

ESA RN 21 Midterm Conference

“Data quality in quantitative research”

Thursday, October 13rd

8:00– 9:00 Registration

9:00-9:15 Conference Opening

Wolfgang Aschauer (ESA RN 21 coordinator)

Savvas Katsikides (Dean of Faculty Of Social Sciences And Education - University of Cyprus)

Nektarios Partasides (President of the Cyprus Sociological Association)

Iasonas Lamprianou (Local Organizer - University of Cyprus)

9:15-10:00 **Keynote Eldad Davidov** (University of Zürich, Switzerland)

Challenges in the analysis of cross-cultural data

I am going to begin with shortly discussing potential problems in the analysis of (particularly but not only survey) cross-cultural or longitudinal data, addressing in particular the risk of comparing ‘apples with oranges’ and possible sources for such problems. Next I will mention traditional statistical techniques to test for comparability of data across groups and expand them to newer methods suggested in the methodological literature to address potential (non-)comparability problems. These methods either identify or try to explain non-invariance. I will question whether some of them may eventually even solve non-comparability problems. Finally, I will relate this methodological discussion to recent suggestions especially in the political science literature to ignore problems of non-comparability, its background and the dangers associated with such a strategy when trying to conduct a meaningful analysis of comparative data.

Chair: Wolfgang Aschauer (University of Salzburg, Austria)

Cross-national surveys: Challenges to achieve equivalent results

Chair: Eldad Davidov (University of Zürich, Switzerland)

Anna Domaranska

National Academy of Sciences of Ukraine, Institute of Sociology, Kiev, Ukraine

anna_doma@list.ru

Testing the measurement equivalence in cross-cultural research:

Address the problem of nonequivalence

It's common knowledge that supporting the presence of equivalence is an absolute necessity for meaningful international comparison. However testing the measurement equivalence can be highly problematic when applied to widely diverse cultural groups. According to B. Byrne, F. van de Vijver (2010) the assumption that all samples derive from the same population is violated as participating countries of survey projects represent different parts of the world. Also modification indices, which usually work well for detecting sources of nonequivalence with a small number of groups, are performing badly when the number of groups is large and diverse. Despite notable progress in the subject field, it is highly likely to encounter evidences of nonequivalence the nature of which is unclear. Thus a key question arises what can be done when equivalence is not supported by the data?

So far the literature suggests several strategies to deal with measurement nonequivalence. Firstly, partial invariance is regarded as acceptable when at least two indicators per construct are equal across groups. Secondly, finding the groups of countries where measurement equivalence holds or clustering countries in some meaningful manner for instance according to affluence, religion, global region, important contextual variables or use an established classification system. Lastly, trying to explain the absent of equivalence by means of a multiple indicators multiple causes (MIMIC) model or a multilevel structural equation modelling (MLSEM) (Byrne & van de Vijver, 2010; Davidov et al. 2012; Davidov et al. 2014).

The purpose of this study is discussing and illustrating these strategies in order to find optimal sequence of research steps for conducting the meaningful comparison in cross-national studies. Based on the ISSP Social Inequality Module collected in 2009 within 38 countries, multiple group confirmatory factor analysis (MGCFA) has been used for testing the measurement equivalence.

Magdalena Burdach

Jolanta Perek-Białas

Warsaw School of Economics/Jagiellonian University, Kraków, Poland

jperek@sgh.waw.pl

Evaluation of using SEM in analysis of relations between trust and public institutions' performance in Poland and Germany

The aim of analysis was to check if in the same way the public institutions' performance is evaluated by two different societies and how it is related with their satisfaction and trust towards the institutions based on data from the European Social Survey (ESS) in Poland (PL) and Germany (GE). The hypothesis of equal coefficients/means within and between countries for adequate variables representing the constructs mentioned above were checked via testing configural, metric and scalar invariance. As a starting point, we refer to the analysis in which models were built for two Rounds of ESS (in 2010 and 2012) separately as well for both Rounds jointly, for single country and between countries. MGCFA were mostly used and as result there were tested two-group factor analysis for PL and GE in 2010, two-group factor analysis for PL and GE in 2012 and four-group factor analysis for PL and GE for both years, 2010 and 2012 with the ADF method of estimation. And the similar analysis will be repeated for the most recent Round of the ESS to evaluate the stability of results obtained earlier. So far the various models without and with restrictions were evaluated at least to obtain the partial invariance, and quality of models were checked (ie. RMSEA for PL and GE for 2012 was equal to 0.043). However, as a follow-up of this earlier analysis, not only comparability with new Rounds will be presented this time but as well there will be given conclusions how it is possible based on ESS to make cross-country comparisons with SEM analysis in the analyzed topics.

Cross-national surveys: Challenges to achieve equivalent results

Chair: Eldad Davidov (University of Zürich, Switzerland)

Tomasz Drabowicz

University of Lodz, Faculty of Economics and Sociology, Dept. of General Sociology, Lodz, Poland

tomasz.drabowicz@eui.eu

An investigation into the theory of social uses of the Internet:

Structured Data Analysis of digital usage space in Germany, Norway, and Poland

This presentation tests certain aspects of a theory of social uses of the Internet (López-Sintas, Filimon, and Gracia-Álvarez, 2012), inspired by the work of Bourdieu (1984, 1986, 1989), outside the Spanish national context in which it was initially formulated. Using the OECD's PISA 2012 data for Germany, Norway, and Poland, it specifically seeks to uncover differences in patterns of Information and Communication Technologies usage among fifteen-year olds and the factors structuring those differences. Simple Correspondence Analysis has been used as the method of investigation, and SPSS version 22 and SPAD version 7.4 were used to carry out the analysis. The results show similarities between usage spaces in Germany and Norway, with the first, dominant dimension representing the frequency of digital use, and the secondary dimension representing the type of frequent digital use. In Poland, the first dimension alone - representing simultaneously the frequency and the type of digital use - is in fact responsible for over 80% of variance in patterns of digital use. Furthermore, in each country under study gender, migration background, family structure, fathers and mothers level of education, material access to the Internet at home, and the number of books at home do not explain variance in digital usage. These results suggest that Beck's (1992) theory of the individualization of risks explains digital usage inequality among adolescents in the countries under investigation better than Bourdieu's inspired social theory of Internet use. The results, however, might be also interpreted as the indirect corroboration of the latter theory, with age being one of the most significant aspects of an individual's social standing affecting digital use. Apart from discussing the substantial results, this presentation discusses the challenges encountered when conducting a comparative research using Correspondence Analysis and PISA data.

How to guarantee data quality for public use? Dealing with official statistics and survey results for pressing issues of contemporary society

Chair: Jolanta Perek-Bialas (Jagiellonian University, Poland)

Wolfgang Aschauer

University of Salzburg, Department of Sociology, Austria

wolfgang.aschauer@sbg.ac.at

Martin Weichbold

University of Salzburg, Department of Sociology, Austria

martin.weichbold@sbg.ac.at

Alexander Seymer

University of Salzburg, Department of Sociology, Austria

alexander.seymer@sbg.ac.at

Monitoring the challenge of migration for Europe based on official statistics

The necessity of conceptual clarity and data quality in applications for government authorities

The Austrian research promotion agency (FFG) has decided to fund a project dealing with the contemporary causes and effects of migration for Europe in general and for Austria in particular. The major research aim of the project is to prepare a software solution to support strategic, collaborative and cross-departmental decision-making processes between certain Austrian Ministries (Federal Chancellery, Ministry of the Interior, Ministry of Defence). In the context of the KIRAS program “critical infrastructure protection”, the project team seeks to develop and implement a prototype of a Foresight-Cockpit. This is a web-based platform, which enables government authorities to anticipate migration movements and future scenarios in a collaborative and cross-departmental manner. Finally, this tool intends to increase the quality of political crisis management in the long-term.

Our role in this interdisciplinary project is to establish a multidimensional concept of the migration challenge for Europe based on publicly available data and to provide certain quality criteria to positively influence the official use of cross-national statistics. Our concept of (statistically) monitoring migration processes encompasses several indicators referring to the extent of migrants crossing the European borders, it includes several societal indicators influencing social cohesion and the public climate on migration in European countries and it addresses certain migrant integration statistics assessing the integration potentials of migrants within European countries. Our role as critical observers of the whole implementation process of the Austrian “Foresight Cockpit” enables us to evaluate the process of (theoretically) conceptualizing, (technically) implementing and (politically) using a broad amount of statistical data from various public sources. Our presentation intends to give an insider view how the unmanageable variety of Europe-wide statistics was edited and refined for official use in this project. It should allow certain conclusions, how we – as critical methodologists – may promote a high level of conceptual clarity and data quality, guarantee an adequate technical implementation of data sources by IT-specialists and are able to affect the way how government authorities deal with statistical data on pressing European issues.

**How to guarantee data quality for public use? Dealing with official statistics
and survey results for pressing issues of contemporary society**

Chair: Jolanta Perek-Bialas (Jagiellonian University, Poland)

Kathrin Gärtner

Statistik Austria, Wien, Austria

Kathrin.Gaertner@statistik.gv.at

Manfred Zentner

Donau-Universität Krems, Austria

Manfred.Zentner@donau-uni.ac.at

**Enhancing the quality of an index measuring quality of life of young people through
participation of experts and young people themselves**

As part of the Austrian youth report (Jugendbericht) 2016 a „Better-Life-Index Youth“ was constructed. This index is based on the ten dimensions of quality of life used for the project „How is Austria?“ (initiative to measure wellbeing and progress not only with the GDP but also with 30 additional indicators from the areas of material living conditions, quality of life and environment) and is calculated on the basis of EU-SILC 2013, which included a special module on well-being. To enhance the quality of this index, the target population (young people from 16 to 30) was directly addressed in form of a survey (conducted online and by interviewers in parks, swimming pools etc., n=1.691) on the relevance of different issues for their lives. Additionally, focus groups with young people from different backgrounds were conducted to find out more about the relevance of different domains for young people in Austria. The final decisions on the inclusion or exclusion of items and variables as well as on the weighting of those items and variables within the index were made by six expert groups. In this paper we are going to compare the final index and differences for various subgroups with a very basic form of a youth index (basic index with equal weights) and simple single measures like general life satisfaction to assess the effect of the described participation process on quality.

**How to guarantee data quality for public use? Dealing with official statistics
and survey results for pressing issues of contemporary society**

Chair: Jolanta Perek-Bialas (Jagiellonian University, Poland)

Henning Silber

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

henning.silber@gesis.org

Cornelia Zuell

GESIS - Leibniz Institute for the Social Sciences, Mannheim Germany

cornelia.zuell@gesis.org

**“My vote does not matter” – Open-ended questions on non-voting behavior in the German
general elections 2013**

Voting in general elections is an essential form of citizens' participation in a healthy democracy and numerous studies have examined mechanisms explaining why people decide to vote or not to vote. However, little is known about personal reasons for individual decisions not to vote. Our paper aims to help closing this research gap by analyzing data from the German Pre- and Post-Election Study 2013 (n_1 and $n_2 \approx 2,000$), based on a representative samples of the German adult population. Non-voters were asked about their personal reasons why they will not vote (Pre-Election) or why they have not voted (Post-Election) in open-ended questions. This data offers unique insights into voting behavior interesting from both a methodological and a theoretical point of view. Methodologically, we will look for systematic differences of a socially desired behavior (voting) and a socially undesired behavior (not voting), and we will investigate the open-ended questions in various aspects of response behavior such as item non-response, length of answers, number of topics, and number of reasons. Theoretically, we will use the open-ended questions as a source to test the rational choice theory by comparing potential and empirical personal benefit and cost structures of voting. Finally, we will investigate voters and non-voters with respect to political interest, political knowledge, political involvement, and political activity.

How to guarantee data quality for public use? Dealing with official statistics and survey results for pressing issues of contemporary society

Chair: Jolanta Perek-Bialas (Jagiellonian University, Poland)

[Roula Nezi](#)

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

[Kostas Gemenis](#)

University of Twente

[Ioannis Konstantinidis](#)

University of Macedonia

Unemployment and economic voting in Greece. Evidence from a survey experiment

The state of the economy has been traditionally considered as an important determinant of voting behaviour (Lewis-Beck & Stegmaier 2007). The theory of economic voting has established a link between citizens' retrospective evaluations about the state of the national economy and their voting intentions. In Greece, much like elsewhere, the frequency of electoral events, the advent of coalition and caretaker governments, and the overreaching consensus about the bad state of the economy, have made it difficult to measure and to explain voting behaviour using the classic questions of retrospective sociotropic evaluations about the state of the national economy, as it is asked in most surveys (Nezi 2012, Nezi & Katsanidou 2014).

Motivated by the discussion above, our study which is part of a wider project examining voting behaviour in Greece during the economic crisis, uses an experimental design to examine the degree to which the economy has an impact on vote choice by focusing on the issue of unemployment. For the purposes of this study we conducted an experiment embedded in an opinion survey. In our experiment we use a split sample design. In each condition, responders were administered the same set of questions, such as their perception of the national economy over the coming and the previous 12 months. Our treatment was the level of unemployment and responders were randomly allocated to one of the three conditions; the "status quo", higher or lower levels of unemployment. Our goal was to manipulate the level of unemployment eliciting vote choice in the general election.

We use descriptive and inferential statistics to compare the findings of our experiment. Based in previous findings, we examined the effect of different levels of unemployment on the whole population but also to different population samples directly affected by the economic crisis, such as unemployed, families with unemployed members and those facing the risk of being unemployed in the future. Our analysis suggests that higher levels of unemployment do have an effect on vote choice for the incumbent.

References:

- Dassonneville, Ruth & Michael S. Lewis-Beck. 2013. "Economic Policy Voting and Incumbency: Unemployment in Western Europe." *Political Science Research and Methods* 1:53–66.
- Lewis-Beck, Michael S. & Mary Stegmaier. 2007. *Economic Models of Voting*. In *The Oxford Handbook of Political Behavior*, ed. Russell Dalton & Hans-Dieter Klingemann. Oxford: Oxford University Press pp. 518–537.
- Nezi, Roula. 2012. "Economic voting under the economic crisis : Evidence from Greece." *Electoral Studies* 31(3):498–505.
- Nezi, Roula & Alexia Katsanidou. 2014. "From valence to position: economic voting in extraordinary conditions." *Acta Politica* 49(4):413–430.

Frederick G. Conrad

University of Michigan and University of Maryland, USA

fconrad@umich.edu

Mick P. Couper

University of Michigan and University of Maryland, USA

mcouper@umich.edu

Joseph W. Sakshaug (corresponding author and presenter)

University of Manchester and German Institute for Employment Research, UK

joe.sakshaug@manchester.ac.uk

Factors Affecting the Reliability of Occupation Codes Derived from Open-Ended Reports

An understudied source of error in quantitative research is the misclassification of open-ended responses in surveys. Essential information for official statistics is derived from open responses that respondents report in their own words. These open responses are coded in order to be quantified and the coding process can introduce error to the survey data. One topic area in which open responses are essential is the measurement of occupation. We report the results of a study that sought to understand the misclassification of open occupation descriptions in the U.S. Current Population Survey (CPS). We analyzed double-coded CPS descriptions to identify which features vary with intercoder reliability. One factor strongly related to reliability was the length of the occupation description: longer descriptions were less reliably coded than shorter ones. This effect was stronger for particular occupation terms. We then carried out an experiment to examine the joint effects of description length and classification “difficulty” of particular occupation terms. For easy occupation terms longer descriptions were less reliably coded, but for difficult occupation terms longer descriptions were slightly more reliably coded than short descriptions. Finally, we observed as coders provided verbal reports on their decision making. One practice, evident in coders’ verbal reports, is their use of informal coding rules based on superficial features of the description. Such rules are likely to promote reliability, though not necessarily validity, of coding. To the extent that coders use informal rules for long descriptions involving difficult terms, this could help explain the observed relationship between description length and difficulty of coding particular terms.

Szymon Czarnik

Jagiellonian University in Krakow, Institute of Sociology, Krakow, Poland

szymon.czarnik@uj.edu.pl

Measurement of Occupational Segregation Using International Standard Classification of Occupations (ISCO)

Categorizing jobs available in modern labor market is a daunting task. One of the tools most widely used for this purpose in survey research is International Standard Classification of Occupations (ISCO), managed by International Labor Organization, and last updated in 2008. This hierarchical classification scheme is utilized in renowned research projects both at international (e.g. European Social Survey, World Value Survey), and national level (e.g. in Poland: Diagnoza Społeczna, Bilans Kapitału Ludzkiego). Data thus obtained are then used to analyze country's labor market in terms of occupational composition, mobility, interplay between supply and demand for particular kinds of jobs, etc.

Since ISCO data on occupation are elicited via open-ended questions, typically they involve subsequent coding. In the first part of the paper I discuss data imperfections due to both interviewer- and coder-originated errors. The second part focuses on the possible pitfalls of gender comparisons based on the ISCO data. The main difficulty lies in the inherent heterogeneity of occupational categories, readily revealed when higher level categories are split into sets of lower-level categories. In particular, the intensity of job segregation depends on the ISCO level at which the analysis is performed. This also poses a serious difficulty for pay comparisons between the sexes with “type of job under control” as we face inevitable analytical trade-off between the precision of job description and the sample sizes of particular occupations.

To be sure, the problems discussed aren't specific for ISCO only – other job classification schemes are vulnerable to the same problems as well.

Georg P. Mueller

University of Fribourg, Switzerland

Georg.Mueller_Unifr@bluewin.ch

Three-Valued Modal Logic for the Fuzzy Classification of Open-Ended Interview Answers

In principle, the coding of answers to an open-ended survey question can be considered as a classification task with *Boolean decision rules*. Based on the presence or absence of facts X_1, X_2, \dots, X_n in the answer of an interviewee, the person is attributed or withhold a certain categorical attribute Y , like e.g. being a liberal person or a good patriot. The advantage of such inference rules is the routinisation of the coding process, which in the best case may even be done by a computer-program.

One of the disadvantages of such coding systems is that the presence or absence of the facts X_1, X_2, \dots, X_n may be unclear, e.g. if they are not spontaneously mentioned in an open-ended interview answer. With classical binary logic, this lack of data quality (missing values) of the input is propagated to the output Y . Hence, the author proposes to use coding rules based on *three valued logic* with an additional truth state i for *indeterminate* situations. Coding rules expressed in three-valued logic obviously accept as input indeterminate truth values for X_1, X_2, \dots, X_n but return in many cases for the categorical variable Y a determinate value true or false and thus increase the data quality of Y . However, also with three-valued logic there are situations with an indeterminate outcome Y .

Thus the author proposes to use in addition *modal operators* for improving the data quality by transforming Y into a possible, necessary, or impossible attribute of the considered case. In order to demonstrate the use of this imputation method, the author analyses a fictitious dataset with open-ended answers from persons, who were asked about ingredients, which they like on their *favourite pizza*, and others, which they don't like at all. On the basis of rules extracted from a cooking-book, the proposed method is used to make inferences about the *category* of the favourite pizza of the analysed person.

The use of fuzzy classification with three-valued modal logic is of course not limited to the coding of open-ended interview questions. It could also be useful for *other classification problems* with fuzzy or partly missing input, like e.g. the identification of political regimes or the analysis of class-affiliations on the basis of personal life-styles.

Keywords: Missing values, imputation, fuzzy classification, three-valued modal logic, open-ended interview questions.

Tasos C. Christofides

University of Cyprus, Nicosia, Cyprus

tasos@ucy.ac.cy

A review of indirect questioning techniques for social scientists

In sample surveys, when the issues under investigation are sensitive, people are reluctant to participate, and even if they agree to participate, false or misleading answers are given by many of them. Indirect questioning techniques offer a solution to this problem. One such technique, the Item Count Technique, can be easily understood by participants and it can be incorporated in structured questionnaires. However, the method has a serious disadvantage related to the protection of privacy. In this talk we propose new versions of the Item Count Technique which protect the privacy of the participants.

[Andreas Quatember](#)

Johannes Kepler University Linz, Austria

andreas.quatember@jku.at

Privacy Protection vs. Efficiency in Randomized Response Questioning Designs

The reason for applying indirect questioning designs such as randomized response (RR) techniques in statistical surveys is the assumption that as a consequence respondents will cooperate and answer truthfully on sensitive questions because of the privacy protecting effect of such designs compared to the direct strategy. Hence, when comparing different techniques, this very important aspect has to be taken in account. A reliable comparison of the efficiency of different RR questioning designs only can be performed under the same level of privacy protection offered by the techniques. This has hardly been considered in the relevant literature on RR techniques.

In the theory on RR strategies for general probability sampling designs including also known true answers, the point of view of privacy protection can be incorporated in the formulae of the estimator's variance showing vividly the direct dependence of the performance of RR methods from the level of privacy protection.

Eleni Manoli

University of Cyprus Nicosia, Cyprus

An improved version of the Item Count Technique

Sensitive issues related to one's life cannot be investigated using ordinary survey methodology approaches. People do not like to participate in surveys dealing with matters of privacy. Even in cases they agree to response to questions related to sensitive or stigmatizing issues, often they provide false or misleading answers. As a result, the conclusions of such a survey cannot be taken seriously. Indirect questioning techniques are designed so that reliable estimates can be produced and at the same time the privacy of the participants is protected to a great extent. Although most of the techniques fall in the category of Randomized Response, one approach, the Item Count Technique is an alternative. The original version of the technique, due to Raghavarao and Federer (1979), Miller (1984) and Miller et al. (1986), does not fully protect the privacy of the participants and in some cases the response provided reveals that the participant belongs to the stigmatizing group. In this presentation we propose a new version of the technique which better protects the privacy of the respondents.

Techniques to increase response to sensitive questions

Session Organizers and Chairs: Andreas Quatember & Tasos C. Christofides

Mark Trappmann,
Ben Jann,
Antje Kirchner,
Ivar Krumpal
Hagen von Hermann
Universität Leipzig, Institut für Soziologie, Germany

Measuring and Explaining Undeclared Work in Germany: An Empirical Study with a Special Focus in Social Desirability Bias

This project contributes to an ongoing debate about how to measure sensitive topics in population surveys. The methodological part focuses on the effectiveness of de-jeopardizing techniques, such as Randomized Response (RRT) and Item Count (ICT), when collecting data on undeclared work. Due to the experimental design of the study, we are able to evaluate the performance of these techniques in comparison with standard direct questioning. In addition, we develop a new technique--Item Sum (IST)--for eliciting responses to sensitive questions, where the responses are continuous variables. This novel technique is a generalization of the ICT, which is only appropriate for binary sensitive characteristics. Furthermore, we developed a regression estimator that allows the analysis of data collected by means of IST.

The methodological results suggest that neither RRT nor ICT increases reports of socially undesirable behavior, whereas the IST results are more promising.

Substantively we investigated within a rational choice framework the role of financial incentives, social norms, networks and opportunity structures as predictors of undeclared work. With respect to the influence of background variables that, according to theory, foster undeclared work, the results indicate that aside from opportunity structures, social norms contribute significantly to the explanation of individual decisions to engage in undeclared work.

Ulrich Pötter
poetter@dji.de

Holger Quellenberg
Quellenberg@dji.de

German Youth Institute, Munich, Germany

Mitigating the Effects of Panel Attrition Processes: The Uses of Refreshment Samples

Refreshment samples are regularly used in panel studies to replenish a dwindling panel population. However, refreshment samples are rarely used to analyse panel attrition processes. Here we investigate their potentials to both contribute to the analysis of attrition processes and to mitigate attrition effects on estimation results.

Our approach combines ideas from the missing data literature with more traditional concepts from sample survey theory. Calibration is the process of adjusting survey sample weights in such a way that weighted marginal survey distributions are close to known marginal population distributions while design weights of the survey are only minimally perturbed. As ‘known margins’, we suggest to use the distribution of (a subset of) variables of the wave 1 sample together with the distribution of the same variables in the refreshment sample at wave 2. The resulting calibration weights can be used to mitigate the effects of attrition on all panel estimates that are based on estimating equations. Since the weights are independent from the contemplated estimators, they can be made publicly available as an add-on for panel data use files. This procedure hinges crucially on a combination of the ‘Missing at Random’ and ‘Selection on Unobservables’ assumptions. We show that it is possible to check the validity of the assumptions by gradually relaxing the requirement of conforming to the (known or estimated) marginals.

The approach is illustrated using data from the recent panel study AID:A (an acronym for ‘Growing up in Germany’) conducted by the German Youth Institute. The interviews of the first two waves of the survey took place during 2009 and 2013/14, nearly 5 years apart. Because of the unusual gap between waves, the proportion of attriters was substantial. At the same time, the refreshment sample was chosen quite large as well, facilitating the application of our methods.

Iasonas Lamprianou

University of Cyprus, Department of Social and Political Sciences, Nicosia, Cyprus

iasonas@ucy.ac.cy

Sampling through gatekeepers in survey research: biases and suggestions

Reaching informants through gatekeepers (e.g. accessing children through parents' consent) has been discussed thoroughly in qualitative research (see, for example, Wanat, 2008). Getting access to participants is important for the validity of qualitative research, but it can be equally important for quantitative, survey studies (Buchanan and Bryman, 2007). Lindsay (2005), is one of the very rare examples of researches which deal with the complex and time-consuming problem of gaining access through gatekeepers in survey research.

Motivated by the discussion above, our study uses an elaborate experimental design to measure the impact of sampling individuals through gatekeepers, on the findings of a survey.

The substantive aim of the wider project was to investigate the degree to which parents are involved in the lives of their children, during their University studies. For the purposes of this experiment, we had three samples:

1. We drew a random sample of University classes, and we asked the students to forward a questionnaire to their parents. We sampled 767 students and 432 of them (around 56%) forwarded the sealed envelope to their parents (note: it was not possible to know if they had used the right address). Out of those 432 parents who had – most likely – received the letter, 125 (29%) actually completed and returned the questionnaire. This is the “gatekeeper sample”.
2. We asked the University to sample 400 random students and post to their parents the questionnaire; 75 parents completed and returned the questionnaire (around 19%). This is the “indirect random sample”. We boosted this sample with additional 28 parent questionnaire from a near-by University using the same method.
3. We also conducted a telephone survey of the general population: out of 5184 telephone calls, 31% persons accepted to participate to the study. Out of those 1595 persons, only 429 (around 27%) had children which had completed University studies. This is the “direct random sample”.

We used descriptive and inferential statistics to compare the findings of the three samples. We also used Multi-group Structural Equation Models to study the structural invariance of the results. We found that the findings from the three samples have significant similarities but also some important differences. We will discuss the messages regarding the potential bias when sampling through gatekeepers in survey research.

Yarrow, M. R. (1963). Problems of methods in parent-child research. *Child Development*, 34, 215-226.

Lindsay, J. (2005). Getting the numbers: the unacknowledged work in recruiting for survey research. *Field Methods*, 17 (1), 119-128.

Wanat, C. L. (2008). Getting past the gatekeepers : differences between access and cooperation in public school research. *Field Methods*, 20 (2), 191-208.

Cresswell, J. W., & Miller, D. L. (2010). Determining validity in qualitative inquiry. *Theory into practice*, 39 (3), 124-130.

Buchanan, D. A., & Bryman, A. (2007). Contextualizing methods choice in organizational research. *Organizational Research Methods*, 10 (3), 483-501.

Uta Landrock

TU Kaiserslautern, Department of Sociology, Germany

landrock@sowi.uni-kl.de

Henning Best

TU Kaiserslautern, Department of Sociology, Germany

best@sowi.uni-kl.de

A sampling strategy for qualitative interviews based on quantitative methods

The project “RiKo – Risk management of corruption” studies the perception and evaluation of the risk of corruption of people in Germany. Methodologically, we use a mixed-mode design. In a first step we conduct a representative survey with 2.003 participants, in order to obtain an overall picture of attitudes and perceptions. In a second step we conduct qualitative interviews to research the subjective interpretations in more detail. Our sampling strategy for the interviews involves several methodological steps. Firstly, we calculate a hierarchical cluster analysis to identify different groups of respondents. We include the perceived occurrence of corruption in different institutions and organizations and the perceived competence of these institutions and organizations to prevent corruption as cluster variables. We identify six clusters of respondents. As additional criterion for sampling we consider the respondents’ place of residence. This criterion includes the distribution over the different geographical regions of Germany, in respect to which we expect variation in the appearance of corruption. Furthermore, we take into account the density of population. We assume that the population density makes a difference for the perception of the risk of corruption. The Federal Statistical Office offers information about the degree of urbanization differentiated in three categories as densely populated areas, intermediate density areas or thinly populated areas. We use this information to assign the degree of urbanization to the respondents of the survey. To complete the composition for the interviewees’ sample we use the affiliation to the six clusters as well as the assignment to the three types of population density, achieving 18 combinations. Lastly, we introduce variation regarding the geographical region by mapping all respondents on a map of Germany, using the program R. This map allows us to select test persons from different regions in Germany to ensure decent regional variation.

Friday, October 14th

9:00-9:45 Keynote Jörg Blasius (University of Bonn, Germany)

Another look at data quality in cross-national comparative research

The quality of survey data is a function of the three levels of actors involved in survey projects: the respondents, the interviewers, and the employees of the survey research organisations. I argue that task simplification dynamics can occur at each of these levels and the effect of such task simplifications is to reduce the quality of the data. The precise form of task simplification differs at the three levels. For respondents it might take the form of utilizing only specific parts of the available response options; for interviewers it can take the form of asking primarily the demographic questions and fabricating plausible responses for the remainder; for employees of research institutes it can take the form of near-duplication of entire questionnaires. I analyse different cross-national surveys to document various task simplification (and time saving) techniques at each of these levels. The analyses also produce results that run counter the established assumptions, for example, the lower the cognitive achievements of the respondents, the higher the reliability of their responses. This finding is, however, consistent within our task simplification framework.

Keywords: data fabrication, interviewer effects, cross-national comparisons, response styles, task simplifications.

Chair: Jolanta Perek-Bialas (Jagiellonian University, Poland)

Uta Landrock

TU Kaiserslautern, Department of Sociology, Germany

landrock@sowi.uni-kl.de

The impact of faked interviews on results of social science research

This paper examines differences between real survey data and data falsified by interviewers. The main research question is whether falsifying interviewers are able to reproduce complex theory-driven multivariate results. We use an experimental data-set with data, partly collected in real face-to-face interviews and partly by interviewers being instructed to falsify. As an example for a realistic theory-driven data-analysis, determinants of political participation are identified in multivariate regression-analyses. The models using real data and those using falsified data differ: In the latter, the explained variances are higher and some indicators have significant effects that are insignificant in models based on real data. This indicates that falsifiers try to invent plausible answers and produce more consistence than found in reality. On the other hand, some significant effects in the real data that are predicted by the underlying theory are not reproduced in the falsified data. These results support the assumption that falsifiers use stereotypes or implicit knowledge for inventing realistic answers to survey questions.

Additionally we analyzed the effects of the interviewers' payment scheme on the political participation. One half of the interviewers was paid per hour, the other half was paid per interview. In the real as well as in the falsified data there is a difference in respect to the payment scheme. In both data sets the corrected r-squared value increases when the interviewer is paid per hour. Furthermore, in the real data the payment per hour leads to more significant effects. In the case of the falsified data the regression models of interviewers, who were paid per hour, are more congruent to social reality. All these results regarding the payment scheme are in line with our theoretical assumptions. The payment per hour enhances not only the quality of real survey data; it even improves the quality of falsifications.

Marc Höglinger
ETH Zurich, Switzerland
hoeglinger@soz.gess.ethz.ch

Andreas Diekmann
ETH Zurich, Switzerland
diekmann@soz.gess.ethz.ch

Evaluating Special Techniques for Surveying Sensitive Topics: An Approach that Detects False Positives

The valid measurement of sensitive issues such as norm-violations or stigmatizing traits through self-reports in surveys is often problematic. Special sensitive question techniques such as the Randomized Response Technique (RRT, Warner 1965) and among its variants the recent Crosswise Model RRT (CM, Yu, Tian and Tang 2008) should generate more honest answers by providing full response privacy. However, the RRT sometimes does not work as expected and evaluating whether particular implementations actually improve data validity is essential before their application in substantive surveys. To this end, studies so far mostly compared prevalence estimates of sensitive traits and behaviors of different techniques. Assuming that respondents only falsely deny but never falsely admit a sensitive trait or behavior, higher prevalence estimates were interpreted as more valid estimates. However, if false positives occur, i.e., if respondents are misclassified as bearing a sensitive trait although they actually do not, conclusions drawn under this assumption are likely wrong. False positives occurred knowingly in one variant of the Crosswise Model RRT. Other RRT variants might be affected too, but studies so far have largely neglected this possibility and did not test for it. We show an evaluation approach that detects systematic false positives without the need of a validation criterion – which is often unavailable. Results from an application in a survey on “Organ donation and health” (N = 1’686) show, that the CM RRT produced false positives to a non-ignorable extent. This finding is in line with a previous validation study, but has not been revealed by several comparative evaluation studies that did not consider false positives.

Puzanova Zhanna Vasilyevna
Peoples' Friendship University of Russia, Russia
Puzanova_zhv@pfur.ru

Tertyshnikova Anastasia Gennadyevna
Peoples' Friendship University of Russia, Russia
tertyshnikova_ag@pfur.ru

Vignette method as a projective technique of studying the sensitive issues in quantitative research

The problematics of the sensitive issues in quantitative research is relevant to all countries and researchers. However, the sensitive nature of some topics depends not only on the characteristics of individuals, social groups, but also cultural, national peculiarities of different countries. The study of such problems is quite a challenge for researchers and it is solved with using projective techniques that have a number of distinctive features.

One of the sensitive topics in the modern world, in our opinion, is the attitude towards Muslims. Taking into account today's realities, use of vignettes have been proposed for the study of this problem, and it is a good solution to use projective techniques. Although use of projective techniques is not always convenient to mass surveys and rather time-consuming at the processing results stage. The advantage of the vignette method is their versatility, practicality, the possibility of formalizing the data before the field phase. The required set of vignettes designed for research purposes, can be represented in the form of individual small blocks in different parts of the questionnaire and across questionnaires. The purpose is to get the required number of representative data from all respondents.

As an example, one of the vignettes, used for the study of Muslim's image. *Malik. She has beautiful blonde hair. You know that she is very religious. Imagine that you were asked to describe her in 1-2 words, what you would choose: Asian / Muslim / Russian woman / representative Caucasian / Christian.* In this case vignette contains a number of criteria related to the name, sex, degree of religiosity (without religion), etc. From vignette to vignette signs on defined criteria change.

Our study showed that the vignettes method gives a good combination of rarely used projective techniques and possibilities of its use in the quantitative research.

Interviewer effects and measurement errors – further discussions

Chair: Jörg Blasius (University of Bonn, Germany)

Philipp Sischka

Université du Luxembourg, Research Unit INSIDE, Esch-sur-Alzette, Luxembourg
philipp.sischka@uni.lu

Alexandra Mergener

Federal Institute for Vocational Education and Training (BIBB), Section 2.2 “Qualifications, Occupational Integration and Employment”, Bonn, Germany
mergener@bibb.de

Kristina Neufang

University of Trier, Methodology and Empirical Social Research, Trier, Germany
s4krneuf@uni-trier.de

Jean Philippe Décieux

Université du Luxembourg, Research Unit INSIDE, Esch-sur-Alzette, Luxembourg
jeanphilippe.decieux@uni.lu

Forced Answering in Online Surveys: Is it really a reactance effect that reduces data quality?

Online surveys are conducted without adequate attention to implementation details too often. One example is the frequent use of the forced answering (FA) option, which forces the respondent to answer questions in order to proceed through the questionnaire. Currently, only a few studies have researched the impact of FA on different quality parameters.

Some studies that evaluated the influence of FA on quality parameters (e.g. drop-out or answer quality) hypothesized that FA leads to reactance in the participants indicated by a higher drop-out-rate as well as lower answer quality. However, no study researched the psychological mechanism behind the correlation of FA on dropout and data quality before. Psychological Reactance Theory predicts that reactance appears when an individual's freedom is threatened and cannot be directly restored. Reactance describes the motivation to restore this loss of freedom. Respondents could experience FA as a loss of freedom, as (s)he is denied the choice to leave a question unanswered. According to Reactance Theory possible reactions in this situation might be to quit survey participation or to fake answers.

This study examines the psychological mechanism that explains higher amounts of dropout and faking behavior in FA condition (compared to non-FA- condition). Our major hypothesis is that forcing respondents to answer will cause reactance, which turns into increasing dropout rates and decreasing answer quality.

We used split-ballot-field-experiments with a forced and non-forced answering instruction. Reactance was measured with a four-item reactance scale. To determine answer quality, we used self-report for faking.

Our Mediation analysis shows that respondents in FA condition report higher amounts of reactance compared to respondents in non-FA condition. In addition to that reactance also is a strong predictor for dropout behavior, faking, or re-participation. Therefore, the influence of FA on quality parameters is mediated through reactance.

Keywords: Questionnaire development, Nonresponse, Forced Answering, Online Survey Research, Reactance, Dropout, Faking, Quality control procedures

Tobias Stark

Utrecht University, European Research Centre on Migration and Ethnic Relations, Netherlands
t.h.Stark@uu.nl

Volker Stocké

University of Kassel, Department of Social Sciences, Germany
volker.stocke@uni-kassel.de

Survey Satisficing and Response Accuracy of Proxy Reports in Survey Interviews

Asking survey respondents proxy questions has become more common in recent years due to progress in ego-centric network analysis. However, validation studies show that answers to proxy questions are very often incorrect and the reasons for respondents' mistakes are hardly understood.

Based on Krosnick's theory of survey satisficing (Krosnick 1991, 1999) we develop hypotheses on the conditions of incorrect answers to proxy questions. According to this theory, respondents can derive answers in different modes of information processing, between which the quality of answers varies considerably. The response mode is determined by respondents' motivation, their (cognitive) abilities and the task difficulty. Furthermore, satisficing theory predicts interaction effects between the different determinants of the respondents' strategies of information processing.

We test these hypotheses in the unique setting of an ego-centric network study in which proxy answers of ego about the alters' characteristics have been validated by interviewing these alters. Assuming that self-reports of the peer group represent the correct answers to the proxy questions; we show that satisficing theory predicts correctly the likelihood of correct survey responses. Higher motivated respondents, those with high information availability, and those with high cognitive abilities are more likely to give correct answers while an increase in the task difficulty decreases this likelihood. Furthermore, only two interaction effects were found to be relevant: Respondents' motivation as well as their cognitive ability are both significantly more relevant in the case of difficult than in the case of easy questions.

Evi Scholz

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

evi.scholz@gesis.org

Cornelia Zuell

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

Social Desirability and Left-right Scale Placement in a Cross-cultural Perspective

This paper is about effects of social desirability on left-right scale placement in a cross-cultural context. Social desirability as a tendency to respond into direction of what is perceived to be the socially desired answer is an issue for survey research. Social desirability might not only distort self-reports on sensitive questions like violence, illegal acts, sexual behavior, or even income but social desirability might also affect answers wherever terms in question wording have a positive or negative connotation. While the intention for a survey researcher is to finally reduce a social desirability bias, the first step is to identify such a bias.

Answers on questions on political ideology in cross-national perspective might be distorted and biased by differences in social desirability of the ideological labels due to differences in political system, history and culture. Left-right self-placement on a uni-dimensional scale is one of the standard questions in many social and political surveys to measure respondents' ideological orientation in a minimalist way. We have asked about respondents' self-placement and tested respondents' individual associations with the terms left and right by asking open-ended probe questions in an experimental online survey fielded in Canada, Denmark, Germany, Hungary, Spain, and the U.S. in 2011. We have automatically coded open-ended answers using an extensive coding scheme covering the multiple dimensions of left and right. We classified the individual categories on left and right according to three potential meanings: positive, negative and neutral or ambiguous. Respondents' aggregated assessment was used to define the "socially desirable" for each country separately.

We tested whether ideological self-placement is influenced by social desirability, in particular if a 10-point scale is used offering no midpoint vs. if an 11-point scale was used offering a midpoint. Preliminary results seem to confirm the idea that respondents who cannot choose a scale center tend to answer into the country-specific social desirable direction.

Thekla Afantiti Lamprianou

Ministry of Education and Culture, Centre for Educational Research and Evaluation, Cyprus

theklaafantiti@hotmail.com

The length of Likert response scale (number of options) and the quality of measurement: a double experimental design

The optimal length of the response scale (e.g. Likert scales from 1-5 or Likert scales from 1-10) has been debated heavily in the last decades. For example, Kieruj and Moors (2010) investigated the length of the response scale and found that it could be related to specific response style behaviors. Preston and Colman (2000) investigated the optimal length of response categories and found significant differences regarding reliability. Lozano *et al* (2008), for example, suggested that longer scales could lead to more reliable measures, however, other researchers do not agree with them (e.g. Jones and Loe, 2013).

Our study used two independent samples of students ($N_1=235$ and $N_2=301$) that responded to two, somewhat similar, questionnaires. The questionnaires asked them to rate the prestige of different occupations.

The first study was an elaborate experimental design where groups of students were administered a questionnaire in a section-by-section style. In each section the same questions were administered to students, but with response Likert scales of different lengths (Section A: 0-8; Section B: 0-3; Section C: 0-1; Section D: mixed 0-3 and 0-8).

The second study, asked the same students to rate each of the occupations on various dimensions, such as occupational prestige, perceived income, education needed to exercise the occupation etc.

The study uses descriptive as well as advanced psychometric methods to investigate issues of reliability and validity of measurement. We found that longer scales indeed lead to more precise measures (smaller standard errors), thus they are more appropriate to rank order either occupations or respondents (for example, reliability reduces from 0.88 to 0.78 and 0.68 for 9 options, 4 options and 2 options). When respondents were asked to reply using shorter response scales, we came out that they had found it very difficult, and some had preferred to skip questions.

However, the second study suggests that there were non-negligible validity and reliability issues which emerged when we used the same length of response scale, but the wording of the question was slightly changed. Thus, our study suggests that researchers should consider more the validity of the questions, rather than the length of the response scale.

Kostas Gemenis
University of Twente, Netherlands
k.gemenis@utwente.nl

Detecting biased and redundant statements in voting advice applications

The online surveys, known as Voting Advice Applications (VAAs), are popular voter information tools that aim to help citizens find the party or candidate closest to their own political preferences. In a typical VAA users are asked to indicate their agreement in a battery of 30 statements using 5-point response scales. Based on the estimated positions of political parties/candidates in the same statements, VAAs are able to communicate the level of match between the user and the parties/candidates by using methods of varying sophistication. As VAAs have attracted millions of users across Europe, and have shown to affect political knowledge and voter turnout, their quality as voter information tools has come under close scrutiny. The critical literature has detected, among others, shortcomings in terms of the selection and formulation of the statements found in VAAs. In this paper I look at the issues of statement selection and formulation theoretically, as well as empirically, by focusing on the distribution of responses among both users and parties. I argue that the less polarized the responses at the user and party level are, the more biased and redundant statements will be respectively. The empirical application of this argument involves a cross-national comparison of two sets of 30 questions used by two popular VAAs during the 2009 and 2014 elections to the European Parliament. The paper concludes with implications for the selection and wording of statements for VAA designers and applied researchers alike.

Vasileios Manavopoulos
Cyprus University of Technology, Limassol, Cyprus
vmanavopoulos@gmail.com

Fernando Mendez
University of Zurich, Switzerland
fernando.mendez@zda.uzh.ch

Vasiliki Triga
Cyprus University of Technology, Limassol, Cyprus
vickytriga@gmail.com

Testing the mode effect hypothesis: A comparison of the response styles of traditional survey respondents and online respondents

Voting Advice Applications (VAAs) are tools that have become increasingly popular across Europe and can be consulted by significant portions of the electorate (up to 40 per cent it is claimed in some cases in Northern Europe) during an election contest. Furthermore, while ostensibly providing feedback to respondents about their congruence with parties across a range of policy themes, VAAs are designed to also mimic aspects of an election study. Yet, despite their growing popularity and the increased use of their data (e.g. in political science), the issue of whether Voting Advice Applications (VAA) create data equivalent to those of traditional surveys remains understudied. Beyond producing datasets non-representative of national populations, VAAs differ from other modes of data-collection in that they involve self-selected individuals, provide immediate feedback on the basis of users' responses and are answered in a self-completing capacity; all three aforementioned characteristics have been connected to socially desirable responding and fatigue-related effects on responses, such as satisficing or primacy effects. For this study data from EUvox2014, a VAA created for the 2014 European parliament elections with more than 1 million respondents, and data from the 2014 European Election Study (EES), collected through Computer Assisted Personal Interviews, are compared to examine differences among respondents from 21 EU member states. We use propensity score matching techniques to create matched for pairs of respondents based on a number of socio-demographic and political orientation variables. We then compare the two groups to assess mode effects in survey administration with a particular focus on overall responding style and party propensity to vote (PTV) scores.

S.C.J. (Bastiaan) Bruinsma
Scuola Normale Superiore di Pisa, Italy
sebastianus.bruinsma@sns.it

Assessing the Data Quality of Items in Voting Advice Applications

Scholars studying party systems rely mostly on either expert studies or manifesto data. However recently, attempts have been made to use the data generated by Voting Advice Applications (VAAs); online applications that ask voters for their position on a number (often 30) of various items using a Likert-scale, and use this to calculate the distance between them and several political parties. While the resulting data set is possibly rich and useful for political scholars, little is known about the quality of the data. This paper aims to partly address this question by looking at the assumption that the responses the voters give to the statements are ordinal. It will do so by looking at the difference between the principal component analysis and categorical principal component analysis solutions of the data (the 'dirty data index' as promoted by Blasius & Thiessen, 2012). As categorical PCA allows us to test assumed ordinality, we can use this difference to calculate the quality of the items in the VAA. The smaller the difference between both solutions, the closer the data is to being metric, indicating data of higher quality. A large difference on the other hand points at violations of ordinality, indicating low quality. The results of this study can be used by scholars to assess the quality of VAA data and estimate its usefulness for their work.

[Marcin Kocór](#)

[Jagiellonian University in Kraków, Institute of Sociology, Krakow, Poland](#)
marcin.kocor@uj.edu.pl

[Szymon Czarnik](#)

[Jagiellonian University in Kraków, Institute of Sociology, Krakow, Poland](#)
szymon.czarnik@uj.edu.pl

How to ask about skills? Different approach to estimate skills demand and supply

Since the human capital idea became widely adopted by social scientists, researchers and policy makers there are many approaches to measure skills resources and skills needs. In the presentation different ways of designing questions will be reviewed to evaluate those assessments (e.g. objectivize, declarative, projective). Next, there will be a discussion about pros and cons of each of possible measurement of these skills resources and needs. And a presentation will be illustrated by examples from one of the unique, representative data from the national survey of the Human Capital Study in Poland project (sample size is 16000 for employees and employers as well) carried out every year over last 5 years. In the summary, there will be shown ways to compare human skills' assets and employers skills' needs to determine the skills mismatches areas. In this way, the presentation will contribute to the discussion about data quality assessment in education research and as well-being crucial for analysis of labor market.

Serhii Dembitskyi

NAS of Ukraine, Institute of sociology, Kiev, Ukraine

e_forge@me.com

SCL-9-NR: a brief test for the measurement of psychological distress in survey research

SCL-90-R is a widely used test of psychological distress that was developed by Leonard Derogatis more than 40 years ago. In accordance with its size (SCL-90-R has 90 items) this test badly fits survey research requirements. On the other hand, SCL-90-R provides good basis for developing shorter versions. This fact resulted in the development of BSI-53, SCL-27, BSI-18, SCL-14 and SCL-K-9. At the same time, in my opinion, the original test has shortcomings, which also exist in its shorter versions: task's formulation in SCL-90-R leads to response biases when respondents tend to choose lower values on the 5-item scale.

For the evaluation of this hypothesis a new brief variant was constructed based of the nationwide surveys of 1997, 1999 and 2014. I've called it SCL-9-NR. At first, one item was selected from each of the 9 sub-scales of the original test according to the strength of its correlation with sub-scales indices and the General Severity Index. Then the task's formulation and response options of the new test were changed. After that two online surveys were conducted (one in Kiev and one in Lviv) in order to evaluate the validity and reliability of SCL-9-NR.

During the analysis of the collected data, the scale reliability and factor validity were evaluated (Cronbach's alpha and the CFA), the structures of two validating surveys were compared (correlation analysis, chi-square test, Student's t-test), normal distribution features of SCL-9-NR's index were verified (graphical methods, Shapiro-Wilk test). Reliability and factor validity were evaluated separately for the respondents from Kiev and Lviv. In addition, the same calculations were done for different gender and age group sub-samples. In general, the results of statistical analysis confirm my hypothesis.

Questions for further consideration, related to factors affecting SCL-9-NR results, include: 1) the sample's composition (particularly its homogeneity/heterogeneity); 2) data collection method; 3) the questionnaire's structure. Also it is important to include SCL-9-NR in the nationwide survey with the aim of convergent validity examination. If the new brief version fits the survey research requirements better than the original version, SCL-9-NR's index will demonstrate higher strength of correlation with adjacent constructs.

Alexandra Petridou

Cyprus Centre for Educational Research and Evaluation, Latsia, Cyprus

petridou.a@cyearn.pi.ac.cy

Yiasemina Karagiorgi

Cyprus Centre for Educational Research and Evaluation, Latsia, Cyprus

karagiorgi.y@cyearn.pi.ac.cy

The Diagnostic Language Tests “Milas Ellinika I”: A Validation Study

With increasing numbers of students with Greek as an Additional Language (GAL) in public primary schools over the past few years, the appropriate assessment of language competence of these students has emerged as a need within the Cyprus educational system. At the same time, appropriate diagnostic tests for the identification of these students' language proficiency levels remained scarce.

The diagnostic tests “Milas Ellinika I” (Parts A and B) addressed to pupils aged 7-11 were developed in Greece by the Aristotle University of Thessaloniki to assess three levels of the Common European Framework of Reference for Languages (A1, A2, B1). The current study aims to present and discuss findings from the a-posteriori validation study on the use of these tests in the Cypriot context. The study involved participation of 382 students from 17 schools from all the districts in Cyprus, who received remedial teaching in Greek language at the time of the study. In particular, validity evidence was collected in relation to four key elements of Weir's (2005) validation framework - namely, scoring validity, criterion-validity, context-validity and consequential validity- as well as on one additional dimension i.e. content-validity. Beyond the tests, student and teacher questionnaires were also employed to provide data on the dimensions explored within the particular framework. Data were analysed through various quantitative techniques, such as frequencies, chi-square-tests and Rasch models.

The validation results indicated that both tests could be used for the identification of GAL students' language proficiency levels in Cyprus schools. Results also highlighted areas for with minor adjustments/improvements. At the same time, the study confirmed that validity remains a multifaceted concept; various aspects need to be taken into consideration, while employing language tests in a context, other than the setting in which the testing instruments are originally developed.

Thomas Glaser

Statistik Austria, Wien, Austria

thomas.glaser@statistik.gv.at

Improvement of sampling and weighting of EU-SILC in Austria by using income register data from administrative sources

EU-SILC in Austria is a voluntary sample survey on income and living conditions that is carried out on a yearly basis via an integrated rotational design. This means that every year about one fourth of the sample is replaced by a new rotational group. EU-SILC is the source of a variety of social indicators such as the at-risk-of-poverty rate (AROP) defined as the rate of private households with an equivalised household income below 60% of the median.

Starting with EU-SILC 2012 income registers from administrative sources have been linked to the sample on micro-level via a pseudonymised key and hence replaced the questionnaire for gathering information about most components of the household income. Income register data are also available for the sampling frame and therefore can be used to enhance sample selection and also improve the weighting procedure. Since data from income registers are known for both respondents and nonrespondents this information can be used for nonresponse analysis and unit nonresponse weighting. Especially in the first wave of EU-SILC income registers have the potential to improve unit nonresponse weighting. Additionally, the calibration of weights on marginal distributions from administrative data ensures that known totals of important socio-economic characteristics can be reproduced by the sample.

Results of works dealing with the abovementioned topics carried out in terms of a Eurostat Grant agreement with Statistics Austria will be presented. The main focus will be on countering unit nonresponse bias by including income register information in the weighting procedure and sampling according to household income components from income registers matched to the sampling frame with the goal to reduce the standard error of social indicators.

Markus Bönisch

Statistics Austria, Wien, Austria; PIAAC Austria

Markus.Boenisch@statistik.gv.at

Data quality in PIAAC – International standards and national procedures

The Programme for the International Assessment of Adult Competencies (PIAAC) is an international OECD survey that compares key competencies of adults (16-65 years) in 33 countries. In order to obtain high quality data and to ensure comparability between the participating countries, the international PIAAC Consortium produced an elaborate set of standards and guidelines for almost all aspects of the national implementation. In Austria, a comprehensive set of procedures was put in place for the PIAAC fieldwork. Some of the international requirements for data collection were not reasonable within the national context and required certain adaptations to accomplish a successful fieldwork. The following fieldwork procedures will be discussed:

- Sampling (person registry VS household sample)
- Interviewer training
- Interviewer payment, -motivation and support
- Respondent motivation and incentives
- Quality control and validation
 - Data checks
 - Validation by phone
 - Validation by registry data

The presentation will talk about PIAAC and its methodological background, describe key fieldwork measures in Austria and discuss how specific measures relate to international data collection standards. Reflecting on this national experience, some of the possibilities and limitations of national compliance to international standards will be discussed.

Furthermore the multidimensional assessment of quality in PIAAC (Response Rate, Non-Response-Bias, compliance with technical standards and guidelines) will be discussed and related to national contexts.

Axel Franzen

University of Bern, Institute of Sociology, Bern, Switzerland

franzen@soz.unibe.ch

Sebastian Mader

University of Bern, Institute of Sociology, Bern, Switzerland

sebastian.mader@soz.unibe.ch

Testing Measurements of Environmental Concern:

Does a simple question outperform multi-item scales?

International surveys like the European Values Study (EVS), the World Values Survey (WVS), and the International Social Survey Programme (ISSP) have measured environmental concern since the beginning of the 1990s. However, the measures employed in these international survey programs lack comparability due to differences in item composition, item wording, and answering categories. This severely complicates international comparative research in environmental sociology. In order to overcome these shortcomings, we search for an easy single item measure of environmental concern that performs just as well as conventional multi-item scales. Such an approach has been successfully applied in the fields of happiness research and public health. In both fields a single item question such as “How satisfied are you with your life, all things considered?”, or “Compared to people of your own age, [how] would you say that your health has on the whole been [...]?” are uniformly used and facilitate comparative and cumulative research. In our study, we suggest a single item to measure environmental concern and compare it to the multi-item scale used in the ISSP. We test both instruments with respect to test-retest reliability and construct validity using a multitrait-multimethod design. Furthermore, we also investigate the predictive validity of both instruments by analysing their relation to empathy and donation behaviour.

Keywords: Environmental Concern, Test-Retest Reliability, Construct Validity, Multitrait-Multimethod Matrix

Friday, October 14th – 16:30-18:00

Session 11:

New methods to generate quantitative data

Chair: Georg P. Mueller (University of Fribourg, Switzerland)

Luis Martínez-Urbe

Fundación Juan March, Madrid, Spain

lmartinez@usal.es

Modesto Escobar

University of Salamanca, Spain

modesto@usal.es

Obtaining non-official data. Quality control for data from the Web

The Web has made possible an ever increasing access to the official statistics produced, updated and revised by national institutes and international agencies. In addition to this, the Web is the biggest data source ever compiled and provides the possibility of scraping millions of datasets to undertake unobtrusive research. These data supplement official secondary data resources and offer the researcher new insights into social phenomena whilst posing new methodological challenges.

This presentation will illustrate some approaches for extracting data from the Web and will delve into the quality challenges that researchers face when gathering and using these un-official data for their research.

Patrycja Antosz

Jagiellonian University in Krakow, Institute of Sociology, Krakow, Poland

patrycja.antosz@uj.edu.pl

Agent based modeling as a research tool in social sciences – example of modeling shirking behaviour in organisations

Simulations, treated as tools to generate quantitative data, have been gaining increasing popularity in social sciences. One simulation technique in particular, namely agent-based modeling, received plenty of attention from scholars. The technique is particularly useful for modelling complex adaptive systems that contain specific elements, i.e.: a number of heterogeneous actors behaving according to well-defined rules, changing their behaviour in time and interacting with one another. Moreover, agent-based modeling easily deals with issues that are at times very difficult to model with other quantitative techniques, such as self-organisation of a system, emergence of higher-level patterns and downward causation. All of those phenomena occur when modeling shirking behaviour in organisations – employees' tendency to exert less effort than previously agreed with their employer. On the basis of the example of shirking the presentation will discuss three topics in detail:

- (1) When to use agent-based modeling? Identification of types of problems that agent-based modeling may be helpful in solving, including a short presentation of several models of shirking in organisations.
- (2) How to use agent-based modeling? Description of using an ODD protocol (Overview, Design Concepts and Details) in designing a model. Possibilities of UML for graphic presentation of main model concepts.
- (3) How to evaluate one's model? A topic of crucial importance, so far often neglected among modellers. Using available secondary data and generating new information to check model's quality in terms of validity and prediction possibilities.

On top of that, both good and potentially improvable practices from the field will be shown. The presentation should be of interest to scholars willing to consider a different approach to traditional techniques of gathering quantitative data.

Alexandra Mergener

Federal Institute for Vocational Education and Training (BIBB), Research Associate, Section 2.2

“Qualifications, Occupational Integration and Employment”, Bonn, Germany

Mergener@bibb.de

Validity of Factorial Survey: A Comparison of vignette-judgements and single-item responses

In survey methodology, experimental designs, like vignette-studies or also called factorial surveys, have gained a broad attention in the social science. The advantages are obvious: Due to controlled settings, experimental designs allow causal interpretations and a higher internal validity. Also experiments can be easily conducted as an add-on of established surveys and the mode of data collection can be internet-based, what minimizes costs. Furthermore, because of the multidimensionality of the vignettes, factorial surveys are considered to be less prone to provoke socially desirable responses in cases of sensitive issues (like discrimination).

However, there are also critical voices. In vignette studies, respondents react on hypothetical descriptions of given scenarios (e.g. situations or persons). Therefore, it is arguable, whether this experimental design is valid in cases of external validity.

This paper compares responses to single-item questions and vignette-judgements concerning recruitment-strategies of immigrants on the German Labour Market. Therefore, I use a merged dataset of a telephone survey (first stage) and a follow-up online survey including a factorial survey (second stage) where more than 5000 companies (Human Resource Managers) in Germany are included. Topics of the first and second stage survey have concentrated on recruitment strategies with a special focus on migrants and possible skill shortages/hiring problems in the companies.

The analysis focuses both determinants that influence real employments of foreign skilled workers and those who influence hypothetical recruitment decisions. Because of the two stage process asking the same companies, it is possible to compare their responses and to gain on this way information about the external validity of the answer within.

Keywords: Factorial Survey, Vignette Study, External Validity, Data Quality

Katarzyna Lisek

Jagiellonian University, Krakow, Poland

kat.lisek@gmail.com

Be an Expert - a new tool for crowdsourced quantitative text analysis

Crowdsourcing is the process in which tasks undertaken previously by organization are outsourced to a wide group of anonymous users of new media (Howe, 2006). It is successfully used by researchers in the strict sciences to collect and analyze data. Numerous experiments proved that Internet users' performance is as good as work of professionals (Robbins et al., 2015, Hutt et al, 2013). The huge potential of passionate amateurs around the world can also be used in social studies. A great field for trying is a content analysis - a method fraught with poor reliability, absorbing high costs and resources, which is nevertheless indispensable within many scientific disciplines (May, 2013 Gibbs, 2011 Goban - Klas, 2004). The challenge is to develop such a mechanism that wipes out imperfections of this method, without losing its hermeneutical and humanistic character.

The answer for this problems could be a new application ,Be an Expert' builded and tested by our team. Developed tool which engages publicity into process of quantitative text coding opens a discussion about new approach for cultural and social studies. Data obtained form the first project conducted via our app were compered with results form traditional methods, what allowed us to confirm an utility of the new method for further research. We also examined a consequences of different types of users' motivators on value of received results. In my paper I would like to present our tool and its implementation, with a strong focus on the quality of gathered data and indicate new opportunities for content analysis to ensure high standard of data collection.

Tobias Gummer

GESIS Leibniz Institute for the Social Sciences, Mannheim, Germany

tobias.gummer@gesis.org

Comparing different selection criteria when applying case prioritization: A simulation study

In order to lower nonresponse bias in a survey, Peytchev et al. (2010) suggested prioritizing low propensity cases during fieldwork. The aim of the case prioritization approach is to balance response propensities across different strata of the sample. I argue that a lacuna in this discussion is that it remains unknown which cases to select for prioritization and how specific selection criteria affect nonresponse bias across a variety of different variables. To shed light on this pressing methodological question, I relied on the Rolling Cross-Section Campaign Survey 2013 (RCS) of the German Longitudinal Election Study. The RCS 2013 was conducted as a computer assisted telephone interview (CATI) with a response rate of 15.5%. Of the 7,882 respondents that were interviewed prior to the election, 67.9% participated in a post-election re-interview. Based on these data, I simulated the application of case prioritization for the post-election interview with varying selection criteria. I further included different scenarios in the simulation that account for different methods of treating the selected cases and how these methods perform for the different propensity strata. Nonresponse bias (prior and posterior to applying case prioritization) was assessed for 120 socio-demographic, attitudinal, and behavioral variables. The findings of the simulation study suggest, first, that the effects of applying case prioritization are estimate-specific and, thus, vary between variables. Second, there are estimate-specific thresholds (i.e., optimal selection criteria) that mark the maximum reduction of bias that can be achieved by prioritizing cases. That is, the optimal selection criterion for one variable is not necessarily the optimal selection to minimize nonresponse bias for another variable. Third, selecting cases beyond that threshold does not further lower nonresponse bias for the respective variable.

Keywords: case prioritization, responsive design, nonresponse bias, response rate.

Maja Rynko

Warsaw School of Economics, Institute of Statistics and Demography, Poland

mrynko1@sgh.waw.pl

Nonresponse bias in PIAAC and postPIAAC surveys in Poland

The methodological documentation of panel studies conducted in Poland is very limited². This is why planning a new panel survey in Poland is a challenge with big uncertainty about the response rates. In a situation where fieldwork agency has to be selected in a public procurement, this uncertainty hinders the preparation of a payment scheme which could maximally motivate the agency to reach high response rate.

This challenge was faced by a research team from Educational Research Institute³ who prepared a postPIAAC survey being a follow up of PIAAC survey. PIAAC⁴ was an international survey coordinated by the OECD with the fieldwork conducted in 2011 and 2012. PIAAC was intended to be a cross-sectional survey with a background questionnaire and a direct assessment of selected skills (literacy, numeracy, problem solving in technology rich environment). Polish postPIAAC survey was conducted in 2014 and 2015 when around 5000 out of more than 9000 PIAAC respondents were contacted. Both PIAAC and postPIAAC were designed to provide information on the relation between education, skills, selected personality traits, situation and mobility on the labour market.

While for PIAAC survey the nonresponse analyses followed the guidelines set by the international consortium and drew on the ESS standards, the postPIAAC research team had to work out own methods of survey quality assessment. As signaled at the beginning, there were no Polish experiences that could be helpful in both response rate prediction and then later in panel nonresponse analysis.

The aim of this presentation is to discuss the quality of PIAAC and postPIAAC results in terms of the fieldwork outcomes and potential nonresponse bias. As PIAAC gathered a lot of information on socio-demographic characteristics of respondents together with the assessment of their skills, several analyses were conducted to show who were the postPIAAC respondents and who were non-respondents. These analyses are probably unique in Poland and their results may be useful in planning future surveys (e.g. oversampling selected groups, adjusting incentives, better planning the resources to reach “difficult” respondents).

Patrick Lazarevic

TU Dortmund University, Germany

Patrick.Lazarevic@TU-Dortmund.de

Reporting Your Pension. Determining the Role of Age-Specific Measurement Error in Reporting Pensions by Using Administrative Data from SHARE-RV

A basic assumption for all comparative analyses is that measurements are comparable between cases. This premise is violated if groups of observations tend to have higher measurement errors than other groups. Since the functioning of a person's memory declines with age, one can be expected that, on average, older respondents have greater difficulties answering factual questions in an interview. These greater difficulties might potentially result in two outcomes: Firstly, older respondents might have greater difficulties reporting the correct answer and therefore rather give a rough estimate based on their best knowledge. In this case, respondents of greater age should display larger measurement errors than younger respondents. Secondly, older respondents, might refuse the answer altogether in order to not disclose their uncertainty or avoid misreporting. In this case, the consequence would be item-nonresponse.

To determine the role of age-specific measurement errors, administrative data of the German Pension Fund, as a reference value, were linked to 1.182 respondents' self-reports of pensions. We compared measurement errors of six age-groups and used logistic and linear regression analyses to explain the probability of sizable measurement errors and their amount as well as the probability of item-nonresponse.

Regarding descriptive statistics, an almost consistent growth of discrepancies with increasing age is evident. This applies also to the share of above-average ($\geq P50$), high ($\geq P75$), and extreme ($\geq P90$) discrepancies between self-reports and administrative data per age group. As a result of multivariate analyses, the occurrence of sizable discrepancies and item-nonresponse as well as the amount of discrepancies are, among other factors, related to higher age. These results have implications for survey methodology, as they draw attention to a source of measurement errors. They highlight the necessity to consider the role of age when measuring factual information.

Keywords: Administrative Data, Item-Nonresponse, Measurement Errors, Record Linkage

Course description

Structural Equation Modelling (SEM) is a powerful tool to analyze latent variable models that are common in social sciences, e.g. the analysis of social and political attitudes, social values, personality factors, and behavioural intentions. SEM combines factor analysis and path analysis by simultaneously estimating causal relations between latent constructs ('structural model') and relations between latent constructs and their corresponding manifest indicators ('measurement models'). Additionally, SEM allows the estimation and control of random and systematic measurement errors. Thus, SEM methodology allows an adequate modelling and empirical testing of measurement models and complex theoretical assumptions.

The course introduces theory and practice of SEM with different software packages (e.g. Mplus, AMOS, Lavaan, EQS) and discusses typical problems researchers have to deal with when adopting their own research models.

The course covers topics like:

- specification problems (e.g. formative vs. reflective indicators; higher order factors, modelling strategies)
- model evaluation (e.g. fit indices, modification indices, correction of attenuation, preconditions)
- systematic measurement errors
- measurement equivalence in longitudinal and multiple group modelling
- mean structure modelling
- causality (e.g. autoregressive cross-lagged models, mediator models, non-recursive models, equivalent models)
- latent variable interaction effects
- bootstrapping and missing value treatment
- estimation problems, different estimators and categorical structural equation models
- comparison of different SEM software packages