental Research (ISER) in Chitwan Valley, Nepal and the University of Michigan conducted the cognitive interviews (n=40). Respondents were randomly assigned to a treatment or a control condition. In the treatment condition, interviewers demonstrated thinking aloud in response to an example question and then asked the respondent to practice thinking aloud to two additional practice questions. Respondents in the control condition did not receive examples or practice questions. Interviewers provided an introduction explaining the purpose of the interview and the cognitive interviewing process to respondents in both conditions. All respondents were asked debriefing questions on their cognitive interviewing experience (e.g. how easy/difficult to think aloud; how comfortable/uncomfortable).

We examine the effect of providing an example and practice thinking aloud on respondents' answers (e.g., the number of mentions unprompted, the number of codable responses, the amount of information learnt) and the results of the debriefing questions. We conclude with lessons-learned and suggestions for future research on the application of cognitive interviewing in cultural contexts outside of the United States and Europe.

Racial and Ethnic Identity of U.S. Born Children of Foreign-Born Parents: How do parents' proxy reports of race and ethnicity for U.S. born children differ from parents' own self-reported race and ethnicity?

Mikelyn Meyers, U.S. Census Bureau

Patricia Goerman, U.S. Census Bureau

Leticia Fernandez, U.S. Census Bureau

While there is a large body of literature that discusses the difficulties that many Hispanic respondents experience when they attempt to complete the race and ethnicity questions on U.S. government questionnaires, relatively little attention has been paid to the unique difficulties these respondents face when attempting to report the race and ethnicity of their children who were born in the U.S. In a small, bilingual usability and cognitive testing study done in 2014 for a census questionnaire (Goerman, Meyers, Simmons, forthcoming), the majority of respondents who emigrated from another country had difficulty answering the race and ethnicity questions for their children born in the U.S., and wrote terms like American, which are considered misreports. Besides the problem that foreign-born respondents encounter when wanting to report that their child was born in the U.S., there is also the more fundamental question of whether a U.S.-born child is considered to be of Hispanic origin by the parent. In addition to examining the results of the 2014 cognitive and usability study of a census questionnaire, we will report on upcoming cognitive and usability testing with both U.S.-born and foreign-born English and Spanish speakers in three modes (CAPI, internet, and paper). Through these studies, we will observe how respondents from various immigrant groups report the race and Hispanic origin of their children, and ask in-depth questions about their interpretation of terms such as Hispanic and American. These results will elucidate the issue of how foreign-born parents perceive the racial and ethnic identities of their children who are born in the U.S. in comparison to their own racial and ethnic identities. This paper will also address whether reporting the race and origin of their U.S.-born children is easier in the traditional two-question structure or in a combined question. Additionally, we will discuss the best way to align some respondents' desire to report that their child was born in the U.S. with the information that U.S. government agencies seek when asking these questions. Finally, this paper will discuss aligning respondents' perceptions of whether their child qualifies as Hispanic with U.S. government agencies' definition of the term. The question of how foreign-born parents report race and ethnicity for their U.S.-born children is one that merits ongoing research, particularly as the development and testing of new, combined race and ethnicity questions moves forward.

What Do Respondents Mean When They Indicate To Be Citizens Of The World? Using Probing Questions to Elucidate International Differences in Cosmopolitanism

Michael Braun, GESIS, Leibniz Institute for the Social Sciences

Dorothee Behr, GESIS, Leibniz Institute for the Social Sciences

Measurement of cosmopolitan attitudes has proved to be challenging. A direct measure targeting the identification of respondents (identity measure) has sometimes been found to be inferior to a composite measure where the stances of respondents towards a broad array of pertinent attitudes are taken into consideration. To explain unexpected findings, a mixed methods approach has been suggested to find out what lies behind this global identity, that is, what people have in mind when indicating to be citizens of the world (e.g. Pichler 2012). To the best of our knowledge, qualitative evidence for such an item has not yet been collected. We report results from a probing study where the closed question of the Eurobarometer (to what extent do you personally feel you are a citizen of the world) was followed by a category-selection probe (Please tell us

why you feel [to a great extent / somewhat / not really / not at all] that you are a citizen of the world). The data come from Web surveys conducted in Canada (English speaking only), Denmark, Germany, Hungary, Spain, and the U.S. in 2011. Respondents were drawn from nonprobability online panels. Any generalizations to the entire populations, therefore, have to be treated with extreme caution. However, we use an acceptable replication of the country patterns found in the Eurobarometer data (for the European countries, only) as a precondition for using the Web survey to shed light on the Eurobarometer data.

Mode

Effects of Survey Administration Modes on Results in Cross-National Research: The Case of the International Social Survey Programme (ISSP)

Theresa Zmijewska-Jedrzejczyk, Polish Academy of Sciences
Marcin W. Zieliski, Polish Academy of Sciences

In literature on survey methodology there is still a relative lack of clear evidence how can valid comparison be made when we use different data collection modes. One of a key problem in cross-national surveys concerns difficulty to separate mode effects from the real differences between countries. Using the International Social Survey Programme (ISSP) Religion Module we examine the effects of applying different modes. We used this module as it deals with sensitive issues, such as attitudes towards past, current religious practices and religious beliefs. For analysis we selected 35 countries where different methods were used: self-completed pen-and-paper forms and face-to-face interviewing. Results of our analysis lead to the conclusion that variation between countries partly comes from mode effects, even if other effects are controlled. We argue that as the data collected on sensitive topics and the results depend on the administration mode, researchers should be cautious of combining data and comparing results from different modes. Keywords: mode effects, cross-national comparisons, International Social Survey Programme (ISSP).

Using Administrative Data in Large Scale Surveys: The Example of the GGP

Tom Emery, NIDI, Netherlands Interdisciplinary Demographic Institute

The Generations and Gender Programme is a cross-national survey that has been conducted in 19 countries. Its primary focus is on families and relationships over the life course and has over 2,000 registered users. Whilst the main survey is conducted in face-to-face interviews, a number of countries have drawn on register and administrative data provided by national statistical agencies to help reduce the burden on respondents, increase the quality of the data and provide cost-efficiencies. The use of such data has been an overwhelming success in many regards but there are significant obstacles to releasing such data alongside the survey data. On the one hand, national statistical offices hold concerns that including such data would breach anonymity and intended use regulations. On the other hand, researchers are reluctant to adopt techniques that invalidate or severely complicate the scientific inferences of such data. Such techniques include treating data linking as a missing data problem and the creation of synthetic data. After much trial and error, the GGP has developed a series of solutions that appear to satisfy the concerns of both the data providers and data users, thus enabling administrative and registry data to be distributed to data users. In this presentation, we discuss how the GGP has adapted to this problem and seeks to provide data that meets the requirements of all interested parties. Specifically attention is given to the principal of data fogging where data is purposefully

degraded to protect individual 's identities whilst maintaining a level quality necessary for scientific research. For example, given that the GGP 's scientific focus is on the relative rather than absolute timing of demographic events we give details on an approach which shifts an individual 's entire life-course by a period 3-6 months. This approach enables registers to avoid revealing the birth month of respondents and their children whilst maintaining the scientific value of the data. To examine whether this approach would be considered valid during peer review, a series of journal editors and senior demographers were consulted and their concerns noted and incorporated within the approach. The result is an approach to administrative and registry data use that satisfies all interested parties without the need for complex statistical procedures.

Measurement

Current Consensus about How to Avoid Bias in Cross-Cultural Research: A Case in Favor Of Addressing Equivalence as a Validity Problem

Jose-Luis Padilla, University of Granada

Isabel Benitez, University of Granada

There is a broad consensus among professionals on key concepts behind any cross-cultural or lingual testing or survey: Equivalence and Bias. Both determine to what extent valid comparisons across different cultural and linguistic groups can be made. Professional guidelines, standards, and recommendations reflect that consensus but differ in how to address equivalence and the extent to which they incorporate the last conceptual and methodological developments. The aim of this paper is twofold: first, to review how the most well-known guidelines and professional standards address bias, equivalence and validity, and secondly, to illustrate an integrated approach to bias as validity issue. We pay special attention to the last edition of the International Test Commission Guidelines (ITC, 2012), and the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). The comprehensive approach to bias is illustrated by an application of the General Health Questionnaire (GHQ-12) among native-born Spanish and immigrants in the National Spanish Health Survey. A total of 1251 immigrants from Morocco, Ecuador, Romania and Colombia as well as Spaniards were involved. Construct, method, and items bias were evaluated, using analyses of measurement invariance, response styles, and Differential Item Functioning (DIF), to address threats to validity of GHQ-12 scores comparative interpretations. The two-factor model based on the positive and negative wording of items achieved the best fit. DIF analyses indicated items with bias, while response styles analyses identified differential levels of acquiescence across groups. Finally, we make a case for including in professional standards and guidelines, explicit recommendations about resorting to mixed-method studies to uncover sources of bias as threats to validity in cross-cultural or lingual testing and survey.

Can Language of Survey Administration Affect Responses to Survey Questions?

Emilia Peytcheva, RTI International

Marjorie Hinsdale, RTI International

U.S. surveys of minorities and immigrants allow respondents to answer in the language of their choice. Psychological and linguistic research suggests that the language of administration may influence responses through its impact on various cognitive processes and through the cultural frame that a language brings to the conversational context. However, the extent to which the

language of survey administration affects survey responding is currently unknown. We use data from New Immigrant Survey (NIS), where statistical methods such as propensity score modeling allow us to estimate the effect of language when respondents self-select themselves into a language of interview. Analyses of the baseline interview show evidence that the language of survey administration has an effect on how Hispanic bicultural-bilinguals respond to survey questions that prime culture-related norms. For example, consistent with more conservative alcohol norms and attitudes among Hispanics relative to Whites, we found significantly lower reports of alcohol consumption for Hispanic bilinguals interviewed in Spanish compared to those interviewed in English. This study repeats the baseline interview analyses using data from the follow-up interview and examines response patterns within the same respondents interviewed in different languages in the two waves of data collection. The implication of such results for current national surveys that sample ethnic minorities and immigrants is that language assignment should be informed by the goals of the survey questions and leaving the choice of language to a bilingual bicultural respondent may affect data quality. Ideally, researchers would be able to inform language assignment among bilinguals based on knowledge about domains where cultural differences and the direction of such differences may be expected, or depending on what respondent cultural identity is of interest. If such knowledge is not available, random assignment of bilingual respondents to a language would at least allow estimation of language effects.

Midpoint Placement in the Left-right Scale in a Cross-national Perspective

Cornelia Zuell, GESIS, Leibniz-Institute for the Social Sciences

Evi Scholz, GESIS, Leibniz-Institute for the Social Sciences

Left-right self-placement on a uni-dimensional scale is one of the standard questions in many social and political surveys to measure respondents' ideological orientation in a minimalist way. Although the left-right scale is a standard question, the design of the scale is not standardized across surveys. One aspect of scale design is the offer of a midpoint. This paper is about design effects on central left-right scale placement in a cross-national context. How do respondents answer if there is no true midpoint? Do respondents who want to express a middle position, in the case of a 10-point scale, use scale middle categories to substitute a true midpoint? Offering a midpoint is much debated among political and survey researchers and quite often, a midpoint serves as a hidden don't know or a missing attitude. Does non-response increase when non-attitudes cannot be expressed by choosing the neutral midpoint to hide non-response? If middle categories in the 10-point scale work as a substitute for a true midpoint in the 11-point scale, non-response will not differ. Methodological research using a German SOEP experiment found that the 10-point and the 11-point scale are similar in non-response and perform well with regard to reliability but that the 11-point scale seems to work better than the 10-point scale in terms of substantive analyses. Are these German findings robust across countries? We tested the effects in a split-half experiment where either a 10-point scale or an 11-point scale was asked in an experimental online survey fielded in Canada, Denmark, Germany, Hungary, Spain, and the U.S. Preliminary results seem to confirm the idea that respondents who are in favor of choosing a scale middle find a virtual center in the 10-point left-right scale and non-response does not show significant differences between 10-point and 11-point scale.

Measuring Acquiescence in Different Cultures: Results of Experiments with Translations and Scale Types

Anna Andreenkova, CESSI, Institute for Comparative Social Research

Acquiescence is well known problem of response style when respondents tend to answer questions in the survey consistently using particular categories of approval/ agreement/ in favor/ support regardless of the question content. In comparative context the differences in level of acquiescence between cultures were often suspected, but not often measured. We present the results of the experiments to measure the size of this effect in different cultures and languages: two Slavic (Russia and Ukraine), three Turkish family language countries (Kazakhstan, Kyrgyzstan and Tajikistan) and three countries in South Caucasus using the dataset Comparative Post Soviet Social Trends. Three statements on different subjects (general political attitude statement, statement on ethnicity relations and statement on gender roles) were asked to respondents in two opposite forms with the same response scale in the beginning and in the end of the interview. The prior hypothesis based on previous findings is that acquiescence effect be stronger in Asian cultures rather than in Slavic cultures. Another hypothesis is that the acquiescence effect is stable regardless of the topic of the statement. The first hypothesis was confirmed, the acquiescence effect was consistently stronger in three Asian countries than in Slavic countries. Ethnic Russians - substantial minority in all surveyed countries - showed similar level of acquiescence effect in all countries. So the culture rather than country plays the major role in producing acquiescence. The hypothesis that acquiescence response style is constant across different topics was in opposite not confirmed. In all countries the effect was the strongest in question on ethnic relations and the smallest on question on gender roles. We propose that the strength of acquiescence effect can be related to the strength of the opinion. It is strongest when public opinion on particular issue is not consistently formed. We did not find substantial gender or age differences in acquiescence in any country. We found that the language of interview plays important role in relation to acquiescence even taking into account the ethnicity. Respondents of the same ethnic origin but interviewed in different languages (Russian or native) showed different level of acquiescence. It can be due to the translation issues or to stronger integration to national culture of those who select native language for the interview. We will present our further proposal on the work with the translation of approval/ agreement/ in favor/ support scale to minimize this response style.

Quality

Prevailing Problems in the Design and Implementation of Comparative Surveys

Gelaye Worku, Stockholm University

Lars Lyberg, Stockholm University

The importance of having a solid infrastructure in place for the design and implementation of comparative surveys cannot be emphasized enough. A solid infrastructure implies input harmonization, a central team making sure that standards and requirements are reasonably met, and that there is an element of continuous improvement of the survey. Most comparative surveys do not have these ingredients in place. In some surveys everything is allowed to vary. The only thing that exists is a source questionnaire and a project leader and countries are asked to conduct the survey. In other surveys there is output harmonization, which in effect means that the project leader expects certain outputs in terms of deliverables but the countries can conduct the survey the way they want. A third variant is input harmonization without a central team, which usually results in some deviations from requirements.

In this paper we will give examples of how quality issues are treated within these four scenarios. The main issue is to achieve functional comparability but this is not possible within each of these scenarios. As a matter of fact the only scenario that is productive in this sense is the first with input harmonization, a central team and continuous improvement. Still there are problems with countries not adhering to specifications, especially some of them including translation, adaptation and pretesting, interviewer monitoring, and quality control. Root causes include an unwillingness to deviate from local procedures, lack of methodological capacity, and a wish to save the expense. We will end with some recommendations on how to improve the situation.

Resources for Conducting Cross-national Survey Research

Tom Smith, NORC at the University of Chicago

Cross-national survey research is a complex and challenging endeavor. Designing a reliable and valid comparative study is notably more difficult than constructing a single monolingual/monocultural survey. Besides the whole array of components that encompass total-survey error in one survey, there is a parallel set of comparison errors that apply in cross-national surveys. Likewise, comparative analysis is complicated by these comparison errors and even well-conducted comparative surveys may produce erroneous results due to failing to achieve functional equivalence at the design, collection, and/or analysis stages. Even locating cross-national data for secondary analysis can be a challenge. This paper describes some valuable resources, especially those available on the web, which can assist in conducting cross-national survey research.

A Summary of the Methodological Criticism of PISA

Can Tongur, Statistics Sweden

Lars Lyberg, Stockholm University

The outcomes of PISA, the Program for International Student Assessment, have triggered methodological criticism of different kinds. The main concern is that the results might be used as a basis for educational reforms. A sample of criticisms includes the following:

- The general survey conditions differ across countries
- The age 15 cohorts differ in actual years in school and other aspects across countries
- Very strange response patterns have been discovered in some countries and organizational fabrication of data is possible
- The suitability of the Rasch model has been questioned for this type of psychometric study
- Bad translations hurt equivalence
- Small deviations from standards can result in dramatic differences in country rankings
- Some countries train their students in PISA testing while others downplay the importance of this study.

We will discuss these and other criticisms and comment especially on the case of Sweden, since Sweden has, according to the latest PISA report from 2012, shown the greatest decline in student attainment among participating countries.

Total Survey Error (TSE)

Cross-national Research in a Total-Survey Error Perspective

Tom Smith, NORC at the University of Chicago

Total survey error (TSE) is a very valuable paradigm for describing and improving surveys, but it can be improved. One key limitation is that TSE was formulated to apply to a single, standalone survey. Yet most survey research combines and compares surveys. TSE can be extended to cover these multi-survey utilizations. TSE needs to be thought of as heavily involving the interaction of error components and the concept of comparison error should be used to extend TSE to cover multiple-surveys including trend analysis, comparative studies, and longitudinal panels. This extension of TSE will greatly improve the design of multi-surveys in general and of comparative (i.e. cross-national/cross-cultural) surveys in particular. Likewise, using TSE can greatly advance the analysis of comparative data by using it to assess and adjust for difference in the error structure across surveys. A comprehensive TSE typology should be used whenever comparative studies are designed and also whenever secondary analysis of comparative studies is carried out. In particular strict application of the TSE paradigm can help to achieve the goal of functional equivalence cross-nationally/culturally. Minimizing TSE is an important goal in survey research in general and is especially valuable for comparative survey research and the TSE paradigm should be used as both an applied application and a research agenda to achieve that goal. Extensive examples from the ISSP and ESS will be used to demonstrate this approach.

Comparing Concept Measurements across Contexts: The Relative Relevance of Wording and Translations

Michèle Ernst Stahli, FORS, Switzerland

Oliver Lipps, FORS, Switzerland

Caroline Vandenplas, FORS, Switzerland

Concepts are measured by different designs, with varying wordings, translations and survey settings. But how far are similar questions comparable through surveys? And how far can single design effects be identified? What makes measurements comparable, what incomparable? What matters more: an identical wording (respectively ‘good translation’) or an identical question format, or other surveys design aspects such as the interview mode, the question context, the way respondents are recruited, etc.? In this presentation we will first compare through five general-population surveys conducted in Switzerland by FORS in 2010 and 2011 (ESS, MOSAiCH-ISSP, SHP-Swiss Household Panel and Selects-the Swiss Election Study) the mean values of two often asked and analyzed items: interest in politics and satisfaction with democracy. The questions differ slightly in wording, but also in the number of answer categories. Moreover, there are important differences on other design features such as cross-sectional versus longitudinal, contact procedures, interview mode, main topic, etc. The results show large differences between the surveys. A trial to mitigate selection bias (coverage and non response) through socio-demographic post-stratification weights reveal to have no or almost no effect on these differences. Most conventional attempts to explain the differences, such as the social desirability effect in interviewer-driven surveys, substantially fail in this non-experimental study. We therefore hypothesize that unmeasured factors may significantly influence the outcomes, such as the way the survey is ‘sold’ to the respondents, the place of the question in the questionnaire, the style and content of the previous questions, etc. (see Vandenplas and Lipps, 2014). In order to isolate the wording and answer scale effects, we repeated the SHP and Selects versions at the end

of the ESS 2014 which is on the field right now. With this test-retest experiment, we will be able to take a step further in trying to identify and disentangle the different sources of the differences. With this contribution, we aim at widening the discussion about comparability beyond translations to additional survey design features. We hope this will give a new impulse to the ‘ask-the-same-question’ versus ‘ask-different-question’ debate.

Vandenplas, C. & Lipps, O. (2014). Robustness of items within and across surveys. FORS Working Paper Series, paper 2014-3. Lausanne: FORS. (ISSN 1663-523x (online); http://forscenter.ch/wp-content/uploads/2014/06/FORS_WPS_2014-03_Vandenplas.pdf)

The Cross-Country Approximate Comparability of Human Values in the European Social Survey Tested By a Bayesian Approach

Jan Cieciuch, University of Zurich, Switzerland

Eldad Davidov, University of Zurich, Switzerland

Peter Schmidt, University of Giessen, Germany

Renè Algesheimer, University of Zurich, Switzerland

The 21-items Portrait Value Questionnaire (PVQ-21) to measure human values is included in all rounds of the European Social Survey (ESS). The collected data provide a unique possibility to explore cross-country differences in value priorities. The precondition of such analysis is measurement invariance of the scale across countries (Vandenberg, Lance, 2000). Measurement invariance implies that the parameters of a measurement model (factor loadings and intercepts) are equal across groups and differences found in means are due to real differences across countries and not due to methodological artefacts. A common way to test for measurement invariance is to impose the (exact) equality constraints on these parameters and inspect the fit of the constrained model. Previous studies of human values using this scale were rather disappointing. It turned out that scalar measurement invariance cannot be established; this precludes any meaningful cross-country comparisons (e.g., Davidov, Schmidt, Schwartz, 2008; Davidov 2008, 2010). Thus the question arose whether this disappointing result is due to the problem of the data or due to the method used to establish measurement invariance which may be too strict. Muthén and Asparouhov (2013) recently proposed a new approach to test for measurement invariance that weakens the strong assumption and proposes to test for approximate (rather than exact) measurement invariance. In the new approach one can permit small differences between parameters, otherwise constrained to be exactly equal in the classical exact approach. We present the results of the approximate measurement invariance test of the PVQ-21 items across countries participated in all six ESS rounds, compare them to the results obtained in the exact approach, and discuss problems connected with the method for the evaluation of models in approximate measurement invariance testing.

Interviewer Effects

Interviewer Effects: Current Issues and Approaches

Stephanie Chardoul, University of Michigan

Kristen Cibelli, University of Michigan

Beth-Ellen Pennell, University of Michigan

Interviewers play a number of essential roles on surveys, but also introduce the risk of error that could affect survey results (Groves et al, 2009). These errors consist of both variance (reflecting estimate instability) and bias (systematic deviations). This presentation will review the findings related to sources of interviewer bias and variance, and discuss the impact of these effects on survey precision (O’Muirheartaigh and Campanelli, 1998). We also will review the challenges of measuring interviewer effects, and techniques to minimize the impact – including interviewer training and supervision, quality assurance, and consideration of interview mode. And, some of the counter-pressures to minimizing interviewer effects, such as questionnaire development, technical and cultural barriers, and other conflicting goals within the survey project. Finally, we will review examples of approaches we have applied on projects here at the Survey Research Center.

Training Interviewers for PIAAC in Germany

Anouk Zabal, GESIS, Leibniz-Institute for the Social Sciences

PIAAC (Programme for the International Assessment of Adult Competencies) is an international large-scale OECD study that measures key adult competencies (literacy, numeracy, and problem solving in technology-rich environments) across the participating countries. The PIAAC survey implemented a complex design and also established elaborate measures of quality assurance and control to produce comparable, high-quality data. The PIAAC interview consisted of a background questionnaire administered as CAPI, and a cognitive assessment which was per default computer-based, with a paper-based administration for those not capable or willing to work on the assessment tasks on the computer; respondents worked on the cognitive tasks by themselves and at their own pace. To prepare the interviewers for the administration of the PIAAC interview, and to ensure that interviewers worked according to the same survey protocols and procedures across countries, the international PIAAC standards required that countries carry out extensive training (ranging from 3-5 days depending on interviewers' experience). The international Consortium prepared comprehensive training materials and trained the countries following a train-the-trainer model (similar to the approach followed by the SHARE survey, see Alcser & Benson, 2005). Germany participated in the first round of PIAAC (2008-2013) with the main data collection taking place in 2011/2012. All interviewers were extensively trained prior to fieldwork start. Training included a sessions on (a) gaining respondent cooperation, (b) the PIAAC survey protocols for the administration of the background questionnaire and the assessment, (c) general information on PIAAC (including contacting, administrative procedures, and quality control), and (d) practice interviews. This presentation will provide more details on the German interviewer training, and will conclude with an evaluation of the value of interviewer training both for PIAAC in Germany as well as potentially for other surveys.

SHARE's Train-The-Trainer program

Frederic Malter, Munich Center for the Economics of Aging
Max Planck, Institute for Social Law and Social Policy

Since the inception of SHARE in 2004, the Train-The Trainer (TTT) program has been integral in achieving one of the philosophical pillars of SHARE, ex-ante harmonization of procedures. Initially crafted by colleagues from SRC at the University of Michigan, the TTT has been continuously developed further to accommodate three constantly evolving aspects of SHARE: first, the electronic contact protocol Sample Management System (SMS) and CAPI software are constantly improved over waves and necessitate the introduction of new training content. An example would be new fieldwork disposition codes and its meaning and how these codes are meant to be used by interviewers. Likewise, new survey items are introduced every wave and need extensive training. Third, modalities of the training have been improved to be more up to current empirical evidence on adult learning. For example, a lot of training modules now contain more engaging modalities such as role play or small group exercises rather than straight-up presenting slides in every session. The basic pedagogic approach of the TTT, however, has been unchanged for all waves of SHARE: the TTT is the template for national training sessions in the participating countries. Survey agencies have to translate the slides and implement the training the way it was done at the TTT, including the teaching modalities. This is essential for cross-national harmonization of procedures. In order to assess how well this implementation was accomplished, we changed the SHARE model contract in two significant ways: we made the delivery of training slides in national language a contractual deliverable. In wave 4, we took the effort and investigated all training slides in all national languages if they represented the content of the TTT. This was published in the wave 4 SHARE compliance Profiles which can be accessed freely online: http://share-dev.mpisoc.mpg.de/fileadmin/pdf_documentation/Method_FRB_FINAL.pdf. The second change was the introduction of yet another deliverable, a standardized report on the national training sessions to be completed by the respective university teams after attending the national training sessions. In my presentation at the CSDI workshop 2015, I will talk about the general TTT philosophy, our approach to evaluating it and developing new content. I will present selected, brand new training content of wave 6 and talk about my experience with being the lead trainer at the TTT. My goal is trying to give participants of the CSDI workshop a feel for what it is like to attend the TTT.

Using Paradata to Monitoring Interviewers' Behavior: A Case Study from a National Household Survey in Ghana

Jennie Williams, University of Michigan

Kyle Kwaiser, University of Michigan

Yu-chieh (Jay) Lin, University of Michigan

Interviewers can be an important source of error in face-to-face surveys. To reduce this error source, survey organizations have used different quality control procedures including calling back respondents, audio-recording the interview, and accompanying interviewer and observing the interview process. These procedures have been recently supplemented by the increased use of paradata to achieve greater efficiency and improve data quality. This presentation will focus on using paradata to monitor interviewer's behavior in a national household survey in Ghana.

EGC-ISSER Socioeconomic Panel Study is the first face-to face survey using a computer assisted personal interviewing (CAPI) mode covering all geographical regions of Ghana. Since interviews are not recorded and this study utilizes unique team management, travel structure and

instrument design which puts the onus of instrument navigation on the interviewer, paradata is heavily relied on to monitor interviewers' behaviors. In this presentation we will demonstrate unique interviewer quality control metrics derived from keystroke paradata at the individual and team levels, and interviewer navigation through parallel block instrument design. We identify and analyze a suite of interviewer, household, and geographic characteristics to determine if interviewer behavior and data quality can be predicted by overall team behavior.

Interviewer Effects: Gender, Islamic Hijab, and Respondents' Sociopolitical and Cultural Attitudes in a Nationally Representative Survey in Tunisia and Beyond

Julie de Jong, University of Michigan

Kristen Cibelli, University of Michigan

Zeina Mneimneh, University of Michigan

Mansoor Moaddel, University of Michigan

Research has shown that interviewers can have important effects on respondent answers, particularly when questions are related to interviewers' characteristics such as race and gender. As social surveys become more prevalent in the Middle East and North Africa, concerns have emerged about potential bias introduced by interviewer gender and Islamic Hijab on related survey items. A small but growing literature has begun to explore these issues. For example, studies found that interviewers wearing the hijab – Islamic headscarf – elicited higher reports of religiosity and adherence to Islamic cultural norms in Egypt (Blaydes and Gillum, 2013) and that interviewers wearing Islamic, rather than secular symbols received higher reports of religiosity from respondents in Turkey (Koker, 2009). Benstead (2010) also found that female interviewers wearing the hijab affected gender-related attitudes among respondents examined the effects of interviewer gender and hijab on responses and data quality inform a nationally representative, partially randomized, survey in Morocco. She found an effect for female interviewers wearing the hijab on gender-related attitudes among respondents.

This presentation draws first on a nationally representative survey among Tunisian adults who were interviewed face-to-face in the spring of 2013. Respondents were asked about a wide variety of religious, cultural, social, and political issues. The survey also recorded the characteristics of the interviewers (including religious wardrobe) as well as their own religious, cultural and political attitudes. Using this dataset we investigate the effect of interviewer gender, hijab, and interviewer attitudes on respondent attitudes toward gender equality, religious beliefs, political values, and morality. This investigation extends the literature by comparing the effect of interviewers' visual cues --hijab or no hijab -- to that of reported attitudes, which may be communicated through more subtle cues during the interview, and their interaction on respondent reports. Multi-level modeling will be used to measure the magnitude of between-interviewer variance on key attitudinal measures and to investigate what interviewer-level characteristics explain this variation.

After examining interviewer effects in Tunisia, we then take a broad view of the differential effects of interviewer gender and veil status on respondents' attitudes in surveys in several Middle Eastern countries using data collected in a cross-national survey. The comparative analysis explores the extent to which there is a pattern in interviewer intraclass correlations that corresponds to characteristics of study countries in general, and whether gender and veil status in particular have greater effects in certain countries and on particular survey topics.

Interviewer effects in the European Social Surveys

Koen Beullens, Centre for Sociological Research, KU Leuven, Belgium

Geert Loosveldt, Centre for Sociological Research, KU Leuven, Belgium

Since the ESS is a face-to-face surveys, interviewers are important agents in the data collection process. Both through their interactions at the doorstep, as well as during the actual interviewing, interviewers can cause effects that may deteriorate the quality of the data. We will provide an overview of our work regarding interviewer effects in the ESS. This comprises the measurement of interviewer effects, how they effect the statistical power of a sample, how interviewer effects can be discerned from area effects and discuss possible strategies to combat interviewer effects during the fieldwork regarding the selection and training of interviewers.

Translation

Selecting and Recruiting Translators for Survey Translation Teams

Alisú Schoua-Glusberg, Research Support Services

How to look for and select translators for Survey Translation Teams is often a topic included in presentations and courses on managing survey translation efforts. In addition to the desired qualifications, experience and background, other factors are important to consider when recruiting a team member. This presentation will discuss the characteristics of the desired candidates but will also present a concrete case in which we had to find a new member for our translation team, which will include concrete examples and lessons learned.

On the Translatability Assessment

Steve Dept, cApStAn Linguistic Quality Control

Andrea Ferrari, cApStAn Linguistic Quality Control

Several authors have emphasized the need to include upstream quality assurance in translation and adaptation designs for multilingual surveys, with a view to reducing reliance on downstream quality control. Along these lines, the authors have developed a conceptual framework for a Translatability Assessment on draft questionnaire items, to be performed before they are finalized and sent to translation/adaptation. The Translatability Assessment was applied in a dozen multilingual surveys, and the results are promising.

Near-final draft versions of items were submitted to a pool of experienced linguists covering the broadest possible range of target language groups. These trained linguists produced draft translations of those items: these helped them identify and describe the hurdles that translators would have faced if no pre-emptive action had been taken.

A set of 13 translatability categories was used to report on the potential translation, adaptation and cultural issues identified. A consolidated translatability report was produced with the help of a senior linguist. For well-documented issues, either a translation note was proposed e.g. to clarify a given term, or alternative wording was proposed, with a view to facilitating localization without loss of meaning.

Item writers used this translatability report to eliminate ambiguities, e.g. Anglo-Saxon idiosyncrasies that may be difficult to render in certain languages, double-barreled questions, to circumvent cultural issues or to avoid unnecessary complexity. In a nutshell, a collective attempt was made to fine tune the initial version of the items so that it became a (more) translatable source version.

The paper presents examples and outcomes of translatability assessments performed on a selection of five multilingual surveys; explains the conceptual framework within which the linguists worked; illustrates the interaction between linguists and item writers; and presents leads for further research, including automated source optimization.

Zooming In On Translation Documentation - A Case Study from the ESENER Survey

Dorothee Behr, GESIS

Steve Dept, cApstAn, Brussels

Elica Krajc̆eva, cApStAn, Brussels

The presentation (Behr, Dept, & Krajc̆eva) looks into translation documentation from the European Survey of Enterprises on New and Emerging Risks. For 47 national versions, the entire translation process (source text and related documentation, multiple translation and assessment steps, including comments) was thoroughly and consistently documented. The authors look at the documentation from different angles: input and output documentation; metadata and paradata; quality assurance and user transparency; translation teams, project managers, and researchers. The presentation zooms in on selected aspects of what will eventually become a book chapter prepared for the second 3MC conference in 2016.

Data Collection Challenges: Sampling in Developing Countries

List and GO! Building a Sample for a Household Survey in Kenya

Sarah Hughes, Mathematica Policy Research

This presentation will describe her experiences on a project in Kenya in which sampling was carried out concurrently with data collection. In essence, the team was able to successfully create a 150,000 household sample frame, implement an in-field sampling strategy, and collect data concurrently. She will focus on the use of census maps and a tablet-based listing application to carry out segmenting, data collection in informal housing areas (shanty/slum), use of GPS and mobile technology, and the complexities of carrying out daily sampling.

Sample Development and Sampling on a Rural Health Survey in Nicaragua

Lauren Doerr, NORC at the University of Chicago

The Nicaraguan Early Childhood Development evaluation baseline survey is an impact evaluation of early childhood development interventions in marginalized regions throughout Nicaragua, conducted for the InterAmerican Development Bank. Doerr is the project director of this survey, realized by NORC in 2013-2014. This talk will discuss how a sample frame and random selection were implemented for this project in Nicaragua, a country where there are virtually no street names, along with other methodological and sampling challenges.

The Gods must be Crazy: Implementing a Longitudinal Study among Southern African Nomads

Kareem Kysia, NORC at the University of Chicago

In this presentation Mr. Kysia will describe his experiences in tracking nomadic and semi-nomadic groups over a 3 year period in Southern Africa with a focus on tracking solutions and response rates. The data being presented were part of the MCA Namibia Indigenous Natural Products Household and Organizational Survey which collected information as part of an impact evaluation of the INP sector in Namibia. The survey focused on increasing rural income among the Himba and Saan populations through better indigenous natural product collection and processing. The focus of the presentation will be on challenges associated with sampling from nomadic populations and tracing said populations over time.

Tools for Cross-National Surveys

Measuring Occupations in the Field: Some Developments in (Semi-)automatic Coding Tools

Eric Harrison, ESS ERIC HQ, City University London

The capture and accurate coding of open text information remains a formidable challenge for any survey, but in cross-national context these problems are multiplied in their complexity. This paper reports on a three year project, part of which was concerned to develop automatic and semi-automatic coding tools which could be applied to the measurement of occupation cross-nationally. Two approaches are presented. Firstly an extension of the existing package CASCO from English into eight languages, primarily for use on very large datasets and for back-office post-coding of information collected in the field. Secondly the development of similar text matching approaches to CAPI and CAWI applications for more accurate and speedy coding of job titles in the field or in self-administered online surveys.

The DASISH Translation Management Tool (TMT)

Maurice Martens, CentERdata

Brita Dorer, GESIS-Leibniz Institute for the Social Sciences

The DASISH Translation Management Tool (TMT) is an online service for supporting translation processes for large multilingual surveys. It has been available since 2004 for the renowned Survey of Health, Ageing and Retirement in Europe (SHARE), but in order to comply with other software developed in the DASISH project it had to be adapted to support the European Social Survey (ESS) and ideally other surveys as well. This called for the use of DDI. Challenges were including multiple surveys, versioning, referencing identifying languages and locales, and attribute pairs defining status of the overall translation process. In addition, for its questionnaire translations, the ESS follows the so-called TRAPD approach, consisting in a multi-step and multi-disciplinary translation strategy, so the TMT should be tailored to support this scheme as well. In the 2013 CSDI workshop the plans for the adaptation of the DASISH Translation Management Tool were introduced. Now we will see where that has led us.

New Tools for Complex Surveys: The DASISH Questionnaire Design Documentation Tool and Question Variable Data Base

Knut Kalgraff Skjåk, NSD - Norwegian Social Science Data Services, Norway

Rory Fitzgerald, ESS ERIC HQ, City University London, UK

Yvette Prestage, ESS ERIC HQ, City University London, UK

Two new tools to enhance the process of documenting questionnaire design and development and survey variables are currently being developed under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project. The primary aim of the questionnaire design documentation tool (QDDT) and the question variable data base (QVDB) is to assist large-scale survey projects in the processes related to questionnaire design and development, as well as in data and metadata production and curation, archiving and dissemination. Secondly, the tools aim to serve researchers and students in exploring metadata from existing projects, or in designing new research. The tools have been developed as an integrated toolset, designed to facilitate easy reuse of metadata components at the different stages and sub-stages of the design and production processes. They take advantage of the gains in efficiency, precision and not at least accessibility that can be derived from better use of technology. Initially planned for the European Social Survey (ESS), which is a complex international biennial cross-sectional survey, the tools are designed to serve other projects, to interoperate with other systems and tools, and to act as reusable models for other tools. Both the QDDT and QVDB are modelled based on the recognised metadata standard DDI-Lifecycle. This presentation gives an introduction to each of the tools and their development.

Metadata across the Survey Lifecycle: A Common Metadata Understanding for the Three DASISH Survey Tools

Hilde Orten NSD – Norwegian Social Science Data Services

Knut Kalgraff Skjåk NSD – Norwegian Social Science Data Services

Taina Jääskeläinen FSD - Finnish Social Science Data Archive

Brita Dorer, GESIS – Leibniz Institute for the Social Sciences

Joachim Wackerow, Consultant

The Questionnaire Design Documentation Tool (QDDT), the Translation Management Tool (TMT) and the Question Variable Database (QVDB) are three tools for questions and variables under development under the Data Infrastructure for the Social Sciences and the Humanities (DASISH) project. The three tools are developed as individual tools that should be able to communicate with each other, so that metadata developed at earlier stages in the survey lifecycle can be reused at the later steps. To facilitate this, work has been carried out to arrive at a common metadata model for the three tools, based on Data Documentation Initiative (DDI) metadata standard. The current presentation gives an introduction to the DDI and its usage in the work towards a common metadata understanding for the three DASISH tools. Topics like metadata flow and reuse, versioning, administrative ownership of metadata and more will be discussed.

Developing a Mobile App for Cross-National Survey Interviewers

Yvette Prestage, ESS ERIC HQ, City University London

With different survey agencies using different systems and methods to monitor fieldwork, the availability and quality of fieldwork progress reports varies considerably. For cross national surveys this makes it difficult to obtain a clear, timely understanding of fieldwork progress. It also means there are barriers to understanding non-response during fieldwork, which means preventative action is often impossible. As part of an EU-funded project, researchers from the ESS and SHARE have developed a prototype mobile application (app) which can be installed on any hand-held device. The app draws on SHARE 's existing computerised sample management system and the paper contact forms used by ESS interviewers, and intends to overcome limitations from both surveys by providing real-time data during fieldwork instead of after fieldwork (ESS) and on a hand held devices instead of laptops (SHARE). The app will enable interviewers to easily and electronically manage their caseload, capturing all of the information currently recorded in the ESS contact forms. By allowing interviewers to log their contact attempts in real time, the app will aid central field supervision, control and monitoring. It will also allow all countries to consistently provide information on a regular basis during fieldwork, whilst retaining the level of detail required. This would provide a more comprehensive understanding of fieldwork progress as it happens, facilitating more timely analysis of the data and the possibility of adopting a responsive design. Whilst it was originally designed with SHARE and the ESS in mind, the benefits of a mobile app that can assist interviewers in managing their workload whilst simultaneously providing agencies with timely information about fieldwork progress could prove invaluable to a range of stakeholders. And, as the app is compatible with smartphones, it is also possible to ensure it makes use of the features often inbuilt in smartphones such as GPS data; call and text functions, and sound recording. Our presentation will describe the development of the prototype app, highlighting successes, as well as problems encountered and how these were overcome. It will also describe our plans for further development, and note the potential that the application has not only for ESS and SHARE, but for other national and cross national surveys. We will also provide a short live demonstration to show its potential.

Metadata and Data Dissemination

Documenting Cross-Country Comparability of Survey Data: The Case of the Generations and Gender Programme

Arianna Caporali, INED - The Institut National d'études Démographiques

Survey data can be usable only if accompanied by comprehensive documentation. In international surveys, metadata are paramount to evaluate cross-country comparability of data. They help identify issues of conceptual or methodological equivalence between country datasets. This is especially crucial for surveys that have a decentralized management model that is relatively dependent upon post hoc harmonization of data. Such surveys require full documentation of countries differences in fieldwork methodologies, as well as of harmonization procedures to provide users with comparable data.

Based on these premises, this paper presents the ways in which metadata are provided within the Generations and Gender Programme (GGP). This is a longitudinal survey of 18-79 years old in 19 countries in Europe and beyond run by a consortium of research institutions (since about 13

years) following a decentralized management model. The challenge of documenting GGP data is the need to combine information from existing surveys with information on the harmonization process. Metadata are provided in compliance with the Data Documentation Initiative (DDI), the international standard recommended by the Consortium of European Social Science Data Archives (CESSDA) for documenting social science data. DDI provides a framework and a format that allow for data and metadata online publication. To publish data and metadata online, the GGP implements the software package NESSTAR (Norwegian Social Science Data Services).

After an account of the GGP, the paper describes how its surveys are presented online and the DDI elements used to document them. Surveys are presented in three ways. To provide users with cross-country overviews of data, consolidated data files are published. In these files, variable metadata account for country specificities, whereas general survey metadata only regard the international guidelines. In order to fully account for cross-country deviations from these guidelines, country-specific data files are also provided. An additional file documents availability of variable across countries and permits to have insights into country compliance with the international questionnaire. GGP national survey producers prepare metadata in advance. The experience of the GGP coordination team in searching and synthesizing missing information is discussed. It is emphasized that in the framework of surveys based on a decentralized management model the collaboration with data producers is crucial for providing high quality metadata. The paper ends with an outlook over planned developments for GGP data documentation.

Comparative Survey Data Dissemination: Archival Challenges

Peter Granda, University of Michigan

With increasing frequency, producers of comparative survey data collections disseminate their data and documentation through specially designed and constructed web portals. These portals may permit downloading of core data files and codebooks but their real value lies in providing users with multiple paths and strategies to engage with all the materials necessary to work effectively with comparative data. A web portal is an ideal vehicle for this purpose because it can organize information to appeal to users from those simply seeking an initial exploration of the data to others with specific research questions in mind. Depending on the objectives of the investigators, these portals can provide the ability to subset the data and retrieve the relevant documentation; permit online analysis without actually downloading any data; compare questions and responses across countries and/or time periods; assess the value of any harmonized variables that may have been produced and see their relation to the original questions; and immediately obtain training resources and other aids to understand how to analyze the data properly. This presentation will describe the long-term consequences of this type of survey data dissemination. It will address such questions as:

- What happens when funding to support the web portal is no longer available?
- What are the challenges of archiving materials that were created specifically for use on the portal?
- What should be preserved from the portal and what should not?
- How does one determine which formats to use when preserving content created for the web?
- What steps can be taken to retain the relationships between source variables and target variables in harmonization projects?

Examples will be used from two projects, one cross-cultural and the second temporal, where a web portal was the primary means of dissemination to researchers. The Collaborative Psychiatric Epidemiology Surveys (CPES), funded by the National Institute of Mental Health, provides data on the incidence and risk factors of mental disorders with a special emphasis on minority groups. The Integrated Fertility Survey Series (IFSS), funded by the Eunice Kennedy Shriver National Institute for Child Health and Human Development, combines data from ten surveys on reproductive health, contraceptive use, pregnancy, and family life over five decades. Both projects are now encountering the challenges of transitioning from a web dissemination environment to one that will enhance long-term preservation.

Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling

Re-Thinking Ex Post Harmonization for Comparative Research: Data Recycling of 22 International Survey Projects

Kazimierz M. Slomczynski, Polish Academy of Science, University of Warsaw

Irina Tomescu-Dubrow, Polish Academy of Science, University of Warsaw

J. Craig Jenkins, The Ohio State University

In this paper we introduce the concept of survey data recycling as a new approach to ex-post harmonization of available surveys, with the aim of increasing confidence in the data, their interpretability and comparability, and the accuracy of empirical results. Data recycling is a complex process, which, in the case of surveys conducted on national samples within well-known international projects, involves the following steps: (a) Organizing all existing materials pertaining to these surveys: general survey documentation, specific description of data, and computer data files; (b) Standardizing existing materials to create common survey documentation, integrated codebook and combined data file suited for methodological and substantive analyses; (c) Evaluating the original materials and creating quality-control variables on how the data were gathered, described, and recorded; (c) Harmonizing technical and substantive variables and creating quality-control variables on the correspondence between source and target variables. To assess this approach we selected 22 well-known international survey projects on the basis of certain criteria, including publication records and the impact that they have on the social-science disciplines: World Value Survey, International Social Survey Programme, European Social Survey, Eurobarometer and its renditions in different parts of the world as well as several other general surveys on social/political attitudes and behaviors. We assume that integrating major international survey projects should be based on the most advanced methodology. The concepts of Total Survey Error (e.g., Groves 1989, Smith 2005, Weisberg 2005) and Total Survey Measurement Variation (Smith 2011) inspired us to look at original survey materials in terms of quality control variables (Blasius and Thiessen 2012). We propose several new variables dealing with survey-quality. We illustrate empirically the distribution of control variables within international survey projects and between them, noting very large variations within projects. We discuss the implication of these findings for substantive comparative research. This paper is part of the project Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling, coordinated by the Institute of Philosophy and Sociology, Polish Academy of Sciences, and the Mershon Center for International Security Studies, The Ohio State University (www.dataharmonization.org).

Weighting Data in Ex-Post Harmonization Process: The Consequences of Different Practices of Survey Organizations for Comparative Studies

Marcin Zielinski, Polish Academy of Science, University of Warsaw

Przemek Powalko, Polish Academy of Science, University of Warsaw

One of the important aspects of survey data harmonization for comparative studies is the issue of data weighting. The most common are three kinds of weights: design weights (for making corrections for the fact of not using a simple random sample), post-stratification weights (for making corrections for unit non-responses), and combined weights (that merge a design and post-stratification weights). The problem of existence of various types of weights in different surveys and the fact that weights are calculated using different procedure and different characteristics of the national populations reduce samples equivalence and, more generally, equivalence of the measurement at the population level. We analyze weighting practices in 22 international survey projects focusing on the consequences for comparative analysis of substantive issues. In the first part of the paper, we assess the quality of weights showing the most common biases and errors, and propose some correction procedures. In the second part we discuss the problem of weights equivalence in terms of the two proposed orientations: the subjective equivalence (internal-oriented approach) and objective equivalence (external-oriented approach). Potential advantages and disadvantages of these orientations are discussed from the perspective of survey data harmonization. This paper is a part of the project Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling, coordinated by the Institute of Philosophy and Sociology, Polish Academy of Sciences, and the Mershon Center for International Security Studies, The Ohio State University.

Quality Control of Variables Harmonized Ex-Post: An Analysis of 22 International Survey Projects

Marta Kolczynska, Polish Academy of Science, University of Warsaw

Matthew Schoene, The Ohio State University

Ex-post harmonization of survey data, that is, the various procedures by which source variables from existing datasets are combined into a target variable, offers a multitude of advantages to researchers. Harmonized datasets extend the scope of analyses beyond the country and timeframe covered by a single survey program, thus allowing researchers to study social phenomena in a broader context. Yet, harmonization also raises many challenges, due, among others, to the different questionnaire features that survey programs employ. These differences can occur on the level of the questionnaire item (such as question wording or response coding), and on the level of the entire questionnaire (like question ordering). They are likely to affect the target variable. In other words, we can say that a target variable is a function of the source variable(s) and in addition control variables capturing such differences. This has implications for the modeling techniques to be used on harmonized data. In this paper we present a theoretical model in which the survey is treated as a separate dimension in the multi-level data structure typically used in cross-national comparative analyses. Specifically, we consider two types of control variables. First, we include information on each harmonized variable, such as the original response scale or the timeframe referenced in retrospective questions. Secondly, we must control for information on the entire questionnaire such as its length and the placement of the target variable in it. For empirical analysis we use an integrated file of 22 international survey projects covering 140 countries and territories from the 1960s until 2013. We demonstrate the effects of these control variables on the target variables dealing with trust in public institution (parliament,

political parties and legal system). The presentation is part of the project "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective", a joint endeavor of the Institute of Philosophy and Sociology at the Polish Academy of Sciences and The Ohio State University.

Documentation Quality Assessment in Ex-Post Survey Harmonization: Implications for Comparative Research

Ilona Wysmulek, Institute of Philosophy and Sociology of Polish Academy of Science

Olena Oleksiyenko, Institute of Philosophy and Sociology of Polish Academy of Science

Anastas Vangeli, Institute of Philosophy and Sociology of Polish Academy of Science

Within the broad taxonomy of survey quality assessment, our approach to comparative analysis of surveys belongs to a type of seeking maximum quality (Lynn 2003). We argue that in ex-post harmonization practice researchers underestimate crucial aspects of quality check of data documentation. Our central concern is the inconsistency between the survey documentation and the data files in 22 well known international survey projects. We analyze the errors that occur on the level of individual variable. For this aim, we have gathered information about such chosen variables as gender, age, education (level of education, schooling years), trust in parliament (opinion), and participation in demonstrations (report on behavior). Information on these variables is extracted from (a) codebooks, (b) questionnaires, (c) SPSS dictionaries, and (d) the data files. From each of the source of documentation we have recorded variable name, question number, exact question formulation, variable label, and value labels and compared it for inconsistencies. We recorded all of the discrepancies that appeared between any of these elements, created the taxonomy of possible errors that can occur between data and documentation and used specific criteria of assessing the significance of each discrepancy. Based on this work, we propose a quality index that assesses the surveys in terms of the average number of errors relative to the the number of items analyzed. We present examples of typical errors and discuss the consequences for comparative research. This paper is a part of the project Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling, coordinated by the Institute of Philosophy and Sociology, Polish Academy of Sciences, and the Mershon Center for International Security Studies, The Ohio State University.

Instrument Technical Design

Computer-Assisted Measurement and Coding of Educational Qualifications (CAMCES)

Silke Schneider, GESIS, Leibniz Institute for the Social Sciences

Verena Ortmanns, GESIS, Leibniz Institute for the Social Sciences

Jessica Herzing, GESIS, Leibniz Institute for the Social Sciences

The individual 's educational qualification is a core social background variable in all standardised surveys. However, the centrality of this variable contrasts clearly with its inadequate measurement, especially in the case of migrants and cross-national surveys. Usually, educational qualification is measured in a closed-ended question format asking for the highest educational qualification achieved. The response format limits the number of answer categories to the most common qualifications in the country surveyed. The given answer categories are normally harmonised post-hoc in classifications schemes which allow cross-national comparison. An inconsistent level of detail of response categories across surveys and countries emerges, due to increasing differentiation of educational systems as well as education and work-related

migration. This complicates the measurement and harmonisation of educational qualification in surveys. In the project computer-assisted measurement and coding of educational qualifications in surveys (CAMCES) a software application is developed to improve the measuring and harmonisation of educational qualifications in computer-assisted surveys, which are cross-national or contain migrant populations. This application is based on an international lookup database containing educational qualification of more than 40 countries. Additionally, the database holds information on the mapping of these certificates towards the international standard classification of education (ISCED) 1997 and 2011, which allows automatic coding. The lookup database can be integrated in the survey software and thus allows respondents or interviewers to enter the exact name of the educational qualification independent of the surveyed country. Two different interfaces of the lookup database (dynamic textfield and dynamic list/tree) were pretested in a CAPI survey in 2014. The pretest focused on the largest migrants groups of Germany such as Italy, Poland, Russia, and Turkey. The results show that the respondents are able to find their educational certificate from their country of origin. The usability of the lookup database was easy and intuitive for the respondents. Currently, the application is tested for the German case in the innovation sample of the German Socio-Economic Panel Study (SOEP). The software application will also be implemented in the migrant sample of the SOEP in 2015.

Adapting Technologies for the Developing World: Case Study in Ghana

Elana Cohen-Khani, Statistics in Netherlands

Beth-Ellen Pennell, University of Michigan

Gina-Qian Cheung, University of Michigan

Yu-chieh (Jay) Lin, University of Michigan

This paper will describe survey and monitoring solutions developed to conduct in-person interviews for large scale community and household panel surveys in Ghana. Survey technologies that have long been used in the United States to enhance data quality were successfully adapted to address challenges introduced by a new context and environment. These innovations included a parallel block instrument design to allow the collection of household, individual, agricultural, and business level data in a highly flexible manner, as well as a case management and data transmission system that allowed researchers rapid access to a rich array of paradata. Project managers were able to make use of timestamps, keystroke data, GPS data, and respondent contact records to track interviewer performance and movement, and map local infrastructure for analyses. Adaptations were also made to address the frequent lack of phone and internet connectivity that is common in rural areas of developing countries. This paper will discuss the challenges introduced by adapting old tools to a new environment, lessons learned while finding solutions to these challenges, and how these solutions enabled a more effective and efficient data collection process.

This paper will particularly evaluate how interviewing navigation through the instrument has been done to determine if the complex instrument designs have accomplished the flexibility of interviewing multiple respondents with varied topics without negatively impacting interviewing efficiencies.

Special Talk

Investigating Cultural Consensus Theory in the CSDI Context

Brad Edwards, WESTAT

Cultural consensus theory (CCT) was developed in cultural anthropology to allow researchers to systematically interview a few members of a cultural group about a topic (say, cooking or kinship), to pool the answers and estimate the answer key, thus arriving at an understanding of the culture informed by the respondents' shared opinions about the topic. A key assumption in CCT is that the way someone experiences a culture is a product of their own direct experience and what they have learned from others. Thus members of a culture more or less share a knowledge and understanding of a particular domain. CCT provides quantitative measures of consensus within the culture, and of cognitive characteristics of its members.

Recently Oravecz, Faust and Batchelder (2014) proposed an extension of CCT applied to social surveys, taking uncertainty ("don't know" responses) explicitly into account, adding covariate information, and deriving statistical inference. They offered two examples, one drawn from General Social Survey data on knowledge of science and the other from an American Association of Retired Persons survey on knowledge of aging.

In thinking about this work in the CSDI context, I explore several questions:

- How does the Extended CCT differ from psychometric work (IRT)?
- Can the Oravecz et al findings be replicated in 3MC surveys?
- Could analysis of ISSP data extend their GSS work?
- The CCT findings are typically robust based on just a few (12 to 20) interviews. Could the Extended CCT be useful in developing 3MC designs?
- Oravecz and colleagues used data from True/False/Don't Know items. They suggest the Extended CCT could be applied to Likert scales. How might one go about doing this?

My hope is to outline a research agenda for investigating the benefits of CCT for CSDI.

REFERENCE

Oravecz, Z., Faust, K., and Batchelder, W.H. (2014) "An extended cultural consensus theory model to account for cognitive processes in decision making in social surveys," in *Sociological Methodology*, 44(1), 185-228.