2018 CSDI Workshop Abstracts

Monday, March 26th

Session: Using Cross-Cultural Cognitive Pretesting for Questionnaire Testing

The Cognitive Process of Reporting Educational Qualifications in Cross-cultural Surveys: A Comparison between Immigrant and Non-immigrant Respondents in Germany

Roberto Briceno-Rosas, GESIS - Leibniz Institute for the Social Sciences

Abstract: Surveys struggle to measure the educational qualifications of immigrants mainly due to the diversity of their educational backgrounds compared to local populations. New technological solutions, like the CAMCES tool, enable surveys to ask all respondents the same educational question while allowing for country-specific reporting of the educational qualifications in multiple educational systems. Can researchers assume that the answering process of immigrants is similar enough to that of non-immigrants? Or are there important singularities in the cognitive process that need to be account for? To address these questions, cognitive interviews were conducted with both immigrant and non-immigrant respondents in Germany using the CAMCES tool. The analysis focuses on the understanding of the concept of formal education and the reporting of educational qualifications. The study shows the similarities and differences in the cognitive process between immigrants and non-immigrants. The discussion of the results highlights the consequences of context effect for questionnaire design, especially how language influences the answers of immigrants.

Cognitive Pretesting of Sensitive Questions for Cross-Cultural Surveys: Methods and Challenges

Patricia Hadler, GESIS – Leibniz Institute for the Social Sciences

Abstract: Sensitive questions are omnipresent in surveys, covering a wide range of topics from income over religious affiliation, sexual behavior, abortion, criminal offenses or alcohol consumption. They include attributes, attitudes or behaviors that are disapproved of, embarrassing or illegal. Such questions may trigger social desirability concerns, be seen as intrusive or raise concerns about repercussions of disclosing the information. Sensitivity affects the quality of survey data in terms of non-response rates (on item and unit level) and response accuracy through misreporting. Asking and comparing sensitive questions cross-culturally poses additional challenges, as the same question may differ in sensitivity, as may respondent reactions to such a question.

In-depth pretesting of (potentially) sensitive questions is all the more important. Cognitive interviewing is a pretesting method used to reveal the cognitive process underlying survey response. It aims to identify respondents' ability and willingness to answer correctly. Cross-cultural cognitive interviewing (CCCI) is used when a questionnaire is administered in more than one language and/or to respondents that differ significantly in their cultural/linguistic background. CCCI has the additional goal of establishing cross-cultural equivalence of survey questions.

However, both CCCI and question sensitivity pose challenges to researchers, especially from a methodological viewpoint. Methodological research for cross-cultural pretesting of sensitive questions is necessary in the areas of data collection and analysis as well as study design.

Prior to testing optimal study designs, a theoretical framework for collecting and analyzing CCCI data is needed. This paper systematically reviews literature on cross-cultural cognitive interviewing and sensitive questions. Drawing on the work of Roger Tourangeau, Gordon Willis, Kristin Miller, Debbie Collins and other leading researchers in the field of cognitive interviewing methodology and practice, a taxonomy for systematic cognitive pretesting of sensitive questions in a cross-cultural context is presented. It consists of question appraisal, corresponding probing techniques and subsequent data entry and analysis.

Aims of CCCI for sensitivity are
1. determining the general level of sensitivity or threat,
2. identifying sensitive elements and dimensions of questions and answers

3. mapping possible bias to the appropriate stage(s) in the cognitive process of survey response and
4. attributing these parameters to generic questionnaire design, translation, and/or cultural adaptation.

Mixed-Method Approach for Understanding Scale Performance in a 3MC Context
Kristen Miller, National Center for Health Statistics

Abstract: This presentation will describe a mixed-method approach currently being explored at the National Center for Health Statistics to better understand the performance of scaled survey items across demographically diverse respondent groups. In so doing, the presentation will 1) illustrate how analysis of cognitive interviews can provide qualitative insight into scales by revealing the specific phenomena considered by respondents as well as the basis for their particular answer, and 2) how cognitive interview findings can be incorporated into a quantitative study to further investigate scale performance. Traditionally, cognitive interviewing has been used as a pretest method to identify and fix problems in questions prior to fielding the full survey. However, a cognitive interviewing study can also function as a validity study in that it reveals the ways in which respondents interpret questions, apply those interpretations to their own experiences and circumstances and, ultimately, formulate an answer. In short, a cognitive interviewing study is able to ascertain the basis for respondents' answers and the explicit, functional meaning of the question's response categories. For example, a cognitive interviewing study examining the question, Do you have difficulty seeing, even if wearing glasses? No, no difficulty, Yes-some difficulty, Yes- a lot of difficulty, or Cannot do at all?, found that respondents based their answers on vision quality as well as the circumstances impacting quality: While respondents who answered "no difficulty" described their vision as always being clear, those who answered "some difficulty" did so because, in some situations depending on what they are doing and their surroundings, the quality of their vision was less than optimal; those answering "a lot of difficulty" experienced poor vision quality regardless of activity or environmental circumstances, and those reporting "cannot do at all" described being completely unable to see, that is, no quality of vision to even consider. These findings, then, present a precise and functional understanding of what is meant by the categories "no difficulty," "some," "a lot," and "cannot do." In a mixed-method design, these findings can be used to develop follow-up probe questions to be embedded into a survey questionnaire, providing quantitative data to verify and further investigate the qualitative findings. In this case, questions pertaining to vision quality, circumstances and activities impacting vision quality can be used to assess the extent to which, as well as consistency across respondent groups, these concepts form the functional basis for the question response categories.

Considerations for Cross-Cultural Questionnaire Testing and Lessons Learned
Cornelia Neuert, GESIS
Patricia Hadler, GESIS
Timo Lenzner, GESIS
Natalja Menold, GESIS

Abstract: The primary aim of cross-cultural cognitive interviewing is to ensure that survey instruments capture the constructs they are intended to measure and to determine whether these constructs are consistently understood and captured across countries and/or languages.

This paper first describes considerations and strategies of the GESIS Pretest Lab for cross-cultural cognitive interviewing and then summarizes experience from two different studies.

In the first study, focus groups were used for cognitive testing in two African countries (Ghana and Togo). To enable the participating countries to conduct focus groups for cognitive testing themselves, interviewers received information and guidance on the methodological background of cognitive testing and were trained remotely in cognitive interviewing and focus groups techniques. The entire application process from recruiting, sampling to reporting the findings was closely monitored by GESIS. Analysis was carried out centrally for both countries by GESIS.

In the second project we tested a questionnaire for refugees and asylum seekers in Germany. Testing was carried out locally, using interpreters that were present during the cognitive interview using online video conference to communicate with participants from various countries. To realize standardized interviews, interpreters were sent both the questionnaire in the target language and the cognitive interview protocol in the source language (German) prior to

the interview. A total of four languages were tested using simultaneous translation; each interpreter carried out simultaneous translation for two interviews in their language.

For both cases, we will present the various challenges we experienced as well as lessons learned and will reflect on our solutions to those challenges taking into account the given conditions.

Session: Questionnaire Design

Ingroup Bias and Its Implications for Using Visual Images in Computerized Surveys
Sunghee Lee, University of Michigan
Mengyao Hu, University of Michigan

Abstract: Computerized questionnaires provide flexibility and convenience for using visual images, a simple yet effective tool that requires less cognitive efforts from respondents than verbal descriptions. Particularly with the increasing popularity of Web surveys, it is appealing to display visual images as part of a questionnaire. Especially for questions about subjective concepts (e.g., general health status) that may mean different things for different people, displaying images along with the questions may clarify/standardize question meanings. However, for cross-cultural surveys, visual images may inadvertently complicate measurement comparability due to ingroup bias, a pattern that people favor or reject members of own group (ingroup) over members of other groups (outgroups).

We explore potential effects of ingroup bias in using images through experiments from two Web surveys. In both surveys, respondents were shown a series of pictures illustrating people in different health-related situations (e.g., jogging, walking on crutches) one by one and asked to evaluate health status or health-related life style of each hypothetical person. For each health-related situation, we experimented the profile of the hypothetical person in such a way that for a random subset of respondents their profile matched with the person in the picture and for the rest not matched.
In the first Web survey conducted by the University of Michigan, we targeted foreign-born Koreans in the U.S. and showed pictures with three health-related situations: 1) overweight young man eating cake; 2) young Asian passed out at a bar; and 3) Asian female jogging outdoors. Specifically, we asked respondents to rate how healthy they think lifestyle of the person in the picture is. In each health-related situation, we varied the profile of the picture person: 1) White vs. Asian for cake eating pictures; 2) male vs. female for bar pictures; and 3) young vs. old for jogging pictures. With a similar logic, we conducted a Web survey targeting the general U.S. adult population with an oversample of racial/ethnic minorities focusing on difficulties around four health domains: pain, affect, mobility and sleep. For each domain, we showed a series of pictures that depict different levels of difficulties. In these pictures, we varied the gender, age, race/ethnicity and body mass of the picture person. Using the data from these experiments, we will examine whether and how ingroup bias plays a role.

Measuring Subjective Health and Life Satisfaction with U.S. Hispanics
Sunghee Lee, University of Michigan
Rachel Davis, University of South Carolina

Abstract: Health and well-being are two important issues not only in research but also in policy. While accurate measurement of these attributes is critical, extant research indicates its difficulty in survey research. The main difficulties stem from the subjective nature of the concepts of health and well-being, as well as the use of response options with vague quantifiers. These difficulties become more evident for cross-cultural studies, where the concepts of health and well-being themselves may not be comparable. Moreover, their measurement instruments may not function equivalently.
This study focuses on the U.S. Hispanics and examines three variables: 1) self-rated health (SRH), 2) life satisfaction (LS) measured with single item and 3) LS measured with five items. On SRH, Hispanics are known to report negative health more than non-Hispanic Whites, and non-equivalent translation of the English response categories of “excellent,” “very good,” “good,” “fair,” and “poor” has been hypothesized as a potential contributor. With the 5-item LS scale, all items are stated in a positive direction and asked with the Likert-type agreement response scale. As Hispanics are shown to be associated with acquiescent response style, the current 5-item LS scale may lead to an overestimation of LS for

Hispanics. Motivated by these specific issues, we implemented the following experiments: 1) on SRH, translation of the response category, "fair," into "regular" versus "pasable"; and 2) on the five-item LS scale, direction of item wordings to be all positive versus balanced. Respondents were randomly assigned to one of the two conditions under each experiment. These experiments were implemented in three different surveys that were designed to be incremental in the scope of target group and in the objective measures of health and well-being that allow us to ascertain better ways to measure SRH and LS: 1) a web survey of Spanish-speaking Hispanics (n=1,416); 2) a telephone survey of Hispanics focusing on Mexican, Puerto Rican and Cuban heritages (n=1,296); and 3) a web survey of Hispanics of all background (n~1,600). Using the experimental data, we will examine simple response distribution, item characteristics based on item response theory, relationships across three subjective measures and between these subjective measures and objective measures. Where appropriate, these examinations will be carried out by considering interview language.

Session: Data Quality and Comparability in Survey Data Harmonization

Comparing Two Measures of Education Harmonized Ex Post: Levels of Education and Years of Schooling

Marta Kołczyńska, Institute of Philosophy and Sociology, Polish Academy of Sciences
Irina Tomescu-Dubrow, Institute of Philosophy and Sociology, Polish Academy of Sciences

Abstract: Education, as an important explanatory factor for many sociological phenomena, is an almost omnipresent variable in statistical analyses using survey data. Although seemingly straightforward, education proves hard to measure accurately in surveys. This, in turn, raises difficulties for using the education variable in comparative studies, and provides additional challenges for harmonizing education measures ex-post.

In this presentation we examine the association between two measures of educational attainment commonly found in general social surveys: education levels and years of schooling. We use the Survey Data Recycling dataset (SDR), which pools information from 22 international survey projects. The SDR dataset contains a number of target variables which have been harmonized ex-post to make them comparable across surveys, including "education levels" and "schooling years". These two education variables have been harmonized independently from one another on the basis of different source variables. Of the total 1721 national surveys, we analyze 1189 surveys that contain both target variables for completed education – levels and years.

We find that in most of the 1189 national surveys the association between the target variables "education levels" and "schooling years" is positive and strong, as would be expected for two operationalizations of the same concept. However, we found 34 surveys where the rank correlation between levels and years is below 0.5, and four where it is even weaker and negative. Next, we explore the effects of harmonization control variables, which bring in information about the design of the survey questions and ways of coding responses that would be lost in the process of harmonization, to see how including them affects the association between the two target variables.

We conclude by discussing (a) possible sources of discrepancies between the harmonized variables "education levels" and "schooling years", and (b) opportunities that examining these discrepancies create for improving the measures of education in cross-national surveys.

Comparing educational attainment from the other way around: A new perspective

Dean Lillard, The Ohio State University

Abstract: It is difficult to harmonize data that have no clear link to an objective physical measurement. This paper works with data on educational attainment. There is great interest in harmonizing data on educational attainment because it figures so centrally both as a resource people use to generate income and because social scientists use educational attainment as a proxy for social and economic position. Several factors make it difficult to harmonize data on educational attainment. The primary problem arises because educational attainment is supposed to correspond to a theoretical construct termed "human capital" but social scientists define human capital in vague terms that are difficult to match to empirical measures. For example, it is common to define human capital as a set of skills or knowledge embodied in a person that firms or others want to "rent" to produce goods or services.

This definition is clear in the abstract. Problems arise when one tries to define an empirical measure of human capital because the "market" demands such a wide set of very diverse skills and knowledge. For example, human capital clearly includes both knowledge of the fundamental laws of physics and how to repair a leaky faucet. As an empirical issue, one needs to group specific human capital knowledge or skills. The practical problem a social scientist faces when trying to harmonize data is that available empirical measures lack the detail one would in principle need to develop truly comparable measures. One lacks such data primarily because it is simply too expensive to collect them. This second feature complicates the task of harmonizing data on educational attainment because researchers lack the necessary inputs.

In this project I explore whether it is possible to harmonize measured educational attainment using estimates of the returns to different types and levels of schooling. This strategy picks a particular calendar year as the base year. It then uses cross-sectional survey data to estimate the rate of return to various levels of attained education. A level of education is then defined relative to a base category according the percentage increase or decrease in the market wage each level of human capital earns (relative the wage earned by people holding the reference category educational attainment). I will then harmonize educational data from other countries using this strategy – defining each level/type of education as comparable if it returns the same percentage increase/decrease around the reference category. The advantage of this approach is that it harmonizes educational attainment against the use of that composite good in the context of the economy in a common calendar year. The strategy has the disadvantage that the return on observed levels and types of human capital is a complex function of the extant labor market and levels of capital that workers use in each time and place.

This study is an exploration of an alternative to harmonization that, until now, researchers have not tried. The resulting harmonized measures will be compared against other harmonized measures that suffer from other factors that weaken the comparability of those harmonized measures.

Inter-survey variability of corruption measures: Implications of harmonization procedures

Ilona Wysmułek, Institute of Philosophy and Sociology, Polish Academy of Sciences

Abstract: Survey data harmonization is the procedure that allows to combine different sources into an integrated dataset with comparable indicators. From the empirical point of view, there are advantages in using multi-source dataset to test research hypotheses. Among other things, using more data sources makes the results robust, as different surveys should not tell contradictory stories about the same countries. Yet, in harmonizing survey data ex-post, researchers most certainly face the challenge of the inter-survey variability of measures, with some countries showing significant fluctuations between results for the same year. Despite the rapid developments in the field of survey harmonization, there are still no easy answers to either sources of inter-survey variability, nor the strategies of dealing with it.

In this paper, I aims to present the analysis of the between survey fluctuations and validity tests of the two corruption measures harmonized ex-post: corruption experience in education and perception of corruption prevalence in education in Europe. To have the insight into the variation between survey projects, I have calculated the fluctuation ratio of between survey results, which is an absolute value of differences in percent points between a pair of survey projects. I than move on by showing whether the differences between survey projects are statistically significant, when controlling for other relevant variables. Further on, I concentrate my analysis on one of the possible sources of the inter-wave fluctuations, which is the differences in response scales. Specifically, I assess the loss in explanatory power of the measures in the construct that has been harmonized ex-post, in this case after the recode of 11-point and 5-point perception scales to binary variables.

This paper is based on the new integrated dataset on corruption in public schools in Europe that I created for the purpose of my dissertation. The integrated dataset consists of the harmonized micro-level data gathered from the Life in Transition Survey, the Global Corruption Barometer and the Quality of Government survey - a total of 96 national surveys conducted in 30 European countries in 2010; and the integrated macro-level data of country level indicators and education system characteristics from the World Bank Education Statistics, the Varieties of Democracy, the Quality of Government Standard Dataset and UNESCO Institute of Statistics Education Indicators. I apply the documentation standards and analytical framework developed by the research team of prof. Kazimierz M. Slomczynski (SDR Harmonization Project) to ensure transparency and replicability of the harmonization procedures.

Validity and comparability of the education measures in the European Values Study 2008

Verena Ortmanns, GESIS-Leibniz Institute for the Social Sciences

Abstract: These days respondents' educational attainment is measured using different questions in a survey and implementing various classifications, mostly done after data collection.
In the European Values Study (EVS) 2008 we also find different kinds of education variables in the dataset. First of all, each of the 47 countries/ regions has its country-specific variable containing the original names of the educational qualification. Secondly, derived from these country-specific categories the International Standard Classification of Education (ISCED) 1997 is implemented and provided as a further education variable. Thirdly, an eight-category education variable developed by the EVS is derived from the ISCED classification. This variable is quite similar to the classification for Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) which was developed to compare educational qualifications across European countries. Next, we find a variable on "low-middle-high" education which is a common-used education variable used as control variable in a large number of studies. Finally, the EVS also ask on the age when finishing education to generate a continuous variable on the years spend in education.

These described education measures differ regarding the underlying concept of the question, the concepts of the classifications, and also regarding their level of detail. This presentation validates the different education measures provided in the EVS 2008. The research question is: How comparable and valid are the different education measures of the EVS 2008? Several regression analyses are calculated, separately for each country and each education variable to answer this question. The dependent variable is on political participation and the placement of respondents in the left-right scale. This variable is chosen because prior research has shown that it strongly correlates with education. As independent variables I use the above-described education variables. The analysis also controls for age and gender of the respondents. For the results, I will mainly focus on the adjusted R squares of the regressions to validate the different education measures, and I will also have a look at the effect sizes.

Looking for ways to characterize countries politically and economically using longitudinal data

Claire Durand, Université de Montréal
Paul Pelletier, Université de Montréal

Abstract: In order to understand contextual effects better, we need to be able to compare countries on a number of relevant characteristics. In trying to do so, we face two problems. The first is the fact that many indices of the situation of the different countries are available only for countries of the "western world". It is very difficult to find information that is equally relevant and available in the different regions of the world and for all the countries. The second problem is that the characteristics of interest change over time. It therefore appears more relevant to characterize countries according to change in their situation over time rather than to their situation at one point in time.

In this paper, we present the results of our longitudinal cluster analyses using K-Means longitudinal analysis (KML) and trajectory analysis of 98 countries of Asia, Africa, Latin America and the WANA region. We performed two series of analyses. The first one aims at clustering countries based on change in institutional trust over two decades (1995-2015). It uses a combined dataset of close to 700 polls conducted by the Barometer projects, LAPOP and the World Values survey. The second series of analyses aims at clustering these same countries over the same period on covariates that may/should be related to institutional trust. These covariates include indicators of the economic situation, like the PIB and the GINI index, the political situation, like the Governance indexes (WGI, QOG indicators), and the social situation (access to education, the situation of women, etc.) of the different countries. We present our different attempts and the criteria we used to select the most relevant and productive clustering in terms of our capacity to understand the variation in the situation of different countries and regions of the world, and particularly in the level of trust for different institutions.

Session: Comparative Standard Demographics

When Standard Wording for Language Spoken at Home Does not Work Standardly across Languages
Marjorie Hinsdale-Shouse, RTI International

Abstract: In 2011, the U.S. Department of Health and Human Services issued Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status for all federally sponsored surveys. This paper analyzes one survey's experience implementing these standard questions related to languages spoken at home and documents a disconnect in respondents' understanding when administered in English versus other languages.

The Outpatient and Ambulatory Surgery Consumer Assessment of Healthcare Provider Services Survey (OAS CAHPS), sponsored by The Center for Medicare & Medicaid Services, includes these standard questions related to languages spoken. However, based on feedback from interviewers and data review, we found that Spanish-speaking respondents misunderstood one question. The question, "Do you speak a language other than English at home?" works well for English, however, when administered in any another language, the question is not framed from the perspective of someone who does not speak English. As such, respondents tend to answer, "No, I don't speak English at home." Without having an interviewer probe for accuracy, the data collected for this question were inaccurate because "Yes" indicates a language other than English is spoken and "No" indicates English is spoken.

Ongoing efforts to clarify this language question continue as the survey is being administered during the national voluntary implementation to patients of U.S. hospital outpatient departments and ambulatory surgery centers. Formatting modifications were made that did not alter the standard wording from the guidelines. The modifications included reading more detailed response options out loud and underlining key words in the question for emphasis. For example, rather than yes/no responses, the responses were: "Yes, I speak a language other than English" and "No, I speak English at home."

This paper explores the ongoing analyses of the modifications made to this question as well as the findings from the use of the standard question before adjustments to improve respondent understanding. The paper also explores alternative wording that could be tested to collect more accurate data from non-English speaking populations. The goals of this review are to better understand the impact of continued use of the standard demographic questions related to language use on the survey data and to identify an acceptable alternative that is better understood by respondents.

Survey Experiments on Income Non-Response in Cross-National Surveys
Steve SCHWARZER, Pew Research Center
Laura Silver, Pew Research Center
Patrick Moynihan, Pew Research Center

Abstract: Income measures are a key factor for understanding how individual attitudes vary within and across nations when it comes to political opinions, attitudes towards minorities and general life satisfaction. Creating comparable, cross-national measures of household income is inherently challenging, given differences in language, context and the often prohibitively high item non-response rates that result.

Using data from a survey experiment embedded in Pew Research Center's cross-national surveys, we explore one mechanism for reducing refusal rates on income questions. We focused on testing the effect of follow-up questions after asking closed-end questions. We tested (1) a closed-ended question with an above/below median income follow-up for those who refused to answer the initial closed-ended question, and (2) a split-ballot experiment that randomly assigned respondents to either the above-mentioned two-step approach or the one-step above/below median-only question.

The paper demonstrates the effectiveness of the different approaches but does illustrate that combining a closed-ended income question with more general, median-based follow-up options significantly reduces item non-response for

income in the majority of countries. However, only asking a median income question can introduce differential bias across demographic groups into the results – strongly suggesting that rigorous testing of any measurement alterations is required before adoption.

Session: Questionnaire Translation in Theory and Practice: Challenges, Solutions – and New Ideas?

Translating questionnaires for cross-national surveys – a text genre and its particularities
Dorothée Behr, GESIS

Abstract: Translation is a highly specialized activity that requires a set of diverse competences, ranging from linguistic and cultural over subject-matter and tools to strategic and service-oriented competences. While many of those competences are general in nature and apply regardless of subject-matter and text genre, it is obvious that the more a translator knows of a given subject-matter – and in close relation to that, of a pertinent textual genre – the better he or she can translate a given text. In this presentation, an attempt will be made to structure the knowledge that a questionnaire translator will need alongside the ISO 17100 categorization of professional competences of translators. The goal is to stimulate a discussion on what a questionnaire translator should know or be aware of in order to produce high-quality translations.

TMT: Questionnaire Translation Management for large international studies. Lessons learned and challenges ahead
Brita Dorer, GESIS-Leibniz Institute for the Social Sciences
Maurice Martens, CentERdata

Abstract: The Translation Management Tool (TMT) is an online service for supporting questionnaire translation processes for large multilingual surveys, originally developed for SHARE. It has recently been adapted to be usable by projects applying a 'team' or 'committee' approach or TRAPD for their questionnaire translations. In the context of the SERISS cluster project, it was used in 2016/2017 by the following surveys in real-time: for the ESS Round 8 in three countries, involving TRAPD, external translation verification and SQP Coding; in addition, the EVS and GGP used the TMT for preparing all of their national versions. In ESS Round 9, that is, starting in spring 2018, the TMT will be used for producing the translated survey instruments in 5-10 language versions. The paper will discuss changes that have been made to the TMT as a consequence of these testing exercised and what the next steps will be.

MemoryLane - A repository and online search engine of legacy materials in large-scale multilingual surveys
Andrea Ferrari, cApStAn

Abstract: For many large-scale multilingual surveys or assessment studies, legacy materials (validated translations from previous rounds or cycles) are a precious resource for all the players involved in localizing the survey instruments for the 'new' round. The ability to easily search legacy translations helps greatly to both ensure consistency of new materials with legacy materials (e.g. recurring instructions, adaptations used in the past) and to provide inspiration to translators/reconcilers/verifiers.
For the attention of interested survey methodologists, cApStAn would like to make a presentation and demonstration at the CSDI 2018 workshop of MemoryLn, its proprietary searchable repository of legacy materials, designed by cApStAn's lead translation technologist Manuel Souto Pico.

PISA2018 is the first PISA cycle for which MemoryLn was made available as a resource firstly to produce the French source version (the PISA localization process is based on a double source design) and then for the translation/adaptation/verification of all national versions of new cognitive items. This innovation was warmly received as a precious help to retrieve entire segments (e.g., "Click on a choice to answer the question.") and to look for specific terminology pertaining to the user interface, such as names of buttons, verbs used in IT (e.g., "drag and drop", "click", etc.) and screen elements (e.g., "tab", "scroll bar", etc.). Translators, reviewers and verifiers also consulted MemoryLn to check how certain adaptations were handled in the past (e.g. "Zedland" and "zed", the fictitious country and currency used in PISA items involving money), and in general as a source of inspiration.

We will provide some details on how MemoryLn is "fed" with materials (bilingual files in various formats, but with some constraints) and the system of credentials put in place to allow selective access to different users as regards confidential materials.

We will touch on developments that are underway or planned, foremost of which are 1) allowing search and display of meta-data (e.g. item identifiers, comments or notes); and 2) building in the ability to "call up" MemoryLn when translating XLIFFs using a CAT tool.

SERISS experiment for testing close versus more adaptive questionnaire translation approaches: first results from the translation sessions and from the field
Brita Dorer, GESIS-Leibniz Institute for the Social Sciences
Ana Villar, City, University of London

Abstract: In 2017, an experiment was carried out (in the context of the SERISS cluster project) in order to test two different questionnaire translation approaches: the so-called Ask-the-Same-Question approach versus approaches that allow more adaptation. The experiment was carried out in two languages, Estonian and Slovene, and the resulting questions fielded in the CRONOS webpanel.
The paper will analyze first findings from the translation activities, comments made during the translation process as well as from the national teams which already showed different understandings of the two translation approaches mentioned by experienced survey practitioners. In addition, the outline for the statistical analyses will be presented and discussed.

Training Translators for a Multi Country Study: Lessons Learned
Alisú Schoua-Glusberg, Research Support Services
Casey Langer Tesfaye, Research Support Services

Abstract: In crossnational studies, translation of instruments is key to comparability of results. National research teams vary in the degree of experience they have selecting and managing translators. To assist the teams and try to ensure that study sites use the same methodology, training of translators must be centralized as a way to strive for consistency.

In this presentation we will report on an ongoing study where we trained translators from 6 cities around the world to translate the study instruments using the Split Committee (or Modified Committee) Approach. We developed training materials and a translator manual, and delivered the training remotely as a webinar.

We will discuss translators' prior expertise and how the training process worked, lessons learned, and recommend changes for future similar efforts.

Tuesday, March 27th

Session: Interviewer Falsification

Interviewer effects and undesirable interviewer behaviour
Ineke Stoop, The Netherlands Institute for Social Research/SCP

Abstract: Respondents interviewed by the same interviewer can be more similar than respondents interviewed by different interviewers. Reasons for this can be perfectly legitimate (neighbourhood effects, selection effects, social desirability). Respondent answers can also be influenced by unauthorised substitution, undesirable interview behaviour (speeding, cutting corners at filter questions), or falsification. The presentation gives an overview of interviewer effects that may or may not be due to undesirable interviewer behaviour.

Identification of Falsifications in Surveys – a Link to the Cross-Cultural Context
Natalja Menold, GESIS

Abstract: A literature review on methods for identification of falsifications by the interviewers is provided. Falsifications mean that interviewers do not conduct the interview at all or ask only some of the questions, and then fabricate the (remaining) data by themselves instead (AAPOR 2003). The focus of the review lies on the comparisons of studies by country/culture. The methods previously used in surveys are 1) re-interview or re-contact, 2) usage of observations or para-data and 3) (multivariate) analytical methods, which are based on certain indicators, derived from the available data. The advantages and disadvantages of the methods for the use in cross-cultural context are discussed.

Session: **Issues of Privacy and Legal Restrictions around Survey Research**

Third Parties in Face-to-Face Interviews: Qualitative Insights from 4 African Countries
Leenisha Marks, RTI International
Charles Lau, RTI International
Matthew Letterman, RTI Intertional
Melissa Baker, Kantar Public
Alexandra Cronberg, Kantar Public

Abstract: One of the main principles of survey research is privacy, where the survey is a one-on-one interaction between an interviewer and a respondent that is conducted in private. There is a growing body of literature that suggests that privatized interactions may not be the reality in face-to-face surveys, especially those conducted in low and middle-income countries (LMICs). The presence of third parties or bystanders can affect respondent reporting. Effects on respondent reporting are often amplified by the type of bystander that is present, impacting the overall quality of the data. Current literature provides insights into bystanders and their effects on data quality, but there is limited research about the interactions that occur before and during face-to-face interviews.

We conducted 8 qualitative focus group discussions with fieldwork staff in 4 sub-Saharan African countries. Our study explored three main research questions. First, we analyzed the context for third party presence – why are third parties present during the interview? Second, we examined the interactions and conversations that occurred during the interviews. Third, we explored the mitigation strategies used by interviewers to manage third parties during the interviews.

Preliminary analysis of the data suggests that power and gender relations play an integral role in the presence of third parties and data quality in these research settings. First, the gender of the respondent and the gender of the interviewer were both identified as motivations for third party presence. Second, the respondent's family role often impacted whether a third party was present during the interview. Lastly, the presence of particular third parties, like a husband or parent, would often drive the responses given by the respondent. We will also report on general mitigation strategies used by interviewers to control for third parties in these contexts.

In-person Interview Privacy Practices, Perceptions, and Preferences: Results from a Mixed Methods Study
Zeina Mneimneh, University of Michigan
Jill Wittrock: Center for Social & Behavioral Research, University of Northern Iowa
Kien T. Le: Social and Economic Survey Research Institute, Qatar University
Engi Assad Ahmed El-Maghraby: Social and Economic Survey Research Institute, Qatar University
Hao-Chun Cheng: Survey Research Center, Institute for Survey Research, University of Michigan

Abstract: Singer et al. (1993: 466) refer to the respondent's concern about privacy as the "desire to keep information about oneself out of the hands of others altogether." During an in-person interview, "others" include not only the interviewer but also any third party who might be present in the interview setting. Thus, managing one's privacy in this context includes not only managing the information reported but also managing all interpersonal interactions during the interview. There are a myriad of contextual factors that could influence the presence of someone during the interview. These factors are driven by how privacy is viewed and practiced by the different participants in the survey process: the

survey firm, the interviewers, and household members (including the respondent) who all are operating in a particular social milieu. This paper examines each of these factors through a mixed methods approach that gathers evidence from survey firms, interviewers, and household members. Data are drawn from an establishment web survey of firms in the Middle East and North Africa and from focus group discussions with interviewers and household members. The findings confirm the importance of environmental factors including cultural norms, survey firm related factors such as emphasis placed on interview privacy during interviewer training, interviewer-related factors such as gender and approach, and household-related factors such as "authority figures" in the family and household structure. The paper also provides suggestions related to survey design protocols and implementation approaches that could enhance the likelihood of private interviews.

Findings from a Worldwide Survey Concerning Freedom to Publish Public Opinion Research
Timothy P. Johnson, University of Illinois at Chicago
Marina Stavrakantonaki, University of Illinois at Chicago
Kathy Frankovic, Independent Consultant

Abstract: Restrictions on the ability to conduct and publish public opinion research remains a worldwide concern. We report findings from a joint WAPOR/ESOMAR study designed to provide updated information regarding current governmental restrictions on freedom to publish public opinion research across the globe. Fielded between July-October 2017, we collected reports from national informants in a total of 133 nations in Africa (n=26), Asia (n=36), Central America (n=7), Europe (n=42), North America & Caribbean (n=7), Oceania (n=3) and Latin America (n=19). The presentation will focus on legal "blackout periods" for reporting public opinion prior to national elections, and restrictions on the conduct and reporting of exit polls after elections. We will also summarize informant reports regarding changes in polling restrictions over the past five years, as well as their expectations regarding future changes in polling restrictions. Finally, we will investigate the degree to which elements of "national culture" are associated with the presence or absence of polling restrictions across nations. Discussion will place findings within the context of earlier Freed to Publish surveys conducted over the past 30 years.

Session: Sampling Approaches in Cross-Cultural Research

Sampling Schemes and Survey Quality in Cross-national Surveys
Marta Kołczyńska, Institute of Philosophy and Sociology, Polish Academy of Sciences

Abstract: The quality of the sample, understood as the degree to which the selected units of observation represent the target population, is crucial for the quality of the resulting survey data. To facilitate comparative research, international survey projects strive to conduct surveys on nationally representative samples of comparable quality. However, the type of sampling schemes they employ to design samples varies, both within and among projects. This variation may affect sample quality, and, ultimately, the quality of the data. This paper documents the diversity in sampling schemes in 22 international survey projects covering 142 countries and territories from the 1960s until 2013, on the basis of survey documentation. We use a classification adapted from Kohler's (2007) study of unit non-response bias in European surveys to code the different types of probability and non-probability samples that the survey documentations describe. This classification takes into account the most important formal characteristics of sampling designs, and at the same time, distinguishes between samples that leave room for interviewers to diverge from respondent selection procedures (such as samples with a random route component) and samples that facilitate tight interviewer control (e.g., probability samples of named individuals). We code separately surveys whose documentation does not provide sufficient information about sampling. To examine the extent to which quality of sampling scheme relates to quality of the survey data, broadly defined, we explore the associations between the indicator for sampling scheme and variables measuring (a) the quality of the survey as reflected in the documentation, (b) the quality of the data records in computer files, and (c) the consistency between the data records and the documentation.

OnDemand Sampling Hub for CATI projects globally

Carsten Broich, Sample Solutions BV
Maja Koceva, Sample Solutions BV

Abstract: In the past, telephone surveys have not been the most attractive method for general population surveys due to the low penetration of telephone owners for most of the countries. Consequently, the face-to-face methodology was the most popular and the oldest way of data collection. Nowadays, depending on the aim of the survey many techniques are available: from face-to face, online to telephone survey. However, the rise of telephone users in the world, especially the dramatic expansion of mobile phone users after 1995, has affected the techniques of data gathering. While face-to-face surveys remain the gold standard in public opinion polling, in more countries telephone surveys gain traction for various reasons. They provide faster turnaround time and lower cost compared to face-to-face survey. In some areas with geographic constraints it is difficult to reach parts of the population.

In order an effective telephone survey to be completed many parameters should be known before it is launched: what is the appropriate share of landline versus mobile phone numbers? Are there different response rates among landline and mobile phone numbers? What is the most appropriate demographic stratification of the landline sample? Are there any issues with the telephone system? How should be the mobile phone sample stratified, etc.?

While many researches have worked on telephone surveys in various countries worldwide, a central knowledge sharing platform that addresses all these issues is currently lacking. These and many other questions, which are prerequisite for high quality telephone sampling frame, are answered by the Sample Solution's online application. The application serves as an open-source consultancy platform which provides open access to research for feasibility checks of telephone surveys together with sampling recommendations. Based on a previous experience as well as research done, the application advises the client regarding the best share of landline vs. mobile phone numbers, usage of specific dialing software (manual vs. predictive dialer), network issues and so on. Subsequently, it gives a recommendation for the stratification of the sample based on the demographic characteristics of the country, mobile providers' market penetration etc. All those parameters help the client to evaluate the choice for a telephone approach to a national representative study. Besides the standard info available, a report can be generated with all required variables (estimated sample size, replicating, country specific info and demographics) which enables researchers to quickly evaluate the feasibility or limitations of telephone surveys within a specific country or region.

Using Big Data to sample ethnic minorities in western European countries: Muslims

Carsten Broich, Sample Solutions BV
Maja Koceva, Sample Solutions BV

Abstract: Random Digit Dialing (RDD) sample consists of randomly generated phone numbers usually stratified by some specific parameters. Thus, RDD sample is usually used for national representative studies. However, very often there is a need of targeting specific audience where the pure RDD sample provides a low incidence rate and hence it is inappropriate due to high cost. Therefore, a need for other sampling methodologies derives. The target audience could differ in terms of age, gender, ethnicity etc. Nowadays, living in a multicultural society, studies upon ethnic groups are of immense importance. Given the political events in the Middle-East and the migration of Muslim populations from The Middle East to Western and Northern European countries, Muslim ethnic group in different non-Muslim countries gains a big attention. Many research agencies are interested in understanding the behavior of Muslim people.

Sampling Muslims populations within a population can be achieved in various ways. In recent years an onomastic approach using the listed sample or targeted RDD sample has been used which has led to large biases.

This paper outlines three main methodologies in target audience selection: list-assisted, geo-targeted RDD sample and social media approach. The list-assisted approach considers the residential listings containing phone numbers for a specific country. It narrows down the target population from the general population. Geo-targeted RDD sample focuses on large cities while small rural areas are excluded since it is difficult to detect the smallest administrative unit populated by Muslims. Social media approach includes a subset of generated cell phone RDD sample which is linked to various social media and public data sets (Big Data). Moreover, the paper provides the results of a pilot project of social

survey upon Muslims in various Western European. Consequently, the theoretical advantages and disadvantages are accompanied by those ones derived from the fieldwork for this quota sampling. Lastly, improvements proposals to overcome some of the disadvantages are being outlined. The proposed recommendations are aiming towards better target audience sampling in future using a true probability sample.

Session: Strategies to Reduce Nonresponse in Cross-cultural Surveys

Worth the effort? An examination of recontact strategies in f2f and telephone surveys

Steve Schwarzer, Pew Research Center
Martha McRoy, Pew Research Center
Patrick Moynihan, Pew Research Center

Abstract: Multiple contact attempts as well as varying calling dates and times are widely considered best practice to secure an interview and reduce the potential for systematic non-response bias. In fact, increasing the number of attempts to reach respondents has been shown to increase response rates and potentially lower non-response bias. However, if respondents who complete a survey after only a small number of attempts are different than those who require more effort, bias may be introduced by restricting the number of attempts by design.

Using Pew Research Center's 2017 cross-national Global Attitudes Survey, we will assess the non-response bias for both face-to-face and telephone designs by the number of contact attempts. In our analysis, we will focus on how "effort" affects the representativeness of the samples.
All Pew Research Center telephone surveys in Europe use a seven-call design, while f2f surveys are based on a three-call design. For our analysis, we will group respondents based on the number of attempts required to complete an interview, while paying particular attention to differences between early completes and latter ones.

Our analysis will focus on how estimated bias varies systematically as a function of fieldwork effort. In addition to contributing to our understanding of how effective fieldwork effort is in reducing nonresponse bias, our study also includes an assessment of how effort may affect substantive survey estimates.

Boosting response and minimising selection bias – recent experiments carried out in a push to web survey across Europe

Andrew Cleary, Ipsos MORI
Alex Cernat, University of Manchester
Peter Lynn, University of Essex
Yvette Boodhna, Ipsos MORI
Tanja Stojadinovic, Ipsos MORI
Sally Horton, Ipsos MORI

Abstract: In a push to web pilot survey commissioned by the EU Agency for Fundamental Rights, Ipsos conducted three experiments. Our focus was on finding solutions that would work in a cross-national context, given the heterogeneity of the EU-28 in terms of response rates and survey culture. Two of the experiments aimed to increase response and a third sought to minimise selection bias and improve data quality by preventing fraud.

To try to increase response, we tested the inclusion of a branded pen in the invitation letter - to add weight to the envelope and generate interest in its contents. This experiment was in three countries using address based samples where letters were addressed to 'the resident/the household'. In our second response boosting experiment, we implemented two different visual designs for the survey branding. This was included on the invitation & reminder letters, the envelopes and the survey website and was carried out across the EU-28.
To try to minimise selection bias and prevent fraud, we tested alternative methods for selecting respondents within households. Our aim was to encourage a random probability sample to complete the survey online and to prevent households from inventing respondents to claim additional incentives. This experiment was in 18 countries using address based samples. Households were asked to select up to two or three adults, to participate in the online survey. Two treatments were tested: (a) an invitation letter provided login details up front for two or three adults; and (b) an

invitation letter requested that any adult household member take part and on completion of the questionnaire online, an additional one or two adults were asked to take part (if there were two or more adults in the household). Within treatment (b), two methods for selecting the additional adult(s) were tested: household choice; and online random selection.

At the workshop, we will present the findings from each experiment as well as recommendations for the mainstage survey. We will include suggestions on how to instruct respondent selection when using address-based samples and postal contact for push-to-web surveys.

Session: Quality in Cross-cultural Surveys

Data-Quality Assessments Within a Cross-National Context: A Multidimensional Approach
Patrick Moynihan, Pew Research Center
Steve Schwarzer, Pew Research Center
Martha McRoy, Pew Research Center
Danielle Cuddington, Pew Research Center
Courtney Nelson, Pew Research Center

Abstract: Ensuring data quality in multinational face-to-face surveys is a challenge given the limited ability to monitor fieldwork staff's adherence to methodological protocols, ranging from proper household and respondent selection to careful respondent interviewing to accurate recording of all substantive and technical information. In our presentation, we explore the utility of a multidimensional approach to assessing data quality that relies on rigorous analysis of both substantive data (e.g., duplicate responses, item non-response patterns, straightlining and so on) and available paradata (e.g., time and date of interview, interviewer workload, geolocations and so on).

Using data from Pew Research Center's 2017 Global Attitudes Survey, we examine the 28 countries surveyed via face-to-face methods (including countries in Europe, Africa, Latin America and Asia) by applying an ex-post approach to identify potential problems within the survey data. We will present how specific indicators of survey quality can inform decisions about data use–in particular, paradata metrics of fieldwork timing and geolocations–while acknowledging the imprecision and unreliability that can be associated with each. Validation of these data-quality measures with others within the survey will be discussed. We will also present validation evidence culled by these same data-quality measures within corresponding discarded case files for each country surveyed. (These discarded case files include those cases that failed to meet the data quality standards of either the local vendor or by Pew Research.)

The paper discusses the sensitivity of each quality indicator to identify "suspicious" cases, the importance of local context in determining the potency of each quality measure, the overall robustness of the multidimensional approach and the limitations of relying on imperfect paradata. We will conclude our presentation by providing recommendations on how different quality indicators can be used effectively to evaluate the quality of survey data.

Improvements in survey quality over time – lessons learnt from Eurofound's pan-European surveys since 1995
Daphne Ahrendt, Eurofound
Sophia MacGoris, Eurofound

Abstract: Eurofound has been preparing and implementing comparative cross-national EU-wide surveys since 1995. That year saw the launch of the European Working Conditions Survey (EWCS) which has been fielded every 5 years since. In 2003, Eurofound launched the European Quality of Life Survey (EQLS), fielded four times so far and 2004 saw the first wave of the European Company Survey (ECS - three waves so far, next wave will be fielded in 2018-2019).
The surveys reflect Eurofound's commitment to answer to the needs of its tripartite stakeholders with a remit to conduct European policy-oriented research in the fields of living and working conditions.

Eurofound has a strong commitment to quality improvement; considering Eurofound surveys' impact at EU, international and national levels, it is important that data collected are sound, robust and of the highest quality. Information on quality is therefore made available to Eurofound stakeholders and the research community. For each

survey, information is gathered and published, assessing the data against a quality assurance framework based on the quality concept of the European Statistical System as developed by Eurostat, as well as other quality frameworks such as the Cross-Cultural Survey Guidelines (and the Total Survey Error Approach.

With preparations for the 7th EWCS commencing, the paper will present an overview of the current quality standards that Eurofound applies to ensure its commitment to quality improvement. The paper will focus in particular on how with each survey round Eurofound strives to learn new lessons for further improvement.

Some examples of quality improvements that will be highlighted in the paper include:

-Questionnaire translation: Eurofound has progressed from back-translation in the early rounds to adherence to the TRAPD method since 2013;

-Translatability assessment which was introduced in 2015;

-Enumeration instead of simple random-route probability sampling in countries where sampling frames are not accessible/available since 2010.

After each survey there now is an external quality assessment report which Eurofound publishes. Full methodological information is available to the research community.

Assessing Quality of Survey Processes and Outputs: The 4th European Quality of Life Survey
Julie de Jong, University of Michigan
Kristen Cibelli HIbben, University of Michigan

Abstract: The European Quality of Life Survey (EQLS), carried out by Eurofound, collects data from the general public across Europe on current life conditions as well as attitudes on a wide range of topics. The 4th wave of the survey was conducted in 2016 in 33 EU and EU Candidate countries. Contracted to do an external quality assessment of the 4th EQLS, we have taken a multi-pronged approach in our evaluation of processes and outputs. We first assess survey process quality using a Quality Assurance Plan developed by Eurofound that includes a number of specific quality indicators that are linked to the quality framework widely used across Europe (Relevance and Timeliness, Accuracy, Punctuality, Accessibility, and Coherence and Comparability). Using the concept of fitness for intended use, we next examine the quality of the key outputs from the survey, including the final source questionnaire in English and the final dataset. We then define a set of 3MC survey best practice guidelines based on the work of comparable 3MC surveys in the European context and recent contributions to the methodological literature. Using these guidelines, we assess the extent to which the 4th EQLS achieves these standards. We conclude our assessment with a discussion of our general findings, highlighting the strengths of the EQLS, and providing a series of recommendations for quality assurance and assessment for future Eurofound surveys and 3MC surveys more broadly.

<u>Session: Institute of Social Science Survey of Peking University Panel</u>

Telephone versus face-to-face interviewing: Mode effects on data quality: A case study of China Family Panel Study
Yan Sun, Institute of Social Science Survey of Peking University

Abstract: This report presents the experience of change in data collection mode from the face-to-face interviewing to face-to-face & telephone sequential mixed-mode in wave 4 of China Family Panel Studies (CFPS). We will also present some findings from assessing the impact changing mode have on data quality and the likely causes of any observed mode effects.

Data Quality Management Based on Computer-assisted Interviewing Method: A case study of China Family Panel Studies
Ding Hua, Institute of Social Science Survey of Peking University

Abstract: Computer-assisted Interviewing method (CAI) has greatly changed the data collection process in comparison with that of the traditional Paper-assisted interviewing mode (PAPI). The collection and application of para-data greatly

expand the quality control methods, improve the effectiveness of data quality management. This report will analyze the factors caused the measurement error, methods to reduce error and improve quality of data in the mode of CAI. The report will focus on the analysis of the quality management methods in China Family Panel Studies (CFPS) which based on the CAI mode and its effect on the improvement of data quality.

China Health and Retirement Longitudinal Study

Xinxin Chen, Institute of Social Science Survey of Peking University

Abstract: With the rapid aging of Chinese population, the problem of providing for the aged population is becoming increasingly important. At present, scientific studies of China's aging problems, like health, retirement and eldercare are still at an early stage, the greatest obstacle being a lack of sufficient micro, longitudinal data. The existing data tend to be small scale in parts of China, not collecting the breadth of data necessary for good social scientific analysis of ageing issues.

China Health and Retirement Longitudinal Study (CHARLS) is the first nationally representative survey of the older population that enables the study of the older population in China patterned after the Health and Retirement Study (HRS) in the U.S., English Longitudinal Study of Ageing (ELSA) and the Study of Health, Ageing and Retirement in Europe (SHARE). It aims to provide a high quality longitudinal micro-data set on a representative sample of individuals 45 and older to enable multidisciplinary study on issues of population aging in China. To ensure national representation, CHARLS survey covered 150 countries/districts, 450 villages/urban communities across the country. The national baseline survey was conducted in 2011-2012 on 17,708 individuals in 10,257 households, and follow-up surveys were conducted in 2013 and 2015. The response rate for the baseline survey in 2011 was 80.5%, and in 2013 and 2015 it was 88% and 87% respectively.

In light of the outdated household listings at the village/community level due to population migration, CHARLS designed a mapping/listing software (Charls-GIS) that makes use of Google-earth map images to list all dwelling units in all residential buildings to create sampling frames.
The core survey consists of the following sections: (1) demographics, (2) family structure/transfer, (3) health including biomarkers, (4) health insurance and health care utilization, (5) work, retirement and pension, (6) relative income, (7) family income, wealth and expenditures, (8) personal income, assets, (9) housing characteristics. All interviews were conducted using the computer assisted personal interview (CAPI) technology.

One special feature of CHARLS that is new to the HRS-type surveys is to collect detailed panel data from community-level informants (e.g., formal and informal community leaders). Basic community information is collected on, for example, land and its allocation and population, the most populous surnames and their numbers. More standard information is also collected, such as details about local infrastructure and public facilities such as roads, electrification, water and sanitation infrastructure and the availability of schools; health insurance and health facilities; pension and prices. The Policy Questionnaire collects details of social welfare programs such as pensions and health insurance, at the county level.

So far, the national data collected in 2011, 2013 and 2015 has all been released publicly, on the CHARLS website (www.charls.pku.edu.cn/en). By January 1, 2018, CHARLS has 23,112 registered users, and over 860 journal papers have been published based on this data set, 30% of which appeared in English journals.

Introduction of the Enterprise Survey of Innovation and Entrepreneurship in China

Dexi Li, Institute of Social Science Survey of Peking University

Abstract: Although private firms are a key engine of Chinese economic growth, primary data for private enterprises, in particular small and medium ones, are scarce. To fill the gap, Peking University has launched the Enterprise Survey of Innovation and Entrepreneurship in China (ESIEC). The questionnaire includes business history of entrepreneurs, start-up process, personal characteristics, operating situation, innovation, business network, and business environment of the chosen location.

After three pilots in Jiangmen (Guangdong Province), Baigou (Hebei Province) and Xiayi (Henan Province) in 2015-2016, Institute of Social Science Survey of Peking University carried out the survey in Henan Province during the summer in 2017. The survey is representative at the provincial level. More than 1,600 firms were successfully interviewed. The success rate is higher than the World Bank Enterprise Survey in China. Building upon the success of the Henan survey, our team will launch a national survey in the summer of 2018. The national survey will cover 161 counties in 25 provinces of China using the same sampling counties as China Family Panel Studies (CFPS), has surveyed about 100 households in each country over four waves. The total sample size is expected to be over 16,000 enterprises.

One key innovation of ESIEC is that it includes the entrepreneurship history of entrepreneurs and network of enterprises. In addition, ESIEC and CFPS share common sampling frame and some common module, which enable us to study what set entrepreneurs apart from common people. The third innovation is to link with administrative data, such as taxation. In doing so, we do not need to collect some sensitive information from the enterprises.

ESIEC mainly hires students from Peking University and other universities as numerators. Students at Peking University can get course credit by participating in ESIEC. More importantly, they can gain valuable field experience, learn more about the real Chinese economy, and improve their communication skills. The biggest challenge in the survey is that it is hard to locate some firms which have wrong registered contact information, have closed doors, or have moved elsewhere.
We have developed several practical methods to look for enterprises. First, we can get the detailed address of sample enterprises on Baidu Map before conducting survey. Second, we can take full advantage of the reputation of Peking University among the public to get help from local people. Third, we use a snowball approach by asking the interviewees to help us contact the next one. Fourth, we ask the local government officials for help.

Session: Paradata Use in Cross-cultural Surveys

Use Item-Level Paradata for Multi-Country Comparison and Interviewer Effect Assessment

Shulin Jiang, Johns Hopkins School of Public Health, Department of Population, Family and Reproductive Health, Gates Institute
Linnea Zimmerman, Johns Hopkins School of Public Health, Department of Population, Family and Reproductive Health, Gates Institute

Abstract: Paradata is widely used to monitor interview quality and manage data collection as automatically generated item-level paradata can indicate potential problems with survey items. Studies using large-scale, multi-country sources of paradata from interviewer surveys are limited, however, even though keystroke paradata provides abundant details with little extra cost.

PMA2020 is a survey platform that uses smartphone assisted personal interviews (SAPI) to monitor key health and development indicators. It collects a nationally representative sample of data from households, females and health facilities annually in 11 countries. Surveys are carried out by female data collectors - resident enumerators (REs), who are local women from or near the respective enumeration areas. PMA2020 uses Open Data Kit Collect (ODK), an open source software that facilitates data collection via mobile-assisted data platform. PMA Analytics runs automatically with ODK when REs conduct interviews on smartphones. Interview process and user interactions, such as time lapse, error message occurrences, screen visits, and answer changes, are time-stamped and recorded in a log. PMA Analytics data is keystroke data generated from survey submissions and their associated logs.

The advantage of PMA Analytics is its large-scale item-level keystroke paradata. The item-level paradata provides rich details to reflect field activities. The multi-country PMA2020 surveys make it feasible to cross-compare aggregated paradata. Data for this presentation was collected from November 2016 to October 2017 in ten countries: Democratic Republic of the Congo, Burkina Faso, Cote D'Ivoire, Niger, Ghana, Nigeria, Ethiopia, Kenya, Uganda, and India.

In this presentation, we will first compare the descriptive analysis results of PMA Analytics, including interview time and key error message occurrence across the ten program countries. We will summarize the similarities while investigating discrepancies in different settings, which provides insight and guidance for training and field activities. Then we will

assess whether there is a change in interviewer effects from the beginning to the end of data collection within one round and compare it across settings, controlling for item and respondent characteristics.

Using paradata to monitor interviewers' instrument navigation behavior and inform instrument technical design: Case studies from a national household surveys in Ghana and Thailand
Yu-chieh Jay Lin, University of Michigan
Gina-Qian Cheung, University of Michigan

Abstract: Many computer-assisted personal interview (CAPI) software captures paradata (i.e., empirical measurements about the process of creating survey data themselves), computer user actions, including times spent on questions and in sections of a survey (i.e., timestamps) and interviewer or respondent actions while proceeding through a survey. In these cases, the paradata file contains a record of keystrokes and function keys pressed, as well as mouse actions. These paradata files are transmitted along with the survey data and can be used for quality assurance checks and reporting, particularly when interviews are not audio recorded.

This presentation uses data from two sources: (1) the Ghana Socioeconomic Panel Study in collaboration with Yale University, University of Ghana, and University of Michigan; and (2) the Evolution of Health, Aging, and Retirement in Thailand in collaboration with National Institute of Development Administration and University of Michigan. Both studies utilize unique team management and travel structures, and have a complex instrument design. In addition, interviewers are allowed to interview respondents within the same sample unit without any particular order and to switch among varied interviewing components in a flexible fashion. This presentation focuses on the analysis of keystroke data to monitor interviewers' instrument navigation behavior.

We first reconstruct and categorize interviewer navigation patterns such as mid-section break-offs through varied interviewing components. These navigation patterns are then inspected for predictive power against data quality indicators such as response changes and non-response. Subsequently, we analyze interviewer, household, and geographic characteristics and identify interviewer quality control metrics (e.g., interview length) to determine if interviewer behaviors and interview efficiency can be predicted by interviewer's overall team behavior or household characteristics, among all other information available. Finally, we will present how these analyses of paradata can be practically applied to improve interview efficiency and data quality of interviewer administered surveys.

Wednesday, March 28th

Improving Cross-National Survey Research by More Coordination and Collaboration across Comparative Survey Research Projects

Making Comparative Survey Research More Comparative: Collaboration and Coordination across Cross-National, Survey-Research Programs
Tom W. Smith, NORC at the University of Chicago

Abstract: Traditionally, large-scale, cross-national programs have essentially operated separately with little attention to other comparative projects. This isolationism is unfortunate since greater collaboration and coordination would benefit each program individually and also advance comparative research collectively. First, managing cross-national projects is a challenging endeavor. Fortunately, tools are being developed to track the development of questionnaires, the translation of questions, and the management of fieldwork. Rather than being designed to service individual projects, these are generic systems that can be utilized across projects. Similarly, the expanding use of DDI-compliant metadata both improves data collection at the front end and data documentation at the back end. Moreover, it makes comparisons across surveys much easier. Second, the sharing of methodological approaches and lessons learned lowers learning curves across projects and contributes to the cumulative development of experienced-based comparative methodology. Third, while separate projects will (and should) continue to have their own substantive focus, some sharing of items (e.g. demographics) would both facilitate questionnaire development and assist comparative analysis. Finally, a number of other advantages of collaboration and coordination will be discussed.

Considering cooperation beyond Europe. Opportunities and challenges for the European Social Survey
Rory Fitzgerald, City University of London

Abstract: The European Social Survey has been considering if, when and how to develop links with existing survey programmes beyond Europe in order to increase substantive comparisons on relevant issues and encourage joint work on methodological challenges. Whilst this is appealing there is a strong sense that any cooperation should strengthen the 'defence of rigour' in comparative survey research and not undermine it simply to generate data. How to high quality, equivalent data in a more global perspective is a big challenge that requires further consideration.

A workshop was held in London at European Social Survey HQ in the autumn of 2017 bringing together PIs from major cross-national regional social surveys and those who lead major national surveys with a cross-cultural dimension.

This paper will discuss the outcome of that workshop focusing on the possible advantages of greater collaboration/coordination among cross-national projects –substantively and methodologically and possible the disadvantages. Some initial examples of cooperation with other surveys will be discussed to outline how cooperation might be achieved in the future.

Comparative Studies: A Proposed Agenda for Future Collaboration
Lars Lyberg, Inizio
Kristen Cibelli, University of Michigan
Julie de Jong, University of Michigan
Timothy P. Johnson, University of Illinois at Chicago
Beth-Ellen Pennell, University of Michigan

Abstract: This presentation will build upon the discussion held last year at the end of the CSDI workshop regarding a possible AAPOR/WAPOR Task Force on quality issues in comparative survey research. We will also draw upon discussions held this past fall among a number of large cross-national studies, hosted by the European Social Survey at City University London. With this as background, we will present a proposed agenda for collaboration with a primary focus first on the Task Force 's goals, deliverables and timeline.

Session: Response Quality

Measurement Error in Proxy Measures of Key Survey Variables across Different Languages
Emilia Peytcheva, RTI International

Abstract: Surveys increasingly need auxiliary data that are strongly correlated with key survey variables in order to adjust for nonresponse bias. Such data are rare in household surveys, but when the design allows, proxy survey variables can be collected. Proxy measures however, are subject to measurement error and its amount can vary depending on the language of survey administration.

We imbedded proxy measures on the household and the selected respondent(s) in the screening instrument of the 2015-2016 California Health Interview Survey (CHIS). We included two key health questions (health conditions, asked at the selected respondent(s), and public health insurance, asked at the household level) to potentially use in the estimation, reduction, and adjustment for nonresponse bias. An initial evaluation of measurement error in these questions revealed significant underreporting of heath conditions, and significant over reporting of public health insurance in the screener as compared to the main interview.

This paper examines whether the proxy health-related variables exhibit differential measurement properties across six CHIS languages (English, Spanish, Chinese, Korean, Tagalog, and Vietnamese). For example, there are strong theoretical reasons to believe that measurement error would be lower in Spanish proxy reports as compared to English, due to familism (strong and close family relationships) and well-studied differences in reporting of health conditions among Hispanics (e.g., the immigration paradox). Findings from these analyses will have consequences not only for the utility of proxy measures for nonresponse adjustments, but also for decisions related to targeting and prioritization of cases.

Identifying of and dealing with item nonresponse in open-ended questions in a cross-national context
Katharina Meitinger, GESIS - Leibniz-Institute for the Social Sciences
Lars Kaczmirek, AUSSDA - The Austrian Social Science Data Archive

Abstract: Open-ended questions are an important addition to closed items in the toolkit of social scientists and they can provide valuable information that closed items often cannot provide (Emde & Fuchs 2012). Due to the elevated cognitive effort associated with open-ended questions, the response burden increases (Bradburn 1978) which makes open-ended questions potentially more affected by nonresponse (Barrios et al. 2010). In addition, previous studies found that the level of nonresponse is affected by factors such as education or age (Andrews 2005; Miller & Lambert 2014; Zuell et al. 2014). Most of these studies are single-country studies with special populations such as employees or students.
One promising strategy to reduce this nonresponse is the implementation of motivational sentences (e.g. Oudejans & Christian 2011; Zuell et al. 2014). The tool Evalanswer can automatically detect nonresponse in open-ended questions in multilingual (English, German, and Spanish) web surveys and attempts to convert initial nonrespondents into substantive respondents by repeating the open-ended answer alongside a motivational sentence (Kaczmirek et al. 2017).

In this presentation, we want to discuss the prevalence of nonresponse in open-ended questions in two Web surveys (Study 1: N=2,685; Study 2: N=2,689) that were implemented in five countries (Germany, Great Britain, Mexico, Spain, and the U.S.) in 2014. Both studies were opt-in panels with respondents selected by quotas for age (18–30, 31–50, and 51–65), gender, and education (lower and higher). In total, we analyze the nonresponse behavior for 29 open-ended questions and distinguish between different types of nonresponse (hard nonresponse and soft nonresponse, such as don't know responses, refusals or unintelligible responses). We will show cross-national differences in nonresponse behavior as well as the impact of sociodemographic factors and survey features on nonresponse. In addition, we will address whether and which initial nonrespondents can be converted by an automatic repetition of the open-ended question alongside a motivational statement.

Questionnaire Answer Agreement: A Method to Detect Lurking Problems and Increase Data Quality
James Pringle, Johns Hopkins University Bloomberg School of Public Health

Abstract: The error rate in data recording during the administration of a survey is difficult to measure in the analysis phase because it is not possible to compare values in the dataset with responses given by a respondent. Mobile devices reduce data entry errors during survey data collection, but they also present challenges. Data collectors must learn a new technology, often with minimal related experience. User interfaces and small screen sizes may make data entry and app navigation rely on very similar physical movements. For example, swiping on the screen to navigate through the survey app may inadvertently change a previously recorded answer.

Performance Monitoring and Accountability (PMA) 2020 conducts surveys in several low and middle-income countries using the Open Data Kit platform. In October 2017, PMA2020 helped train 15 supervisors in Ghana for an upcoming primary health care facility survey. They broke into three groups and visited three facilities separately and administered the survey together as a group. The recorded responses of fellow group members agreed only 88% of the time on average, even though they all heard the same responses and made the same observations within their groups. The questions with the most disagreement were reviewed, and this led to valuable corrective training.

We will explore the error rate in data recording on mobile devices in PMA2020 trainings by conducting a simulated data collection activity in Burkina Faso, India, Kenya, Niger, and Nigeria with over 100 enumerators and 25 supervisors. As part of training, the enumerators and supervisors will administer a survey in groups to people selected and given a standard script. Recorded answers will be compared against the script to calculate correctness. Within-group agreement will be calculated question by question. We will compare results across countries and between supervisors and

enumerators. We will also describe the difference between correctness and agreement and determine the adequacy of majority rule as a replacement for correctness when the truth is not known during analysis. We will identify and characterize questions that have the most disagreement, provide training, and quantify improvement after administering the survey again.

Discovering data recording errors and providing corrective training is one way to improve data quality. The results from this study will help inform future survey trainings and heighten awareness of errors due to data recording.

Session: Interviewer Effects

Impacts of interviewer's personal characteristics and team composition on household survey: an empirical analysis based on China Household Finance Survey

Qing He, Survey and Research Center for China Household Finance, Southwestern University of Finance and Economics
Xin He, Survey and Research Center for China Household Finance, Southwestern University of Finance and Economics

Abstract: Many studies have looked at the interviewer effects in the household survey, such as interviewers' gender, wearing, religious attitudes, and so on. However, there are relatively few research studied the impacts of interviewer's personality traits and the composition of survey team on household survey quality. During the 2017 wave of China Household Finance Survey (CHFS), we collected very detailed information about the interviewer, including demographics, individual characteristics, and family background information. In addition, we applied the Catell 16 Personality Factor Test (16PF) to evaluate the type of interviewer's personality.

The discussion will mainly focus on the impacts of interviewer's characteristics, especially the personality traits, on the survey quality, which is measured by indicators including refusal rates, item response rates, response errors, and so on. Furthermore, the fieldwork was implemented in the form of survey team. We will discuss the relationship between the composition of a survey team and survey quality. At last, we will discuss the effects of quality control intervention on data quality.

Interviewer Effects and Social Desirability: A Cross-Cultural Analysis of Face-to-Face Surveys

Martha McRoy, Pew Research Center
Courtney Nelson, Pew Research Center
Danielle Cuddington, Pew Research Center

Abstract: With face-to-face interviewing, responses and results can be influenced by many confounding factors, including interviewer effects. While the interviewer's presence can introduce a social desirability bias of its own accord, the interviewer's demographics – including gender, ethnic or racial background, education, and age – can create an additional interviewer effect, adding to this bias. This is concerning given that an interviewer's traits can bias the results and findings of the study.

Interviewer effects can have an extreme impact on cross-cultural surveys, particularly when an interviewer's characteristics and the survey context overlap. For example, an interviewer's gender may influence how the respondent answers certain questions in some countries, while in others the interviewer's gender is paramount to garnering survey participation. Therefore, to better examine interviewer effects cross-culturally, details about the interviewers themselves must be collected. However, local vendors, or even interviewers themselves, can be hesitant to provide personal information, making the analyses of interviewer effects challenging.

Using Pew Research Center's 2017 cross-national Global Attitudes Survey and previous waves, we will assess the impact of interviewer characteristics on responses to sensitive, substantive questions on topics such as politics, religion, and social attitudes. The analyses will focus on the impacts of interviewer-respondent genders, levels of education, ages, and the interviewer's level of experience. In addition to examining the social-desirability of responses, changes in item-nonresponse will be reviewed. These analyses will also be used to assess whether interviewer characteristic effects vary across subject matters and countries.

Session: Mixed Methods and Survey Modes

Using a responsive design to conduct a survey under unclear outcome conditions

Tobias Gummer, GESIS - Leibniz Institute for the Social Sciences
Pablo Christmann, GESIS - Leibniz Institute for the Social Sciences
Sascha Hähnel, GESIS - Leibniz Institute for the Social Sciences
Christof Wolf, GESIS - Leibniz Institute for the Social Sciences

Abstract: Conducting large-scale face-to-face surveys has become increasingly challenging in most countries due to rising nonresponse rates and, thus, increased costs. Recent developments in split questionnaire designs might make it feasible to conduct these General Social Survey-type studies in cheaper self-administered modes. Split questionnaire designs aim at dividing one questionnaire into several shorter questionnaires. However, applying this method to the problem at hand rests on the assumption that we are able to collect high quality data with split questionnaires in self-administered modes.

In recent years much effort has been invested in improving face-to-face surveys. As a consequence little is known about how (mixed-mode) self-administered surveys should be designed and fielded in countries like Germany. In our case, we identified a lack of empirical evidence on effective incentive strategies and on designing the sequence of mode choices for respondents. Setting up adequate pre-testing was deemed not viable due its costs. Consequently, we were in severe need of a survey design that allowed fielding the survey under unclear outcome and cost conditions. To cope with the lack of empirical evidence, we implemented a responsive survey design to adjust the data collection protocols based on knowledge that was gained during data collection. Consequently, we designed our mixed-mode (MAIL and WEB) self-administered survey as a two-phase process. Our probability-based gross sample was randomly split into two subsamples. The first subsample was then randomly split into 2×2 experimental groups across which we varied incentive strategies (prepaid vs. postpaid) and sequence of mode choice (simultaneous vs. sequential). After fielding the first sample with the different experiments, we evaluated their outcomes in terms of nonresponse, nonresponse bias, and cost. Based on this knowledge, we then re-adjusted our data collection protocols and started fieldwork for the second sample. We will present findings from our experiments in the first phase of the survey and discuss how we used outcome and cost indicators to operate the responsive survey design.

Questionnaire split following a matrix design to run long surveys on the web: a viable solution for international comparative surveys? A real-life experiment with the European Values Study

Michèle Ernst Stähli, FORS (Swiss Centre of Expertise in the Social Sciences)
Alexandre Pollien, FORS (Swiss Centre of Expertise in the Social Sciences)
Michael Ochsner, FIRS (Swiss Centre of Expertise in the Social Sciences)
Dominique Joye, University of Lausanne
Patricia Milbert, FORS (Swiss Centre of Expertise in the Social Sciences)

Abstract: Relevance & Research Question: The challenge of long surveys on the web
In the tradition of the social sciences, many high-quality general population surveys consist of long questionnaires, lasting up to one hour or even longer (for example ESS, SHARE, EVS, to mention some international projects only). Recently, eroding coverage (in the case of telephone surveys), declining contact and response rates, as well as rising costs push surveyors to consider the web mode to replace interview modes such as telephone and personal interviews. But how can such long surveys be successfully fielded on the web, where short questionnaires are recommended?

Methods & Data: an experiment with the EVS
Several countries of the EVS project (European Values Study) have decided to run in parallel to the traditional face-to-face interviews a separate web version, where the one-hour questionnaire has been split into several modules, each respondent thus answering only to a part of the whole survey. In Switzerland, the option of a paper version is offered to

non-respondents at the second reminder (push-to web design). In a second phase, the respondents were invited to answer the complementary part, aiming so at full responses. In addition, a control group received the complete one-hour questionnaire. Moreover, the Swiss sampling frame offers socio-demographic background information on all sample units, allowing for precise checks of the final sample composition by experimental group.

Results: innovative survey design and insights on feasibility
We will present the different designs and provide insights on the feasibility of fielding long web(-paper) surveys, compared to the face-to-face mode, either as full one-hour surveys or following a matrix design. The Swiss data show also the relative quality in terms of response rates, break-offs and sample composition of the different designs tested, varying length and mode.

Added Value:
While usual web survey practice concentrates on short surveys, the feasibility of long and very long web surveys has not yet been explored in a systematic way. The EVS online project provides valuable insights into issues, potential options and feasibility for all long surveys considering moving to the web mode.

Using In-depth interview at interviewees' homes with nationwide quantitative survey for understanding Southeast Asian lifestyles

Midori Aoyagi, National Institute for Environmental Studies
Vu Huy, Vietnamese Academy of Social Sciences, Institute for Regional Sustainable Development
Aya Yoshida, National Institute for Environmental Studies
Aya Yoshida, National Institute for Environmental Studies

Abstract: For the South East Asian Lifestyles study, National Institute for Environmental Studies, Japan teamed up with the Vietnamese Academy of Social Sciences, Institute for Regional Sustainable Development to carry out surveys for analyzing sustainable consumption of Vietnamese people. Our study focuses not only consumption itself but extends to understand how people in Vietnam see their future. We used mixed method, combining the in-depth interview at interviewee's home, and a nationwide quantitative survey, using multi-stage sampling with "maximum variation approach" proposed by International Labor Organization LO (ILO, 2009) for this study.

As Vietnam is North-South long narrow country, we divided regions into six, such as North, Central, and South (Mekong Delta), and each has coast side and mountainside, roughly. For our in-depth interview, we chose three sites, Ha Noi (for the North), Hoi An (for the Central), and Ho Chi Minh (for the South). In each site, we chose both urban and rural clusters (communities). We had 30 interviews in total. The 4 to 6 researchers of NIES and VASS-IRSD visited local government office (first district level, then ward or commune level), to ask for interviews support, then officers or the head of the commune introduced us their candidates. On average, interviews lasted 2 hours.

As for the quantitative survey, we used multi-stage sampling with "maximum variation approach". The reasons are as follows. Vietnamese population is not homogenous since it has many ethnic groups. Secondly, the population is unevenly distributed across the country with high level of concentration in large cities, such as Hanoi and Ho Chi Minh. Furthermore, there are huge gaps between rural and urban people in terms of lifestyles. Multi-stage sampling with "maximum variation approach" in this case would ensure better representativeness with reasonable costs. We chose six provinces (one from each region), then 54 sampling points (9 from each province, "ward" in towns, "commune/village" in rural areas). Then 22 to 24 households in each sampling point are selected using systematic sampling from the list provided by the head of the commune. Respondents who are between 18 and 75 years old in each sampling point were chosen by Kish grid at the respondent's home. Finally, our interviewers approached approximately 1200 respondents in total in nationwide. This part of our survey is currently on going at the time of writing this abstract. We will report the results including quantitative part of our survey on March.

What Mobile Phone Survey Mode Produces the Most Representative Data? A Comparison of SMS, IVR, and CATI in Nigeria

Leenisha Marks, RTI International
Charles Lau, RTI International

Alexandra Cronberg, Kantar Public
Ashley Amaya, RTI International

Abstract: Mobile phone surveys in lower income countries have become increasingly common in recent years. The most common modes are computer-assisted telephone interviewing (CATI), interactive voice response (IVR), and short message service (SMS, or text messaging.) Each mode has strengths and weaknesses. CATI might achieve higher response rates because of live interviewers, but is time consuming and expensive. SMS surveys are inexpensive and fast, but may suffer from non-response because of illiteracy, distrust of surveys, language issues, among other reasons. Because IVR uses automated voice recordings, it can include illiterate people – but may experience substantial non-response errors like SMS.

There is virtually no research comparing the representativeness and bias of CATI, IVR, and SMS. To address this gap, this study compares the representativeness of CATI, IVR, and SMS surveys in Nigeria. We conducted random digit dialing (RDD) mobile phone survey using three modes: CATI (n = 3,785), IVR (n = 1,818), and SMS (n = 2,759).

Using these data, we investigate three questions about the representativeness and bias of mobile phone surveys in Nigeria. First, what are response rates in CATI, IVR, and SMS surveys? Second, how representative are the respondents from each mode? We investigate this question by comparing the socio-demographic composition of each mode with a gold standard, the 2013 Nigeria Demographic and Health Survey. Third, to what extent do weights correct for bias in each mode? To answer this question, we post-stratify the sample to population totals. We then apply the weights and compare estimates of voting behavior between the mobile phone survey modes to official statistics. By answering these three questions, we aim to provide practical guidance in selecting the optimal mobile phone survey mode in Nigeria and other lower income country.

Session: Data Collection Challenges

A Mechanism For Understanding The Effects Of Repeated Survey Research On Communities: Findings From A Qualitative Study Of PMA2020 Surveys In Burkina Faso

Shani Turke, Johns Hopkins Bloomberg School of Public Health (JHSPH), Gates Institute
Abigail Greenleaf, JHSPH, Gates Institute
Fiacre Bazié, Institut Supérieur des Sciences de la Population (ISSP)
Nathalie Sawadago, ISSP
Caroline Moreau, JHSPH
Georges Guiella, ISSP

Abstract: While panel conditioning has been widely studied among longitudinal survey research conducted in high-income countries, less is known about the phenomenon among research conducted in low and middle-income settings. In addition, the existing body of literature relies largely on quantitative methods to elucidate the magnitude of panel effects on data quality, but provides little information on the underlying mechanisms for respondent or community-level change over time as a result of the survey. The current research aims to contribute to both gaps in the literature by qualitatively exploring the community-level effects of PMA2020 surveys in Burkina Faso, a cross-sectional, nationally representative survey collecting data on family planning and reproductive health behaviors over repeated rounds within the same enumeration areas. We conducted sixteen in-depth interviews and four focus group discussions with resident enumerators, women recruited from the sampled communities in the survey, to better understand how community perceptions of the study change over time, and what effects this may have on interview quality. Data were analyzed using Atlas.ti, following an iterative approach, gradually moving from emic themes to etic theory development. Through this qualitative analysis, we describe how communities perceive the study at introduction and after multiple survey rounds, detailing a proposed mechanism of change at the community level. We also discuss how these changes may influence interview quality. Resident enumerators see difficulties with study introduction as due to community-level distrust and misunderstanding of the study's purpose, in addition to their own lack of experience as interviewers. Over time, enumerators build rapport and integrate into their communities while also improving their professional skills. Concurrently, community members begin to see enumerators as sources for family planning information, while also sometimes exchanging information amongst themselves about the study. After multiple rounds, enumerators describe

communities that are more trustful and have greater understanding of the study, often associating this change with richer and more complete respondent reporting. At the same time, expectations of immediate benefit remain and respondents and community members alike still see enumerators as family planning resources, both perspectives that may negatively affect respondent-interview interactions over time. While we cannot make direct inferences on how this mechanism of change impacts data quality, our findings do provide valuable insight into how and why repeated exposure to survey research may influence communities in West Africa over time.

Treading through India's data challenges
Prerna Mukharya, Outline India

Abstract: Given Outline India's focus with data as the epicenter of research, the abstract will revolve around the following listed topics on the CSDI website –
1. Questionnaire development and testing
2. Translation, adaptation and assessment
3. Data collection challenges and solutions
4. Innovative use of technology

Despite the fact that ground level data is foundation to policy making, in the development sector, little importance has been attributed to field work which is the basis of data collection. Engagement of researchers in field work and field training is limited which also limits them from monitoring the data collection process which jeopardizes the data quality.

Our presentation is an attempt to address this gap by discussing the challenges faced in data collection, documenting learnings from the field and underlining the importance of a thorough phase of pre-testing survey tools in relevant settings to develop a contextually pertinent study design and identify anticipated and unanticipated inconsistencies. Additionally, given India's diversity and plethora of languages and dialects in use, questionnaire development and testing is a core component to ensure cross cultural comprehension. It is also crucial to consider nuances and innuendos that vary across regions in order to design questions and tools accordingly which feeds into data analysis.

In terms of unanticipated consistencies, in our own experience, we were refused and turned away from conducting surveys in a village in the state of Rajasthan because some unknown NGO had collected data there a few months ago with an unfulfilled promise of transferring money to their bank accounts. On the contrary, during our recent field experience in Gujarat, due to the community's exposure to different NGOs, their participation in our research was more forthcoming.

Therefore, in this presentation, we would like explore the importance of executing fieldwork through our experiences, documenting processes and errors and developing a nuanced understanding of research that is true to its scientific approach but is still executable on ground. The presentation will briefly touch upon the use of technology in social sector research and fieldwork that Outline India employs, and the ways in which the confluence of human capital and technology interacts with the processes of knowledge production.

<u>Session: Managing Production and Quality issues in Large Scale International Assessments</u>

Dealing with a rare target population: out-of-school youth survey of PISA-D in Paraguay
A. Figueredo, *Ministry of Education and Science (Paraguay)*
Veronica Heilborn, Directorate for Evaluation of Quality of Education, Ministry of Education and Science (Paraguay)
S. Suarez, Ministry of Education and Science (Paraguay)

Abstract: After the advent of democracy in 1989, Paraguay has made important efforts to improve its educational system, expanding coverage, implementing educational reforms and increasing investment in teachers, equipment and infrastructure of public educational institutions. However, it maintains important challenges regarding keeping children and young people in the educational system, as well as the education´s quality it provides.

In 2015, the Ministry of Education and Sciences signed a Participation Agreement with the OECD for PISA for Development, and later during 2016 the country agreed to participate in the new Component of the Program: Strand C

aimed at assessing the competences of 14 to 16 year-old out-of-school young people as well as those that have several years of lag, in 6th grade or lower grades.

Strand C's design defines a house survey methodology, with a self-administered test on Tablet support, and background questionnaires for young people and another one for parents or caregivers. The Items Response Theory used by PISA tests, required a minimum of 1200 cases for the field trial and 2700 cases for the main study. The study evaluates the same competencies and skills as PISA tests, but only in Mathematics and Reading domains.

In Paraguay, cognitive items were adjusted for computer- based assessment, and a household survey was designed covering 137 primary sampling units, throughout all eastern region. The design involved a random sample in four strata: high/low concentration of target population, and rural/urban areas. The operation was deployed in the field for 11 weeks. In addition, tests were applied to students under 7th grade in selected Schools.

The implementation had to face several obstacles derived from: the application of the study design to the data availability and national context, the management of data collection through computer equipment, the field strategy used for data collection, the data monitoring and validation processes, and other directly related issues on being an innovative Strand, implemented for the first time by PISA-D International Consortium.

During Field trial, learned lessons and alternative solutions for Strand C´s main survey, scheduled for 2018, have been identified.

Innovative approaches in survey implementation – overcoming data collection challenges in U.S. PIAAC

Nina Thornton, Westat
Leyla Mohadjer, Westat
Tom Krenzke, Westat

Abstract: Sponsored by the Organization for Economic Co-operation and Development (OECD), the Programme for the International Assessment of Adult Competencies (PIAAC) measures literacy, numeracy, and problem solving in technology-rich environments. Twenty four countries participated in the first round of PIAAC in 2010; an additional nine countries joined PIAAC (Round 2) in 2012 and five countries have joined Round 3 with data collection in 2017-2018. The national samples include about 5,000 in-person interviews in households. Eligible adults aged 16 to 65 living in households (or noninstitutional group quarters) are administered a background questionnaire followed by a self-administered assessment that is either computer-based or paper-based. The results of Round 1 and 2 are available on http://www.oecd.org/site/piaac/publications.htm.

The main goal of PIAAC, as in other international large-scale assessment surveys, is to collect standardized data across all participating countries. The data needs to be representative of the target population, of the highest quality, and comparable across countries. Managing data collection and quality in any survey is challenging but even more so in face-to-face surveys involving individual assessments in households.

The U.S has participated in all three rounds of PIAAC data collection. We present experiences gained in dealing with various challenges encountered with household data collection, strategies implemented to overcome these challenges, and the lessons learned that will enable us to better plan for future rounds of data collection. We also present innovative technologies through the use of paradata to monitor field work and strategically enhance efficiencies in production.

During Round 3 data collection, we debuted the use of survey dashboards in order to maximize effective field management – use of dashboards facilitated early detection of cases with invalid data, more precise management of household contact attempts, times, and travel routes to maximize production, and real-time monitoring of hours worked by interviewers. The system provided at-a- glance summary graphics related to production and the ability to investigate at the case level if needed.

In addition, we used adaptive data collection strategies through the application of paradata to develop sample monitoring dashboards which displayed the status of key response indicators; this allowed case prioritization for interviewers in order to achieve a balanced sample, provided guidelines for the number of contact attempts, and allowed projections on whether the target sample size would be achieved.

These collective measures enabled field data collection to progress in a more streamlined manner that was efficient, cost-effective and achieved survey objectives.

Looking back and looking forward: From PIAAC and PIAAC-Longitudinal in Germany to PIAAC Cycle 2

Anouk Zabal, GESIS
Beatrice Rammstedt, GESIS

Abstract: In the first cycle of the Programme for the International Assessment of Adult Competencies (PIAAC), key competences of adults (16-65 years) were measured in over 40 countries. Germany participated in the first round of data collection in 2011/2012. To ensure a high quality of the data and the equivalence of measurement across countries, the PIAAC Consortium established an elaborate set of standards and guidelines for the national implementation of the project. Reconciling these internationally defined standardizations with national constraints was a challenging – and also rewarding task – for the national project managements. This presentation will reflect on limitations to the straightforward implementation of certain standards and on local adaptations in the German context. In addition, we will look to experiences from the German PIAAC-Longitudinal project (PIAAC-L), which was created as a national longitudinal extension to the PIAAC, and discuss some possibilities for improving data collection processes and the quality of fieldwork for the upcoming second cycle of PIAAC in Germany.