

ESRA 2009 conference, Warsaw  
29 June – 3 July 2009  
**Book of Abstracts**

**ES**  
**RA** European Survey Research Association



**Institute of Philosophy and  
Sociology  
of the Polish Academy of  
Sciences**



**Faculty of Law and  
Administration  
University of Warsaw**

June 25, 2009

---

**We would like to thank our institutional members for their support of the organisation:**



The Cathie Marsh Centre for Census and Survey Research (CCSR), UK  
<http://www.ccsr.ac.uk/>



Centro de Investigaciones Sociológicas, Spain  
<http://www.cis.es/>



City University London, UK  
<http://www.city.ac.uk/>



Swiss Foundation for Research in Social Sciences (FORS), Switzerland  
<http://www.unil.ch/fors>



Leibniz-Institut für Sozialwissenschaften (GESIS), Germany  
<http://www.gesis.org/>



Harris Interactive, US  
<http://www.harrisinteractive.com/>



infas Institut für angewandte Sozialwissenschaft, Germany  
<http://www.infas.de/>



ESADE Business School, Survey Research Centre, Spain  
<http://www.esade.edu/>



University of Essex, UK  
<http://www.essex.ac.uk/>



Universitat Pompeu Fabra, Spain  
<http://www.upf.edu/>

# Quick overview

<b>1</b>	<b>Substantive Applications of Survey Methodology</b>	<b>1</b>
1.1	Analysing Attitudes Towards Migration with Large Comparative Survey Data (I)	2
1.2	Analysing Attitudes Towards Migration with Large Comparative Survey Data (II)	3
1.3	Basic Human Values (I)	6
1.4	Basic Human Values (II)	9
1.5	Emotional Intelligence Research	12
1.6	European Values Study (I)	15
1.7	European Values Study (II)	17
1.8	European Values Study (III)	19
1.9	Health Surveys in Europe	22
1.10	Issues in Business Survey Methodology	25
1.11	Macro Mechanisms and Macro Hypotheses	28
1.12	Measurement of Social Capital	31
1.13	Measuring the Quality of Democracy by means of Survey Data	33
1.14	Network Survey Measurement	35
1.15	The study of left and right in political science	38
1.16	Time use surveys: emerging issues (I)	40
1.17	Time use surveys: emerging issues (II)	42
1.18	Trust and Trustworthiness (I)	43
1.19	Trust and Trustworthiness (II)	45
<b>2</b>	<b>Sampling and Nonresponse</b>	<b>47</b>
2.1	New Challenges in Sampling (I)	48
2.2	New Challenges in Sampling (II)	51
2.3	New Challenges in Sampling (III)	56
2.4	Nonresponse	59
2.5	Selection Bias in Panel Research (I)	63
2.6	Selection Bias in Panel Research (II)	65
2.7	Strategies for Nonresponse Adjustments (I)	67
2.8	Strategies for Nonresponse Adjustments (II)	69
2.9	Understanding nonresponse and attrition: Research from the UK Survey Design and Measurement Initiative	71
<b>3</b>	<b>Data Archive Teaching</b>	<b>73</b>
3.1	Access to Survey Data on the Internet (I)	74
3.2	Access to Survey Data on the Internet (II)	76
3.3	How to facilitate in order to compare - infrastructures for data cooperation	78
3.4	Quality Enhancement in Repeated Surveys	80

<b>4</b>	<b>Design and Quality of Survey Questions</b>	<b>83</b>
4.1	Can multiple question testing and evaluation methods improve survey questions or predict sources of measurement error? . . . . .	84
4.2	Implementation Issues in Internet-based Surveys . . . . .	88
4.3	Interviewer and Respondent Behaviour in Survey Interviews . . . . .	91
4.4	Methods and approaches for coding textual survey variables . . . . .	94
4.5	Methods of Evaluating the Quality of Data Collection in Surveys . . . . .	96
4.6	Questionnaire design in panel surveys . . . . .	98
4.7	Questionnaire Developing and Testing . . . . .	101
<b>5</b>	<b>Adding Data to Surveys</b>	<b>105</b>
5.1	Connecting data from independent surveys . . . . .	106
5.2	Media-reported events: the context of surveys . . . . .	108
<b>6</b>	<b>Mode of Data Collection and Data Enhancement</b>	<b>111</b>
6.1	Analysis and Evaluation of Survey Paradata . . . . .	112
6.2	Detecting and Adjusting for Nonresponse Bias using Contextual and Paradata . . . . .	115
6.3	Mobile Phones and other ICT in Survey Research: Implications for Data Quality . . . . .	117
6.4	Use of the Internet and Mixed-Mode Surveys to Survey the General Public (I) . . . . .	119
6.5	Use of the Internet and Mixed-Mode Surveys to Survey the General Public (II) . . . . .	122
6.6	WEBSURVNET: Network on challenges in survey methodology for web surveys . . . . .	123
<b>7</b>	<b>Social Indicators</b>	<b>125</b>
7.1	Comparative social surveys in Europe: issues in methodology and implementation . . . . .	126
7.2	Economic indicators and indices . . . . .	129
7.3	Integration of objective and subjective indicators: methodological and technical issues . . . . .	131
7.4	Measuring Civil Society - What are the issues? . . . . .	134
7.5	Quality of measures for concepts of Social sciences (I) . . . . .	136
7.6	Quality of measures for concepts of Social sciences (II) . . . . .	138
7.7	Social Indicators of Trust in Criminal Justice . . . . .	140
<b>8</b>	<b>Data Analysis</b>	<b>143</b>
8.1	Analysis Strategies for Cross-Cultural Research . . . . .	144
8.2	Causal Analysis Using Multi-Wave Panel Data: Problems and Solutions . . . . .	147
8.3	Comparing and Evaluating Autoregressive, Latent Trajectory, Autoregressive Latent Trajectory, and Continuous Time ALT Models (I) . . . . .	149
8.4	Comparing and Evaluating Autoregressive, Latent Trajectory, Autoregressive Latent Trajectory, and Continuous Time ALT Models (II) . . . . .	151
8.5	IRT: Item Response Theory in Survey Methodology (I) . . . . .	153
8.6	IRT: Item Response Theory in Survey Methodology (II) . . . . .	156
8.7	Marginal models for dependent data . . . . .	159
8.8	Methodological Issues in Multilevel Analysis for Cross-national Research . . . . .	161
8.9	Testing Structural Equations Models . . . . .	164
<b>9</b>	<b>Qualitative Methods</b>	<b>167</b>
9.1	Advancements in Cognitive Testing Methodology . . . . .	168
9.2	Enhancing Survey Methodology with Qualitative Methods . . . . .	171

9.3	Quantitative and qualitative approaches to validate psychological and educational questionnaires in cross-cultural research . . . . .	174
<b>10</b>	<b>Comparative Research</b>	<b>177</b>
10.1	Comparing Social Survey Data in a Global Perspective. The cases of the CSES and the ISSP. . . . .	178
10.2	Designing and testing survey instruments for comparative research . . . . .	181
10.3	Investigating social change with surveys: problems of comparability, harmonization and cumulation . . . . .	183
10.4	Language, Communication, and Research in the Comparative Context . . . . .	187
10.5	Quality Monitoring Challenges in Cross-National Data Collection (I) . . . . .	189
10.6	Quality Monitoring Challenges in Cross-National Data Collection (II) . . . . .	191
<b>11</b>	<b>Fieldwork Monitoring</b>	<b>193</b>
11.1	Fieldwork monitoring . . . . .	194
11.2	Interviewers as Agents of Data Collection (I) . . . . .	196
11.3	Interviewers as Agents of Data Collection (II) . . . . .	198
11.4	Non-response bias in cross-national surveys: an evaluation of Designs for detection and adjustment (I) . . . . .	201
11.5	Non-response bias in cross-national surveys: an evaluation of Designs for detection and adjustment (II) . . . . .	203
<b>12</b>	<b>Special Issues related to Survey Research</b>	<b>207</b>
12.1	Analysing children born of war across time, nations and disciplines (I) . . . . .	208
12.2	Analysing children born of war across time, nations and disciplines (II) . . . . .	210
12.3	Climate Change Survey: substantive and methodological issues . . . . .	212
12.4	Cognition in survey research (I) . . . . .	214
12.5	Cognition in survey research (II) . . . . .	217
12.6	Data quality in surveys among the elderly . . . . .	219
12.7	Demonstrations of new resources . . . . .	221
12.8	Special Issues . . . . .	222
12.9	Surveying sensitive subjects . . . . .	225
	<b>Index of authors and coordinators</b>	<b>228</b>

## **Chapter 1**

# **Substantive Applications of Survey Methodology**

## 1.1 Analysing Attitudes Towards Migration with Large Comparative Survey Data (I)

### Coordinator:

Marcel Coenders, University of Utrecht (Netherlands)

### 1.1.1 Anti-immigrant attitudes in a comparative perspective

Marcel Coenders, University of Utrecht (Netherlands)

### 1.1.2 A comparative analysis of socialisation, context and individual effects on European attitudes to immigrant ethnic minorities

Robert Ford, University of Manchester (United Kingdom)

**Abstract** Mass migration to Europe is a historically recent phenomenon, having commenced within the past fifty years in all the countries which have experienced it. As a result, all of Europe's diversifying societies contain cohorts who grew up in an ethnically more homogeneous, pre-migration society and cohorts who have grown up in a more diverse national (though not necessarily local) context since mass migration began. These differences in socialisation conditions could have an important impact on attitudes to minorities, analogous to the socialisation effects on attitudes found by researchers such as Mannheim (. However, cohort differences in attitudes towards minorities, and the socialisation effects which may produce them, have only been examined in individual country case studies. Ford (2008) finds evidence that cohorts socialised since mass migration began in Britain have become progressively less hostile to minorities, while Coenders et al (2008) find that in the Netherlands levels of migration and unemployment in youth have a lasting impact on attitudes to immigrant minorities.

This paper seeks to extend these national studies into a comparative context, utilising data from the 2003 European Social Survey on Western European countries which have experienced large scale migration settlement from outside Europe. The analysis examines hostility to immigrant minorities, measured using items which identify discriminatory preferences in housing, employment and marriage and opposition to laws combating ethnic and racial discrimination. The cohort distribution of these discriminatory preferences in each country is examined and then multilevel models are constructed to test the impact of both present social and political conditions and the conditions prevalent during the respondent's youth on these attitudes. The social conditions examined include the effects of contextual diversity, economic insecurity, and migration rates. The political conditions tested include the impact of incumbent governments pursuing restrictionist or assimilationist immigration policies and the effects of the presence of prominent anti-immigration political actors, such as extreme right parties or anti-immigration press campaigns. As well as these contextual factors, the impact of individual factors such as education, direct social contact with immigrant minorities, unemployment and social class are also examined.

### 1.1.3 Why Do 'Good' People Think That Immigrants are 'Bad'?

Florian Pichler, University of Surrey (United Kingdom)

**Abstract** Media and academic research on attitudes towards immigrants suggest that anti-foreigner sentiments are part of everyday life in many European countries and become more and more presentable in all sorts of contexts, including politics. The causes of xenophobia, though very complex, are often identified in political orientations (e.g. leaning towards the far-right, authoritarian minds) and personal experiences (for instance, fears about social security) in the literature. Nevertheless, we know rather little about perceptions of the more specific perceived harm done to society by immigrants. Furthermore, it is not clear whether only bad people are against foreigners, where bad usually refers to right-wing party voters, holders of anti-democratic views or those opposing equality and integration. But what about anti-immigrant sentiments in general as these views become more widespread? Here, we also set out to portray good people and their views with respect to anti-immigrant sentiments. Are they so much different? Using data from Round 3 of the European Social Survey, we offer a study differentiating between the perceived impact of immigrants on economy and culture on an overall assessment of the consequences of immigration for the host country. Applying multilevel regression techniques, the cross-national comparative analysis explores a series of objective and subjective determinants of anti-foreigner sentiment. We tentatively conclude that xenophobia is not an attitude of only bad people any longer but that the immigrants are bad belief is a widespread one across a large number of European countries.

## 1.2 Analysing Attitudes Towards Migration with Large Comparative Survey Data (II)

### Coordinator:

Marcel Coenders, University of Utrecht (Netherlands)

### 1.2.1 Pupils' Endorsement of Assimilation in Belgium's Multicultural Capital: Combining nonparametric IRT with a multilevel approach

Celine Teney, Université Libre de Bruxelles (ULB) (Belgium)

**Abstract** The dataset on which this contribution is based is composed of a sample of seventy (70) schools in Brussels Region. Three thousand one hundred and twenty one (3121) pupils attending in 2007 the last grade of the secondary education in one of the selected schools participated in the study. About half of the sample is composed of second generation migrants from about 100 different countries. The most represented origin countries in our sample are EU countries, Morocco, Turkey and the Democratic Republic of Congo. Thus, our sample is composed of a highly culturally diverse youth population.

Acculturation models were measured by three items that could be answered by a 5-points likert scale. They referred to the segregation, assimilation and integration models (Berry, 1997, 2001). From previous studies, it seems that respondents from both the majority and minority groups consider those models as unidimensional. Explorative factor analyses across the cultural groups of our sample confirmed the unidimensionality of the items.

Because of the hierarchical structure of the sample (pupils aggregated within schools), the culturally diverse population of our sample and the unidimensionality of the items, the endorsement of acculturation models by pupils will be analysed in two steps.

In the first step, a Mokken scale analysis will be performed on the three items. This will allow us to test the hierarchical property among the items across the majority and minority groups. If the three items can be ordered by degree of difficulty (or degree of agreement) and if this ranking is valid among each cultural group, the underlying Mokken scale can be used as outcome in further analyses.

In the second step, a multilevel linear model with the Mokken scale as outcome will be performed. This will allow us to control and assess the influence of schools on pupils endorsement of the adaptation strategies. At the individual level, the power of several variables in explaining ethnic differences on the outcome will be tested. However, the interpretation of the results should be done carefully, because the outcome variable and some possible explanatory variables can have completely different meanings when they refer to second generation migrants or to the majority group.

### References:

- Berry, J. W. (1997). Immigration, Acculturation, and Adaptation. *Applied Psychology*, 46(1), 5-34.  
Berry, J. W. (2001). A Psychology of Immigration. *Journal of Social Issues*, 57(3), 615-631.

### 1.2.2 Measuring experienced discrimination across minority and majority groups: Levels, dimensions and predictors

Veronique Vandezande, Katholieke Universiteit Leuven (Belgium)

Marc Swyngedouw, Katholieke Universiteit Leuven (Belgium)

Karen Phalet, Katholieke Universiteit Leuven (Belgium)

**Abstract** Discrimination has been advanced as a plausible explanation of persistent ethnic minority disadvantage in the second generation. Yet, survey studies of second-generation attainment have not usually measured experienced discrimination. The TIES surveys cover second-generation minority samples and majority comparison samples in the same ethnically diverse and relatively disadvantaged urban areas (TIES 2007-08, CeSO/CSCP). Using structural equation modelling (SEM), this study compares personal experiences of discrimination across minority and majority samples in the TIES surveys, defining discrimination as unfair or hostile treatment because of ones origin or background. The valid comparative measurement of discrimination in survey research is not without its challenges. One of the concerns is the internal validity of discrimination measures, in light of the



known effects of question wording in surveys and the multidimensional nature of real-life experiences of discrimination. Compared to existing surveys, the TIES surveys include more extensive questions that provide multiple measures of discriminatory experiences across key institutional settings and domains of social life. Moreover, the TIES surveys use standardised measures of discrimination across different groups in the same cities. Hence, the survey allows us to address key issues of equivalence in the comparative measurement of discrimination. Taking a perspective from intergroup relations, we expected that both minority and majority groups may experience discrimination in intergroup encounters. Minority and majority discrimination experiences differ, however, not only in level but also in meaning. To examine these differences, we analysed the experiences of discrimination in three groups: the Turkish and Moroccan second generations and the native inhabitants of ethnically diverse urban areas in Belgium. Using multi-group measurement models with latent means in SEM, we indeed found that levels of experienced discrimination varied considerably with group positions in the ethnic hierarchy. We tested and confirmed a multidimensional model of experienced discrimination, distinguishing between the socio-economic and the social domain, with the latter referring mainly to negative intergroup contact in public spaces. This multidimensional model differed significantly between the two minority groups on the one hand and the majority group on the other, confirming the expectation that the meanings of discrimination experiences differ between minority and majority group positions. In a last step, we used (multi-group) MIMIC models to test the external validity in minority and majority groups, relating discrimination experiences to objective and perceived disadvantage, intergroup relations, and gender as predictors.

### 1.2.3 Public Opinion Formation on Ethnic-Targeted versus Income-Targeted Employment Policy in the Netherlands

Marieke van Londen, Radboud Universiteit Nijmegen (Netherlands)

Marcel Coenders, University of Utrecht (Netherlands)

Peer Scheepers, Radboud Universiteit Nijmegen (Netherlands)

**Abstract** In this paper we examine whether ethnic-targeting increases opposition to government efforts to assist ethnic minorities on the Dutch job market. Most empirical studies on opposition to ethnic or racial policies have been carried out against the background of Americas race relations. Therefore, we derive our hypotheses from American studies on opposition to affirmative action. These studies show that Whites evaluate race-targeted policies more negatively than parallel income-targeted policies. Assessing the effects of ethnic-targeting is best done by analyzing opinion about parallel ethnic-targeted and non-ethnic targeted policy. Thus, taking advantage of a computer-assisted survey experiment, we analyze the opposition of the Dutch majority towards ethnic-targeted and income-targeted employment policy. We examine whether employment policy in the Netherlands is less opposed to when the policy reaches beyond ethnic minorities. In addition, we examine to what extent effects of self-interest, ethnic group threat and political ideology on opinion about employment policy are contingent upon target-group.

We applied data from round 2005 of the Dutch, cross-sectional representative survey Social and Cultural Developments in the Netherlands. While American studies show low levels of support for affirmative action policies, almost ninety percent of the Dutch majority members support ethnic-targeted employment policy. Moreover, we found no significant differences between opposition to income-targeted and ethnic-targeted employment policy among members of the Dutch majority. The effects of political ideology on opposition to employment policy were not contingent upon target-group. In contrast, we did find stronger (positive) effects of personal ethnic threat and ethnic group threat on opposition to ethnic policy as compared to opposition to income policy. Perceived group threat from ethnic minorities was more consequential for opposition to ethnic policy than perceived personal threat from ethnic minorities. In the discussion a comparison is made between Dutch and American welfare policies.

Keywords: survey-based experiment, interethnic attitudes, affirmative action

### 1.2.4 Racial prejudice and opposition to anti-racist policies in Europe: individual and contextual predictors

Alice Ramos, University of Lisboa - ICS (Portugal)

Cícero Pereira, Instituto de Ciências Sociais - University of Lisbon (ICS) (Portugal)

Jorge Vala, University of Lisboa - ICS (Portugal)

**Abstract** This paper explores the predictors of prejudice and opposition towards anti-racist policies in Europe, using data from European Social Survey 2002. Two different levels of predictors are considered: individual level, and contextual level. At the individual level threat perception and social values were selected as predictors, and at the contextual level we considered three dimensions: economic factor, immigration factor and political factor. The questions addressed are: 1) what are the predictors of prejudice at both individual and contextual level?; 2) what is the impact of prejudice on opposition towards anti-racist policies?; 3) is there a mediating role of threat perception on the relation between prejudice and opposition to anti-racist policies? Results show that 1) differences between countries concerning racial prejudice and opposition to anti-racist policies mainly occur at the individual level—only a small part of the differences between countries occur at an aggregate level, being share vote in extreme right the contextual variable that showed some predictive power, but only on opposition to anti-racist policies; 2) at the individual level, the adherence to the values of self-transcendence and conservation are the best predictors of prejudice; 3) Concerning opposition to anti-racist policies the best predictors are prejudice, perception of symbolic threat and adherence to self-transcendence values; 4) Although there is a direct effect of racial prejudice on opposition to anti-racist policies, this relationship is partially mediated by perceptions of threat, mainly by symbolic threat.

### 1.2.5 Opposition to immigration and to naturalization: data from ESS and ISSP

Cícero Pereira, Instituto de Ciências Sociais - University of Lisbon (ICS) (Portugal)

Jorge Vala, University of Lisboa - ICS (Portugal)

Rui Lopes, University of Lisboa - ICS (Portugal)

**Abstract** This work analyses the meditational role of threat perception on the relationship between prejudice and discrimination (opposition to immigration and opposition to naturalization of immigrants). In the first study, using probabilistic samples of 21 European countries ( $N = 36566$ ) from European Social Survey (2002), we showed that the relationship between prejudice and opposition to immigration was more strongly mediated by realistic than symbolic threat perception. In Study 2, using representative ISSP samples of two countries with different traditions of immigration (Switzerland,  $N = 940$ ; Portugal,  $N = 1514$ ), we replicated results of the first study concerning opposition to immigration and we demonstrated that symbolic threat perception mediates the link between prejudice and opposition to the naturalization of immigrants. The different meditational effects suggest that realistic threat legitimates the link between prejudice and opposition to immigration, while symbolic threat perception legitimates the link between prejudice and opposition to naturalization.

## 1.3 Basic Human Values (I)

### Coordinator:

Eldad Davidov, University of Köln (Germany)

### 1.3.1 Testing the equivalence of an instrument to assess Schwartz' human values: How harmful are translations?

Alain De Beuckelaer, Radboud Universiteit Nijmegen (Netherlands)

Eldad Davidov, University of Köln (Germany)

**Abstract** Language is (at least to some extent) culture-specific. It serves as a means to carry over personally or collectively held representations of basically all phenomena related to ones life. One major challenge in international survey research is to ensure the equivalence of translated survey instruments (e.g., reducing its cultural connotations) without compromising on their validity and reliability. In addition to intrinsic differences in language, other aspects of culture (e.g., differences in the context in which language is used and/or structural differences causing objects/entities to be non-equivalent) may also jeopardize the validity and reliability of cross-cultural comparisons (see, for instance, Smith, 2003).

In this study, we examine empirically the extent to which (high) equivalence of survey instruments to measure human values (Schwartz 1992) can be established across cultures sharing the same language as opposed to cultures having a different language. Our examination made use of a short (i.e., 21-item) survey instrument to measure Schwartz human values (as examined in Davidov et al., 2008, and Davidov 2008) based on data from the European Social Survey (rounds I and II).

### References:

- Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods*, 2(1), 33-46.
- Davidov, E., P. Schmidt, and S. H. Schwartz (2008). Bringing Values Back In: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, 72(3), 420-445.
- Schwartz, S. H. (1992). Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology* 25, 1-65.
- Smith, T. W. (2003). Developing comparable questions in cross-national surveys. In J. A. Harkness, F. J. R. Van de Vijver, and P. Ph. Mohler (Eds.), *Cross-cultural survey methods*. NJ: Wiley.

### 1.3.2 Explaining Noninvariance in the Universalism Value across 25 Countries Using Multilevel Analysis

Hermann Dülmer, University of Köln (Germany)

Eldad Davidov, University of Köln (Germany)

Elmar Schlüter, Wissenschaftszentrum Berlin für Sozialforschung (WZB) (Germany)

Peter Schmidt, University of Giessen (Germany)

**Abstract** Testing for invariance of measurements across countries and over time is necessary before meaningful comparisons of correlates and means may be conducted. In this respect one differentiates between three important levels of invariance: configural, metric and scalar. Configural invariance indicates that the same items load on the same latent variables across groups (which may be different countries or time points). Metric invariance indicates that the factor loadings of the indicators are equal. It allows a meaningful comparison of the latent variables correlates across countries. Scalar invariance requires in addition that the intercepts of each indicator are identical allowing also a mean comparison. Different authors have proposed several alternatives to deal with noninvariance: (a) to resort to partial invariance (where only two indicators are required to have equal loadings and intercepts) (Byrne Shavelson and Muthen, 1989; Steenkamp and Baumgartner, 1998); (b) to introduce method effects to account for the differences in the loadings or the intercepts (Knoppen and Saris 2008); or (c) to

refrain from cross-country comparisons. We propose a fourth strategy in which contextual effects are introduced to account for the differences in the parameters across countries.

Using data on 25 countries from the second round of the European Social Survey (ESS), we firstly test the invariance of the universalism value scale (Schwartz 1992). Two of the three indicators that were proposed by Schwartz to measure it display substantial variation in factor loadings and intercepts across countries. In a second step we introduce two theoretically-driven contextual variables to account for this variation: a human development index (HDI) and a dummy variable indicating whether the country is catholic (Norris and Inglehart 2004). As a result, measurement invariance of universalism is evidenced, accommodating its use in comparative analyses across the 25 countries.

### 1.3.3 Determining respondent scale scores along a circle: a cross-cultural comparison of priorities for human values.

Wijbrandt van Schuur, University of Groningen (Netherlands)

Sipko Huismans, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** Schwartz (1992) showed that ten clusters of values sharing the same motivational basis can be ordered in a circle. The order of the value domains along this value circle was shown to be cross-culturally valid. However, individual respondents with their priorities for the different domains could not be represented. The first author has developed a procedure for determining ordinal scale scores for both value domains (items) and respondents. This procedure is an extension of a previously developed ordinal unidimensional unfolding model (Van Schuur, 1993) that in turn is based on Mokken scale analysis (c.f., Van Schuur, 2003). It uses the response profiles directly to find a representation of both items and respondents along the circumference of a circle. This IRT- procedure differs from previous ones, in which a measure of similarity based on these profiles was used for a multidimensional scaling representation of only the items. Our new procedure allows the cross-cultural comparison of respondents scale scores in different societies. Using the ESS-data of 2003, in which the ten value domains power, achievement, hedonism, stimulation, self direction, universalism, benevolence, tradition, conformity, and security – were operationalised by 21 questions, we will compare the distribution of respondent scale scores in the European countries in which the data were collected.

### 1.3.4 Basic Human Values: Comparing Russians and Other Europeans on the Country and Individual Levels

Vladimir Magun, Russian Academy of Sciences (Russia)

Maksim Rudnev, Russian Academy of Sciences (Russia)

**Abstract** The authors rely on the Schwartz values module data from the third round of the European Social Survey held in 2006-2007 and conduct comparison of Russians with the populations from the other 24 European countries. The Russians mean rating on the integral value dimension Openness to Change Conservation is in the middle of the European spectrum (going together with more than a half other countries means) while the Russians mean rating on the Self-Transcendence Self-Enhancement dimension is the extreme one and is located at the Self-Enhancement pole of the European spectrum. The multiple regressions corroborate the results of country means comparisons and are even more sensitive to the value differences between Russia and other countries.

The common approach to cross-country comparisons through comparing country means keeps uncovered the within-country heterogeneity of values. Taking into account this heterogeneity one should compare the within countries distributions in more detailed manner. Having in mind this goal we put aside the country identity of the respondents from all the 25 countries and classify them into 4 clusters building our classification upon their responses to the 21 value items (k-means algorithm in the SPSS). In terms of two value dimensions each cluster is described by the extreme score on one dimension and medium score on the other: first cluster is characterized by the highest Openness to Change (and consequently by lowest Conservation) and the medium score on Self-TranscendenceSelf-Enhancement dimension; members of the second cluster have the highest Self-Enhancement score, members of the third have the highest Self-Transcendence score, and members of the fourth one have the highest Conservation score.

The fact that each country inhabitants have a share in each of the 4 clusters proves the within-country heterogeneity of values in all the countries studied. Still the between-clusters distributions of the respondents from the post-socialist countries are more unequal (more polarized) than the distributions of the respondents from the old capitalist countries.

The majority of the Russian respondents settled down in the second (44% of the Russian sample) and fourth (36% of the Russian sample) clusters where the post-socialist countries are overrepresented. In addition to these two majority groups there are two Russian value minorities whose members don't share the typical for Russian population values. One minority is represented by the first cluster members (14% of the Russian sample, the biggest shares in the cluster belongs to Denmark, Sweden, Switzerland and Germany) and the other is represented by the third cluster members (6% of the Russian sample, the biggest shares in the cluster belongs to France, Switzerland and Sweden).

So, relying on the country means we deal with one of the Russian value majority only while ignoring small but influential value minorities which have their counterparts (more or less numerous) in the other European countries. By the same token even those European countries which have the most different from the Russians mean values appear to have smaller counterparts of the Russian value majorities among their populations.

### 1.3.5 Cultural Basic Human Values – Western Europe vs. Eastern Europe: Accordance or Discrepancy

Indrek Tart, Tallinn University (Estonia)

**Abstract** Basic values are important to catch the motivational space of different cultural communities. Europe under its enlargement and integration processes has become overloaded by immigrant populations who carry with them the human value preferences that took shape in quite unlike living environments than those of their current host country. Thus, in general European countries acquire more and more ethnically flavoured layers into their national basic value structures. If values can be treated as quite stable after the process of socialization then we are facing a need to find mechanisms of how they will function and be politically-socially-culturally correctly dealt in changing human environments. But first of all there must be clear understanding which basic value constellations Europe already has. Do European ethno-linguistic basic value fields look more unified or diverse? Do Eastern and Western Europeans share only partly basic human values constellations and which values show the watershed? Does ESS deal with the possible problem? In the paper we will use the Shalom Schwartz Portrait Value Questionnaire 21-item version from the three rounds (2002, 2004 and 2006) of the European Social Survey presenting 39 ethno-linguistic communities of more than 120 thousand respondents. This paper will describe the value constellations of cultural similarity networks/clusters in the multidimensional scaling space, looking at higher dimensional solutions for more sensitive analysis, using several versions and various dimensionalities of MDS. Analytic possibilities proposed by the UCINET software will also be in use.

Keywords: basic human values, cultural communities, multidimensional scaling

## 1.4 Basic Human Values (II)

### Coordinator:

Eldad Davidov, University of Köln (Germany)

### 1.4.1 Value orientations of Muslims in Europe

Eldad Davidov, University of Köln (Germany)

Peter Schmidt, University of Giessen (Germany)

Shalom Schwartz, Hebrew University of Jerusalem (Israel)

**Abstract** In this study we compare basic value orientations as defined by Schwartz (1992) between Muslims and non-Muslims in Europe. For the data analysis we apply three waves of the European Social Survey (ESS 2002-2003, 2004-2005, and 2006-2007). Seven values are compared across the two groups based on findings in previous studies, suggesting that only 7 out of the 10 postulated values can be identified with data from the ESS (Davidov, Schmidt and Schwartz 2008; Davidov, 2008). First, we conduct a multigroup confirmatory factor analysis (MG-CFA) to assess whether these seven values are comparable across Muslims and non-Muslims. High invariance is supported by the data, allowing a meaningful comparison of value means between them. In the second step, mean comparisons between the two groups are conducted controlling for other sociodemographic characteristics, such as age, gender, level of education and household income.

### References:

- Davidov, E. (2008). A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey. *Survey Research Methods* 2(1): 33-46.
- Davidov, E., Schmidt, P. and Schwartz, S. (2008). Bringing Values Back In: Testing the Adequacy of the European Social Survey to Measure Values in 20 countries. *Public Opinion Quarterly* 72: 420-445.
- Schwartz, S.H. (1992). Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology* 25: 1-65.

### 1.4.2 Human Values and Active Citizenship: a Multilevel Approach

Massimiliano Mascherini, European Commission - Joint Research Centre (Italy)

Anna Rita Manca, European Commission - Joint Research Centre (Italy)

**Abstract** The current European climate and the revitalized Lisbon strategy have put social cohesion at the heart of the European policy agenda. An important dimension of social cohesion is the active citizenship with its determinant of values, representative democracy, civil society. Mascherini and Hoskins, (2009), measured Active Citizenship through a composite indicator based on a list of 61 basic indicators drawn from the European Social Survey of 2002.

Following this framework, in this paper we estimate the effects of human values dimensions taken from the theory of basic human values, Schwartz (1992), Davidov et al. (2008), on active citizenship together with a set of socio-demographic variables, (as age, gender, education, etc.), behavioral variables, (as the media impact, the politics interest, etc.) and contextual variables (GDP pro capita, Gini index, etc) by using a random-effect multi-level regression model.

Although the statistical legitimacy on comparing these effects across countries need to be discussed, preliminary results show the importance of the impact of human values on active citizenship. In particular the self-transcendence dimension displays a positive effect on active citizenship as well as the stimulation and self-direction dimensions. On other hands conservation present a negative effect. This result is robust also by controlling the several individual and contextual variables included into the model.

### References:

Hoskins B.L., Mascherini. M. (2009) Measuring Active Citizenship through the Development of a Composite Indicator. *Social Indicators Research* 90:3, 459-488

Davidov E., Schmidt P., Schwartz, S.H. (2008) Bridging Values Back In: the adequacy of the ESS to measure values in 20 countries. *Public Opinion Quarterly* 72:3, 420-445.

Schwartz S.H. (1992), Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In Zanna P. (eds). *Advanced in the experimental social psychology*. 25. San Diego Academic Press, 1-65

### 1.4.3 Religion and values from a comparative perspective. The relation between religious involvement and value priorities in 30 European countries.

Bart Meuleman, Tilburg University (Netherlands)

Eldad Davidov, University of Köln (Germany)

Bart Cambré, Tilburg University (Netherlands)

**Abstract** Since the inception of the social sciences, scholars have been convinced of the theoretical and empirical usefulness of the concept values. Values can be seen as desirable transsituational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity (Schwartz 1994: 21). Despite its paramount importance, the value concept has not always received the scientific attention it might deserve. It is remarkable that certain aspects of value theory have been tested only very rarely.

In this contribution, we focus on the relation between religious involvement on the one hand, and value priorities on the other. There are several good theoretical reasons to expect that there indeed exists a link between religion and value systems. Through processes of socialization, religions influence the value systems of their adherents. An alternative explanation can be sought in self-selection processes: value priorities might also guide the choice to involve in a particular religion. Yet, we expect the relation between religious involvement and specific values to vary according to the cultural context. After all, the structural position a religion holds in society can have an effect on the specific values it is associated with. The religion-values connection is a highly relevant topic because it can provide deeper insight in what it exactly means to be a religious person, and how this meaning can differ along with the cultural setting.

This paper provides an empirical test of the connection between religious involvement and value patterns. This is done from a comparative perspective: using data from the first three rounds of the European Social Survey allows us to compare this relation across religious groups (including Catholic, Protestant and Orthodox Christians, Muslims and Jews) in more than 30 European countries. This wide variety of cultural settings makes it possible to determine how various contextual factors shape the relation between religion and value priorities.

### 1.4.4 Values and Social Classes: Differences and Similitude in Europe

João Ferreira de Almeida, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Rui Brites, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Analia Torres, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

**Abstract** Following Schwartz, who advanced the basis for structural theory of values, values express motivational goals and differentiate, precisely, by those expressed goals (Schwartz, 1996). Values analysis assumes a central stance in social research, conceptualizing them as organized and relatively durable systems of preferences. Values are expressed in a definite culture, understanding the latter in the anthropological sense; that is, as an enlarged and diverse aggregate of characteristics that at the limit is synonym of the notion of society.

Bearing in mind the actionalist properties of values understood as their behavioural repercussion the research program that the authors have being developing in this domain identifies patterns of values connected to social classes. However, other values settings may equally show relevant distinctions, either in an isolated way or combined with social classes, as it is the case of gender.

Analysing European Social Survey, round 3, the aim of this presentation is to illustrate how social class and gender structures contribute to shape patterns of life.

### 1.4.5 Human values and subjective well-being. A study in an Italian context

Caterina Sciarra, Università degli Studi di Firenze (Italy)

Maurizio Cappelli, Università degli Studi di Firenze (Italy)

**Abstract** In the ambit of a quality-of-life project, firstly supposed to be used in order to help local authorities to improve their policy standards, two different surveys have been carried out in two middle small urban contexts in Italy (Fucecchio and Atri) by using the same questionnaire design. Moreover, collected data allowed researchers to investigate thoroughly the relationships between subjective well-being and human values, observed in terms of importance that each individual attach to life ambits.

In this perspective, one of the questionnaires areas aimed at studying human values by applying the well-known Thurstone scaling approach by asking interviewees to rank eight ambits (friendship, physical aspect, career, culture, family, earnings, social relationship and health).

The results of the analytical approach allowed us to point towards three main goals:

- testing the methodological soundness of Thurstone scaling approach in studying human values
- studying the relationship between human values and subjective well-being in cross-cultural comparison. This will be done in the perspective of managing Thurstone data as importance weights in building complex well-being indicators (F. Maggino, 2007)
- analysing differences and similarities between the two cultural contexts: data allowed us to understand how interaction between subjective well-being and human values may vary by shifting from a context to another, though these contexts can be considered similar at a macro level.



## 1.5 Emotional Intelligence Research

### Coordinator:

Joan Manuel Batista-Foguet, ESADE, Universitat Ramon Llull (Spain)

### 1.5.1 Development of a Questionnaire to measure Emotional and Social Intelligence Competencies

Aláide Sipahi Dantas, ESADE, Universitat Ramon Llull (Spain)

Joan Manuel Batista-Foguet, ESADE, Universitat Ramon Llull (Spain)

Ricard Serlavos, ESADE, Universitat Ramon Llull (Spain)

Richard Boyatzis, Case Western Reserve University (United States)

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

Laura Guillén, INSEAD (France)

**Abstract** Practitioners and researchers have pointed out the influence of emotions on decision-making processes, on willingness to take risks, on the definition of strategies, and on the role of managers in ethical issues inside organisations, amongst others. The influence of emotions on everyday organisational life has called for the assessment and management of emotional and social intelligence competencies. The substantial increase in research on Emotional Intelligence (EI) in the field of management is another indication of the growing recognition of the importance of the individuals emotional facet to better understand professional outcomes (Fernandez-Berrocal & Extremera, 2006). Furthermore, higher performance in work settings is related to the ability to develop cognitive as well as emotional and social intelligence competencies and to applying this capacities in organisations (Boyatzis, Stubbs & Taylor, 2002).

The assessment of the emotional facet of individuals in organisations, through emotional and social intelligence competency mapping, follows McClellands (1973) proposition that competencies are better indicators of professional performance. Besides, according to Spencer (1997), the competency movement has been responsible for a better assessment of the competencies that are present in the best performers. However, while cognitive competencies have historically received great attention, only in the last two decades has there been a growing attempt to grasp the emotional dimension of workers.

The aim of this article is to present the developmental process of a measurement instrument to assess workers emotional and social intelligence competencies. In drafting this questionnaire, the foundational literature and the most recent literature on Social and Emotional competencies measurement were reviewed along with recognised questionnaires on Emotional Intelligence. A pilot study was conducted in order to make a preliminary assessment of the questionnaires construct validity.

### 1.5.2 A Motivational Model of Emotional Intelligence Competency Display

Aláide Sipahi Dantas, ESADE, Universitat Ramon Llull (Spain)

Joan Manuel Batista-Foguet, ESADE, Universitat Ramon Llull (Spain)

Ricard Serlavos, ESADE, Universitat Ramon Llull (Spain)

**Abstract** This paper proposes a motivational view over the display of different emotional intelligence competencies. The efficacy of organizations is rooted on the emotional intelligence competencies of workers. As this idea becomes widely accepted organizations have invested on the development of individual competencies best suited for their businesses. However, in general people present different levels of facility on developing competencies. A possible explanation for these differences lays on the fact that people have different implicit motivations. Such motivations, main drivers of the behaviours, are more stable and can facilitate the development of particular competencies. For this reason, the identification of the implicit motivations that underlie the different emotional intelligence competencies is important. It can help to explain why people present different competencies and better direct the investments to its development. The paper here proposed bases itself on this premise to suggest a model of emotional intelligence competencies display based on the three social motives (affiliation, power and achievement).

The first section of the paper will explore the competency movement. This construct, despite its controversial status among academics, have definitely entered the practitioner world and impacted the way organizations are managed. The discussion around the competency movement will settle the terrain for the central subject of the

paper that is the correlation between emotional intelligence competencies and implicit motivations. In a sub-section, it will be discussed the emotional intelligence competency approach to competencies. The recognition that skills other than the technical ones are fundamental for the personal and organizational effectiveness widened the research agenda towards emotional and social competencies. Since emotional intelligence competency, at this paper, will be treated as consistent emotional and social behaviour that contributes to effective performance, the second section of this paper will concentrate on the role of motivations on driving peoples behaviour. On the following sub-section, the discussion will be centered on the implicit motivations that drive behaviours, specifically the 3 social motives theory. Finally, theory linking the three social motives to the display of effective professional profiles will be used to derive a model of emotional intelligence competency display based on the three social motives. An agenda for future research and a final section of conclusion will end the paper.

### 1.5.3 Different ways to approach a management development course: The ESADE Business School experience.

Ricard Serlavos, ESADE, Universitat Ramon Llull (Spain)

Joan Manuel Batista-Foguet, ESADE, Universitat Ramon Llull (Spain)

Alaíde Sipahi Dantas, ESADE, Universitat Ramon Llull (Spain)

Richard Boyatzis, Case Western Reserve University (United States)

**Abstract** Individual differences in the way people approach a management development program are still a blackbox although they have a significant influence in the learning process and the outcomes. The use of a web platform not only facilitates a customised approach to the learning process but a good opportunity to explore within the blackbox. This study presents the experience of Esade Business School in running a leadership development course offered in a whole range of MBA and Executive Education Programs.

The course is supported by a web designed to easy participants self management of a process that invites them to reflect on their visions and current capabilities and to elaborate a learning plan. The process is inspired in the WSOM experience and includes a wide range of exercises created by Richard Boyatzis.

The study is based on the data collected through the web in the period 2006-2008, which gathered information on more than 2000 individuals from 50 different MBA and Exec. Edu Programs. This data include vision statements, declared values and operant philosophies, competencies profiles measured with the Emotional and Social Competencies Inventory (ESCI), learning styles, and learning goals.

A main conclusion is that although there are many ways to navigate through the process there is a great coincidence in the consideration of self awareness raise as the most important outcome. And yet self awareness is the most difficult competency to measure. Other findings are related to the comparisons between individuals and groups from a wide range of backgrounds, nationalities, and industries, and the difficulties when conducting an outcome research project.

Some practical implications can be derived to improve the design and evaluation of management development courses.

### 1.5.4 Split Ballot Split-ballot multitrait-multimethod designs'Effect of Response Scale on Assessment of Emotional Intelligence Competencies

Joan Manuel Batista-Foguet, ESADE, Universitat Ramon Llull (Spain)

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

Richard Boyatzis, Case Western Reserve University (United States)

Ricard Serlavos, ESADE, Universitat Ramon Llull (Spain)

**Abstract** Most of the research on emotional and social competences has been done in English-speaking countries. Most of the survey measurement instrument of behavioural competencies use a five-point Likert scale, without considering that the response modality may affect the findings. Batista-Foguet, Saris, Boyatzis, Guillen & Serlavos (2009) have shown using a Split-Ballot design that an eleven-point response scale with mainly European MBA students, that the eleven-point scale lead to composites with greater reliability, validity and less invalidity, than the commonly used 5-point scale. A complete four-group Split.Ballot (SB-MTMM) design for two methods and their replications (exact repetitions of the same items are included in the design after sufficient time has elapsed to avoid memory effects) is used to compare both response modalities. On this paper, the SB-MTMM design allows to evaluate the occasion effects as well as measurement bias, reliability, and validity simultaneously, while reducing response burden a usual drawback of MTMM studies.

### 1.5.5 Facing the adversity: The role of personal and social competences on well-being

Celeste Simões, Technical University of Lisbon (CMDT-LA/FST) (Portugal)

Margarida Gaspar de Matos, Technical University of Lisbon (CMDT-LA/FST) (Portugal)

Mafalda Ferreira, Technical University of Lisbon (CMDT-LA/FST) (Portugal)

Gina Tomé, Technical University of Lisbon (CMDT-LA/FST) (Portugal)

**Abstract** Aim: The aim of the present work is the analysis of personal and social competences associated to well being, in a population of adolescents with special needs facing adverse environmental situations.

Methodology: The sample included 494 adolescents with special needs, mean age 14 years old. Pupils attended 77 public schools from all Portugal. Data collection was held within the HBSC (Health Behaviour in School aged Children) survey (Currie et al, 2004; Matos et al, 2006). For the purpose of this specific study, the questionnaire included questions about well-being (Mental Health Index Kidscreen , Gaspar et al., 2006), personal and social competencies (Resilience Module - California Healthy Kids Program Office, 2000), and life events (Life Events Checklist, Johnson, 1986).

Results and Discussion: Results showed that adolescents with special needs had a set of internal resources (personal and social competencies cooperation and communication, empathy, problem solving, self-awareness, self-efficacy, and goals and aspirations) that functions as protective factors regarding their well being, since the adolescents that have higher scores on these dimensions present higher levels of well-being comparatively with the ones that have medium or low scores on these dimensions. However, when adolescents face several negative life events, only problem solving have a significant protective effect. Adolescents that have high levels of problem solving maintain their levels of well-being even the presence of multiple negative life events.

Problem solving seem thus to be a resilience factor and a determinant factor for the positive development of adolescents with special needs (ASN), when they face multiple negative environmental events.

Keywords: Personal and social competences; Well-being; Negative life events; Protective factors; Resilience Factors.

## 1.6 European Values Study (I)

### Coordinators:

Paul M. de Graaf, Tilburg University (Netherlands)

Wolfgang Jagodzinski, Central Archive (Germany)

Ruud Luijkx, Tilburg University (Netherlands)

### 1.6.1 Religious trends in Dutch society 1981-2008. Results from the European Values Study.

Loek Halman, Tilburg University (Netherlands)

**Abstract** The paper reports on the empirical findings of the four waves (1981, 1990, 1999, 2008) of the European Values Study in the Netherlands. We will focus on the trajectories of change with regard to religion and explore whether secularization continues or whether the trends in Dutch society have to be understood in terms of neo-secularization. The neo-secularization thesis states that there is a religious decline at the societal level, but not at the individual level.

In the paper we will describe the main trends in religious beliefs and practices in the Netherlands and the causes and consequences of these trends. Using a cohort perspective the paper explores whether the religious changes can be attributed to population turnover or to intra cohort developments. The secularization thesis asserts that religion has lost its societal significance and thus that religion is no longer a source of inspiration for civic behaviour and therefore considered a threat for social cohesion and democracy. This idea will be elaborated on by studying the impact of religious beliefs and practices on key notions of civil society, such as solidarity, social trust, societal engagement and tolerance.

### 1.6.2 Russians' attitude towards social and political issues

Elena Bashkirova, Bashkirova and Partners, CEO (Russia)

**Abstract** This survey is devoted to Russians attitudes towards various social and political issues. We compared WVS1995, EVS1999, and EVS2008 data. The analysis revealed that opinions on some issues have changed significantly e.g. attitudes towards role of the government in economic and social life of the country: comparison showed that in 2008 the share of those who believed that people should take more responsibility for providing themselves went up. There are also changes in political behaviour: citizens in 2008 were less active than in previous years.

Level of confidence in some social institutions has also changed e.g. the shares of those who trust churches a great deal increased by 6.3% since 1995; share of the respondents who trust Russian parliament doubled in 2008. The share of those who trust the police also increased. Besides, people became less wary of major companies: in 2008 21.7% of the respondents answered that they did not trust such companies at all, while in 1999 this figure equalled 38.6% and in 1995 - 34.2%. Russians also tend not to trust the United Nations: the level of trust dropped more than twice in a period from 1995 to 1999, for the next decade (1999-2008) the situation has improved, but level of trust was almost by 13% lower as compared to 1995.

Out of all the options - democratic political system, having a strong leader, army rule or experts making decisions - people prefer democratic political system.

Russians also would like to see less emphasis on money and material possessions in public morality.

### 1.6.3 The relationship between autonomy and social behavior. The Netherlands 1981-2008.

Paul M. de Graaf, Tilburg University (Netherlands)

**Abstract** Much longitudinal research on value change has shown that the strong economic development in western society has given way to a persistent shift towards more appreciation of personal autonomy. The decreasing power of religious and other traditional authorities have gone together with an increasing focus on self-expression and personal autonomy. The question I set out to answer in this paper is to what extent the increasing focus on personal autonomy has gone together with decreasing levels of social behaviour and solidarity. Using the four Netherlands waves of the European Values Study (1981, 1990, 1999, 2008), I will use a cohort perspective to study the contributions of population turnover and intra-cohort developments to the explanation of macro-level changes in personal autonomy and in social behaviour. In the second part of my analysis, I will study (developments in) the association between personal autonomy and social behaviour on the individual level. Social behaviour will be measured by (a) solidarity, (b) support of civic values, (c) unpaid work for voluntary organizations, and (d) political participation. The results will be presented during the conference.

#### 1.6.4 When values change

Rikke Skovgaard Christensen, University of Copenhagen (Denmark)

Peter Gundelach, University of Copenhagen (Denmark)

**Abstract** This paper will explore situations where there is a shift of values due to social events. Data is the Danish combined 1981-1990-1999-2008 European Values Survey data set for Denmark and the special Danish EVS panel study 1990-2008. The paper will first trace the value consequences of major social events in the general population. Examples of such events are the value consequences of fear of AIDS in the late 1980 and the political value consequences of the insistence of a so-called value struggle initiated by the Danish conservative government from 2001. Second, the panel study is used to trace consequences of individual life story event such as marriage or divorce on family and well-being variables.

## 1.7 European Values Study (II)

### Coordinators:

Paul M. de Graaf, Tilburg University (Netherlands)

Wolfgang Jagodzinski, Central Archive (Germany)

Ruud Luijkx, Tilburg University (Netherlands)

### 1.7.1 Is Religion Really a Cultural Marker? The Impact of Religion on National, Supra- and Sub-national Identities

Wolfgang Jagodzinski, Central Archive (Germany)

**Abstract** Since the seminal book of Huntington on the clash of civilization it is widely accepted that religion is an essential element in the formation of collective identities. Indeed, religion seems to play a key role in a larger number of recent inter- and intra-national conflicts. Religion may not only be a source of conflict, however, it can also foster societal integration. In Europe quite a few politicians attempt to create a European identity by appealing to the common Christian heritage. It is an open question whether these attempts are successful. If religion promotes societal integration at all, it may rather strengthen the national, or, below the national level, the regional identification than the integration with an overarching supra-national culture. It is also unclear whether religion plays the same role in religious and secular societies.

In order to answer some of these questions, the measurable aspects of collective identities have to be elaborated. Furthermore, the possible channels through which religion might influence these identities have to be specified. This is done in the first part of the paper which ends with a number of middle-range hypotheses about the influence of religion on regional identification in West European societies. The hypotheses are tested with data from the German part of the European Values Survey which also allows the comparison of the highly secularized, former Protestant East German regions with Protestant and Catholic regions in the West. The data also allow a test of the assumption that religion becomes increasingly important as a cultural marker.

### 1.7.2 Trends in gender beliefs in Romania: 1993-2008

Malina Voicu, The Research Institute for Quality of Life, Romanian Academy (Romania)

Paula Andreea Tufis, The Research Institute for Quality of Life, Romanian Academy (Romania)

**Abstract** Attitudes towards the appropriate role for men and women have changed in all European countries during last decades. Everywhere in Europe people share more equalitarian values now, as compared to fifty years ago. There are two main theoretical frameworks which try to explain changes in gender attitudes, one stressing the effect of cultural change (Inglehart, Norris, Welzel, 2002; Inglehart & Norris, 2003) and the second emphasizing the role of economic and technological development (Wilensky, 2002) on support for gender equality. The cultural approach considers the shift towards more gender equality as part of a broad cultural change, while the economic approach allots the transformation to the increased demand for female labor force in industrial and post-industrial societies. According to the first approach, changes in gender beliefs occur mainly as result of cohort replacement, younger cohorts socialized in a more equalitarian culture replacing the older ones. The second theory emphasizes the role of economic and institutional context on gender attitudes variation.

Romania, as other communist countries, has registered higher level of womens participation on the labour market and an increased support for gender equality on the labour market in the beginning of the 90s. However, the gender ideology was rather a traditional one, because communist regimes have introduced progressive ideology of gender equality in very traditional cultures (Hanson, Wells-Dang, 2006). While the higher support for womens participation on the labour market was the result of full time employment policy, practiced by the communist regime, gender ideology reflects the level of social and economic development since the state did not intervene in issues related to the private area. Transition to market economy produced an economic decline and increased the unemployment rate. In the same time, the generous public childcare system existing during the communist time disappeared. These processes are likely to determine a trend towards less equalitarian gender beliefs.

This paper focuses on the dynamic of gender beliefs in Romania, during the last 15 years, analysing the variation in attitudes towards women involvement on the labour market and in gender ideology. We expect to find a significant return to more traditional gender ideology, as well as a decrease in support for women employment outside home, which are mainly the effect of contextual factors (higher unemployment rate and lack of coherent family policy). The analysis will employ standard cohort decomposition of social change methods,

using European Values Survey data from three different waves (1993, 1999, 2008) which allow longitudinal comparisons.

### 1.7.3 Does the decline in church religiosity foster alternative spiritualities and/or unbelief?

Pascal Siegers, University of Köln (Germany)

**Abstract** Religious change has profoundly modified the religious opportunity structure for individuals living in Western European countries. In this regard, the decline of church-religiosity in Europe brought to evidence by proponents of secularization theory is most noteworthy. According to secularization theory, in the long run, modernization is expected to bring about atheism and religious indifference.

But is church religiosity really declining in Europe? Are alternative religious beliefs stepping into the breach left by church-religiosity?

In a first step, my paper aims to map trends of church religiosity for six Western European Countries (Belgium, Denmark, France, Italy, Germany, and the Netherlands). The four waves of the European Values Study (1981, 1990, 1999, and 2008) provide data over four decades that allow mapping trends in church-religiosity and unbelief or atheism.

Proponents of religious individualization, however, challenge the view that declining church religiosity necessarily results in unbelief. In contrast, they point to the growing importance of alternative spiritualities in Europe.

Alternative spiritualities are characterized by (1) an emphasis on personal experiences of transcendence, (2) strong religious individualism, and (3) great distance to traditional (Christian) churches and other religious organizations. With respect to the conception of the sacred, holistic and impersonal beliefs are dominant.

Evidence supporting the growth of alternative spiritualities, however, is mostly based on qualitative studies. In order to quantify the importance of alternative spiritualities, some items were included in the fourth wave of the European Values Studies (2008) that are designed to measure core aspects of alternative spiritualities: (1) an item on religious individualism and (2) a scale of spiritual self-assessment.

Thus, in a second step, I will combine these new indicators with established measures of church-religiosity in order to draw a comprehensive picture church religiosity, alternative spiritualities and unbelief in the selected countries. I will use Multiple-Group Latent-Class-Analysis to test whether a single measurement model fits for all the selected countries or whether substantial heterogeneity with respect to the indicators can be observed for religious orientations across countries. This is a necessary precondition to establish conceptual equivalence in order to compare belief patterns across countries.

My hypothesis is that alternative spiritualities are particularly strong in those countries where the decline in church religiosity (1981-2008) has been particularly strong. If this turns out to be the case, this challenges the assumptions of secularization theory.

### 1.7.4 Translation Problems in International Surveys: On the Difficulties of Translating a Western Language Master Questionnaire into Asian Languages

Kazufumi Manabe, Aoyama Gakuin University (Japan)

**Abstract** I have been involved in comparative and collaborative data analyses of the ISSP Religious Module, the EVS and the WVS. The results have shown that while religiosity is strongly associated with peoples beliefs, attitudes and values in Western countries, that tendency is not evident in Japan. The results, I have hypothesized, are based on Western religious concepts and views, and the Westernised translation problems of such question items as religious faith, denomination and participation. To inquire into these translation problems in international surveys, I have done comparative surveys on religiosity and social values in Germany and Japan from an independent methodological standpoint. In this presentation, I will discuss the difficulties of translating a Western Language Master Questionnaire into Asian Languages.

## 1.8 European Values Study (III)

### Coordinators:

Paul M. de Graaf, Tilburg University (Netherlands)

Wolfgang Jagodzinski, Central Archive (Germany)

Ruud Luijkx, Tilburg University (Netherlands)

### 1.8.1 Presentation of the European Values Study 2008: an overview

Ruud Luijkx, Tilburg University (Netherlands)

**Abstract** The European Values Study (EVS) is a unique research project into Europe's basic values. It now spans a period of almost 30 years with surveys in 1981 (13 countries), in 1990 (31 countries), in 1999 (33 countries), and in 2008 (45 countries). The survey has gradually expanded from mostly Western European countries in 1981 to the whole of Europe in 2008. The 2008 European Values Study is conducted in all European countries with a population over 100,000: Albania, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Great-Britain, Greece, Hungary, Iceland, Ireland, Northern Ireland, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Republic of Macedonia, Malta, Moldova, Montenegro, the Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine. It is expected that the fieldwork will be in almost all countries will be finished in the summer of 2009.

The 2008-questionnaire is to a large extent a replication of the earlier questionnaires. Even though several items have been changed in the consecutive waves, EVS still includes an impressive number of unchanged questions. The questionnaires pertain to a very broad spectrum of life domains: family and marriage, economics, work, leisure, politics, religion, morality. This allows to introduce domain-specific and overarching concepts and to examine the relationships between basic values and attitudes in different fields. Given the long time span, the European Values Study is a unique database for longitudinal values research.

The 2008-wave was improved in many ways. Substantively, replication was guiding, but all questions were evaluated and adapted where necessary; the set of background variables (also of social origin and partners) was elaborated. The translation process was controlled by using a web-based translation programme (WEBTRANS). The sample size was enlarged to at least 1500 cases and a strict random sampling was used in all countries.

### 1.8.2 Value change in Eastern Europe - what is happening there?

Horatiu Rusu, Lucian Blaga University of Sibiu (Romania)

Mircea Comsa, Babes-Bolyai University of Cluj Napoca (Romania)

**Abstract** Large-scale studies that developed tremendously the last three decades or more, offered the opportunity to more and more longitudinal and cross-national analyses. European Values Study Series (among other like WVS, ISSP, ESS, and EES) is one of them. Consequently, many comparative studies using EVS data were published. A lot of them, mainly dedicated to Western European countries, search for value similarities / dissimilarities and value change (Arts, Hagenaars & Halman, 2003; Arts & Halman, 2004; Ester, Braun & Vinken, 2006; etc). Although starting with the beginning of 90's EVS extended to East, there are not as many comparative studies dedicated to the Eastern European countries. In our paper, we employ both a longitudinal and cross national analysis, dedicated to value change in Eastern Europe. "What are the trends in value orientations over the time passed from the fall of the communist regimes in the Eastern European countries and how do they vary between countries?", "Is there a dynamic towards cross-national value convergence between the Eastern European countries?", "Has Eastern Europe, the tendency to witness a general farewell to traditional values, like Western societies?" are the main research questions in our study. We move beyond a "particularist view" of values, that is focused on a specific value domain (i.e. religious values, family values, work values, political values, etc.), and employ a "dimensionalist perspective", focusing on a value system. We will use EVS / WVS 1990-2006 data sets from Belarus, Bulgaria, Czech Republic, Estonia, Germany East, Hungary, Latvia, Lithuania, Moldova, Poland, Russian Federation, Serbia, Slovakia, Slovenia, Ukraine, Romania. A special attention is given to Romania, where 2008 data will be also included in the analyses. The main method employed is multi-group confirmatory factor analysis (MGCFA); we are testing for measurement invariance (configural, metric and scalar) and then compare the means of the latent variables (values). On one hand we expect no clear dynamic towards either value convergence or divergence. On the other hand, we expect to find - due to the reign of transition



that brought with it a growth in the social risks - a revival of the traditional values, after a short period (at the beginning of the 90's) of modernization and post modernization (see Inglehart & Welzel, 2005).

### 1.8.3 Job Preference Orientations in Europe: Tests for Equivalence and Explanations for Cross-National Variation

Maurice Gesthuizen, Radboud Universiteit Nijmegen (Netherlands)

Ellen Verbakel, Tilburg University (Netherlands)

**Abstract** As opposed to extrinsically motivated persons, who value the material rewards of a job, intrinsically motivated persons are motivated by a concern to make use of ones abilities. It has become increasingly important for modern economies to have an intrinsically motivated work force. The shift to a service economy and the upgrading of the occupational structure, for example, have made it harder to monitor the employees work, because nowadays higher autonomy and more specific knowledge characterize their jobs. As a result, employers are more dependent on their employees concern to do the job well. In addition, intrinsically motivated employees are believed to be more satisfied with their jobs, and could therefore be expected to be more productive, and more happy in general. In this sense, intrinsic job preferences are beneficial for the employee, the employer, the economy, and society as a whole.

Little is known about differences in job preferences between countries, despite the potential beneficial impact job preferences may have on the societal level. Recently, based on 5 countries and single-level single response models, Gallie published a piece of comparative research on this topic in the *European Sociological Review* (2007). We plan to follow his new route, but improve on it in several ways. Firstly, we propose an improved measurement of work values that has been extensively validated for cross-country comparability, using scalar invariance models that test for cross-national equivalence. Furthermore, in stead of using single response models, our multiple response models take into account that intrinsic and extrinsic job preferences are usually highly positively correlated, which will result in more reliable estimates. Finally, we will considerably extend the number of countries under study (from 5 in the Gallie-study to 23), and perform multilevel analyses to determine to what extent variation in job preferences can be attributed to differences between individuals or differences between countries, so that stronger tests can be performed of hypotheses that explain country differences.

On the country level, we hypothesize that a high level of wealth, a high level of social welfare, and a low unemployment rate positively affect intrinsic job preferences and negatively affect extrinsic job preferences. These country characteristics reduce insecurity, which allows people to relax their focus on extrinsic job features and strengthen their focus on intrinsic ones. Moreover, we hypothesize that characteristics of a high-quality production regime (e.g. high levels of R&D and good work conditions) positively affect intrinsic job preferences and negatively affect extrinsic job preferences.

We analyse the integrated dataset of the European Values Study 1999/2000. Preliminary results show that countries differ substantially in the extent that their populations hold intrinsic and extrinsic job preferences. Our hypotheses generally find support as far as extrinsic job preferences are concerned, but are generally less able to explain differences in intrinsic job preferences.

### 1.8.4 The European Social Model: formal institutions and individual attitudes, 1990- 2006. A multilevel study

Monika E. Kaminska, University of Amsterdam (Netherlands)

Ferry Koster, University of Leiden (Netherlands)

**Abstract** In this paper, data from the EVS, the ESS, Eurostat, ICTWSS and UNCTAD, are combined to examine, through multi-level modeling, the cross-national and over-time variation in attitudes towards values underpinning the European Social Model (ESM).

While not a unitary concept, the ESM can be indirectly defined through its core characteristics: welfare provisions, non-discrimination in workplace, social and economic rights, and social partners participation in economic governance. So far, the related research has been geared towards examining these formal institutions of social models in EU member states. This study extends this perspective and provides insight into how much public support the model receives, by focusing on EU public opinion attitudes towards values underpinning the ESM: sharing risks and opportunities, social solidarity and cohesion, protection through active social intervention, consultation in industry and social and economic citizenship rights (as defined by Giddens 2007). First, individual attitudes from the period 1990- 2006 are aggregated to the national level to investigate trends over time and cross-national variation. Second, national level indicators are used to examine national social models with

reference to the core institutional characteristics of ESM. Next, the study establishes the degree of commonness (convergence, clustering or hybridization) between national publics attitudes towards ESM values, as well as the degree of commonness between formal institutions of national social models. Finally, the paper examines the patterns of relations between attitudes and formal institutional setting at the national level, as well as the level of economic and political integration into the EU. While the general discussion covers old and new member states, particular attention is paid to the 2004 entrants.

### 1.8.5 Trends and expectations: First results from the Austrian survey

Regina Polak, University of Vienna (Austria)

Katharina Renner, University of Vienna (Austria)

**Abstract** Trends and explanations.

For the Austrian part of the study we already have very interesting outcomes. In all parts of the study work, family, religious and political values the year 1999 differs significantly from the other waves of the survey Austria took part in (1990 and 2008). At the workshop we will discuss possible explanations for some of the outcomes.

Regarding the perception of values in general: The implicit meaning of value in the EVS is worth a discussion since it is not used in a consistent way. Values are measured in the context of socio-demographic structures attitudes, norms, etc. However, it remains unclear what perception of value is meant. Does it follow a more philosophic-ethical or a sociological concept? A profound discussion on that matter would be crucial to the EVS itself.

Index of Authoritarianism.

In the Austrian questionnaire, out of a series of variables and with the method of factorial analysis we established an index of authoritarianism. In past surveys, the variables that were relevant for authoritarian attitudes remained constant. In this years survey, however, the nature of authoritarianism has changed. Authoritarianism today is different than ten years ago. This has many consequences. For instance, changes in political values can be explained by this shift of authoritarianism. At the workshop we will present possible explanations for this phenomenon.

Religion.

Two aspects will be discussed that mainly focus on methodological issues and might deserve attention in future surveys:

- 1 Religious transformation processes are happening all over Europe. However, the questionnaire does not provide variables to measure these processes. What is measured is the level of religiousness and (on the opposite) the number of people attending or leaving religious communities. Yet the questionnaire does not measure the pluralism of church-related religiousness, nor does it justice to religious pluralism itself.
- 2 We can see changes over the last twenty years regarding the nature of religiousness. An analytic model (on the basis of a theoretic model) shows that next to the cognitive dimension of religiousness, the religious praxis and experience is a relevant factor. Nevertheless the variables in the questionnaire concentrate on the cognitive dimension. A supplement of experience-related questions would be preferable.

## 1.9 Health Surveys in Europe

### Coordinator:

Mathis Schröder, University of Mannheim (Germany)

### 1.9.1 Sampling, recruitment and participation in recent health examination surveys in Europe

Hanna Tolonen, National Institute for Health and Welfare (THL) (Finland)

Päivikki Koponen, National Institute for Health and Welfare (THL) (Finland)

Liv Grotvedt, Norwegian Public Health Institute (Norway)

Sidsel Graff-Iversen, Norwegian Public Health Institute (Norway)

Johan Heldal, Statistics Norway (Norway)

Kari Kuulasmaa, National Institute for Health and Welfare (THL) (Finland)

**Abstract** The FEHES (Feasibility of a European Health Examination Survey) Project made a review of recent health examination surveys (HES) in Europe and prepared recommendations for future surveys.

A national HES has been carried out during 2000-2006 in 11 EU Member States (MS). Surveys focused mainly on adult population aged 25-64 years. The target population varied from all persons living in the country to the citizens of the country, depending on the available sampling frames. In some countries, people who could not communicate in the national language of the country were excluded. Most countries also excluded institutionalized people.

The sampling frames used for drawing individuals or households varied from population registers to census lists, address files, telephone directories, and electoral registers. About half of the surveys were individual based. The other half were household surveys, where usually 1 or 2 persons per household were examined.

The proportion of selected people found not-eligible varied from 1% to over 30%. Also the proportion of persons who could not be contacted varied considerably (1%-25%) between the surveys. The participation rates of the surveys varied from 20% to 93%.

Mass media was used in almost all surveys to inform public about the survey. Many surveys also had Internet sites with more information about the survey and a telephone line for further questions. The first contact to the selected persons was by mail, either through information or invitation letter in all surveys, except in one, where the first contact was a telephone call.

Varying definition of the target population and low response rates in some of the national HESs make comparative analysis difficult. The FEHES recommendations for future national HESs in Europe have set guidelines concerning the target population, eligibility criteria, sampling frame, and recruitment efforts as well as the definition and target level of participation rates. These aim to enhance the comparability of future national HESs in Europe.

### 1.9.2 Striving for the best of both worlds: integrating EHIS questions in the Belgian Health Survey

Stefaan Demarest, Scientific Institute of Public Health (Belgium)

**Abstract** National health interview surveys organized throughout the European countries provide a treasury of information on the health status, the lifestyle characteristics and the medical consumption of their populations. Unfortunately, up till now the international comparability in this domain was rather restricted due to differences in the wording of the questions, in the mode of data collection and in the sample design.

Since 2002, DG SANCO and EUROSTAT have taken different initiatives to implement a European Health Survey System (EHSS) including the development and implementation of the European Health Interview Survey (EHIS). The EHIS aims at measuring the health status, life style (health determinants) and health care services use of the EU citizens in a harmonized and highly comparable way. It is important to stress that the EHIS is not conceived as a new, separate survey but as a standardized tool intended to be integrated in national surveys. It goes without saying that this integration can jeopardize the comparability of the results with the previous national surveys, due to changes in the wording of the questions, the reference periods,... As a result, one of the major goals of (national) surveys - monitoring trends in health, lifestyle and medical consumption - is under

siege. On the other side, one of the main advantages of the EHIS is to allow making comparison between the EU member states and publishing results at the EU level.

Although the integration of EHIS will become compulsory in 2013, it was decided already to implement (parts of) EHIS into the Belgian 2008 survey. The paper describes the problems related to this implementation and how they were dealt with. Based on a pretest executed in 2007 in which both the current Belgian questions and the similar EHIS questions were used evidence based decisions could be taken for the development of the 2008 survey in Belgium. Yet, provisional results of this survey show that still decisive steps must still be taken for a full integration of the EHIS.

### 1.9.3 The Survey of Health, Ageing and Retirement in Europe (SHARE)

Mathis Schröder, University of Mannheim (Germany)

**Abstract** The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-nationally comparable panel database of micro data on health, socio-economic status and social and family networks of more than 45,000 individuals aged 50 or over in 14 European countries and Israel. The data are available free of charge to the entire research community. Two waves of data have already been collected in 2004-05 and 2006-07, respectively. Data collected include health variables (e.g. self-reported health, health conditions, physical and cognitive functioning, health behaviour, use of health care facilities), bio-markers (e.g. grip strength, walking speed, body-mass index, peak flow), psychological variables (e.g. psychological health, well-being, life satisfaction), economic variables (current work activity, job characteristics, opportunities to work past retirement age, sources and composition of current income, wealth and consumption, housing, education), and social support variables (e.g. assistance within families, transfers of income and assets, social networks, volunteer activities). In addition, the SHARE data base features anchoring vignettes to adjust cross-national subjective data for differential item function. The survey's third wave of data collection, SHARELIFE, will collect detailed retrospective life-histories of all panel respondents. Future waves of SHARE will include more biomarkers collected from dried blood spots and other measurement, such blood pressure. Methodologically, SHARE uses ex-ante harmonization of survey content as well as a common survey management system to ensure comparability across all participating countries.

This presentation aims at introducing the many innovations of SHARE in the realm of cross-nationally comparable health surveys to a wider audience and encouraging the use of the data in their own research. Some examples demonstrating the research potential will be shown.

### 1.9.4 The German Health Interview and Examination Survey for Children and Adolescents

Martin Schlaud, Robert Koch Institute (Germany)

**Abstract** The aim of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS) was to improve the information available on the health of the upcoming generation in Germany and to fill gaps in knowledge. This was to be done using a nationwide representative health interview and examination survey of children and adolescents aged 0 to 17 years. From May 2003 to May 2006, 17641 girls and boys took part in the study; along with their parents' input, they provided a unique pool of information.

Enrolment of study subjects was carried out in two steps: first, 167 study locations (sample points) were chosen; second, subjects were selected from the official registers of residents of the local residents' registration offices. The percentage of quality-neutral drop-outs was comparatively low (5.3%). The participation was 66.6% of all eligibles and showed only little variation between age groups and sexes. The completeness of data sets was good. The main reasons for non-participation were failure to appear at an agreed appointment time (or cancellation of the appointment at short notice), refusal of the child/adolescent himself, or lack of interest of the parents.

Questionnaire data were obtained on physical health and diseases, mental health and problems, social health and living conditions, lifestyle factors and health behaviour, nutrition, health risks, and health care utilisation. Tests and measurements were taken on anthropometry, perseverance, vision ability, blood pressure, heart rate, motor activity and co-ordination and a sonography of the thyroid gland was done. Blood and urine samples were analysed for micronutrient deficiency (e. g. folic acid, vitamin B12, vitamin D, transferritin), sero-epidemiology of infectious diseases and immunisation status (e. g. hepatitis markers: anti-HAV, anti-HBs, anti-HBc, HbsAG, antibodies against measles, mumps, German measles, herpes, *Helicobacter pylori*), risk indicators or risk factors for chronic non-communicable diseases (e. g. specific IgE, antibodies against respiratory agents, HDL cholesterol, total protein, TSH, fT3, fT4). Retain samples were frozen for future analyses.

An overview of methods and selected topics of health among children and adolescents in Germany will be presented.

## 1.10 Issues in Business Survey Methodology

### Coordinator:

Ger Snijkers, Statistics Netherlands (CBS) (Netherlands)

### 1.10.1 Recent Research in Business Survey Methodology

Ger Snijkers, Statistics Netherlands (CBS) (Netherlands)  
Jacqui Jones, Office for National Statistics (United Kingdom)  
Gustav Haraldsen, Statistics Norway (Norway)

**Abstract** In 2006 the first International Workshop on Business Data Collection Methodology (BDCM workshop: ONS, London), in 2008 the second took place hosted by Statistics Canada (Ottawa). The goal of the Workshop is to exchange information and learn from each other. Issues that can be discussed are research papers, but also and this is even more important research in progress, unanswered questions, hands-on experience, mistakes and pitfalls. Thus, the Workshop is aiming for the development of current best practices and increase of expertise.

This paper will summarize the key issues and findings of the 2008 BDCM workshop. The issues deal with:

- Business data collection designs and strategies, including the combination of register data and survey data in one design for data collection.
- Contact strategies and Data collection procedures. This includes research on response burden, understanding the response process within businesses, strategies on motivating businesses to participate, and non-response reduction strategies.
- Electronic data reporting, including the design of electronic questionnaires.
- Standards and guidelines for questionnaires that are specific to business questionnaires, and pre-testing of business questionnaires.
- Mixed-mode survey designs, e.g. the implementation of mixed-mode designs, including the generation process of questionnaires for a number of modes, and the contact strategy.
- Data quality and costs, which includes e.g. research on timeliness of response, mode-effects, weighting, reduction of bias, and efficiency improvement.
- Post survey evaluation.

Also the developments in business survey methodology for the last 15-20 years will be discussed. When looking at the conference agendas of the three International Conferences on Establishment Survey (ICES 1993, 2000, 2007), we can see a shift in the methodology of business surveys. In the first conference a lot of attention was given to the design of individual surveys, frames and sampling, and data analysis. In the last conference the focus was on methodological research in general with regard to data collection, the use of administrative data, and business frames and sampling.

### 1.10.2 The Challenges and Opportunities of carrying out Research and Analysis in Business Tax

Katherine Fox, HM Revenue & Customs (United Kingdom)  
Jane Taylor, HM Revenue & Customs (United Kingdom)

**Abstract** HM Revenue & Customs (HMRC) was formed in 2005, following the merger of Inland Revenue and HM Customs and Excise departments. HMRC collects and administers tax revenues, pays and administers benefits and credits and enforces border protection.

The Knowledge, Analysis and Intelligence (KAI) directorate is a multi-disciplinary unit within HMRC that exists to provide research and analytical services to help deliver better, evidence-based policy and improvements to the service HMRC offers its customers. KAI's research is used to develop policy, support the departments core function, inform public debate and monitor changes over time.

This paper will set out the key challenges of carrying out research and analysis for business taxes within HMRC. Results from a methodological review of issues for large business tax research will be presented alongside issues of relevance to tax research with small and medium enterprises. Solutions to the challenges of business tax

research in the areas of population definition, sampling (including both internal administrative data and external data sources), data collection and maximising response rates will be proposed. Lastly, the paper will outline the aims of a new strategic approach to carrying out tax research with businesses, and the added value of developing a more long-term, coherent and intelligent evidence base that will provide lasting insights and transferable findings that can be used across the Business Taxes directorate.

### 1.10.3 How useful are administrative data for nonresponse adjustment in a business survey - evidence from the German job vacancy survey

Hans Kiesl, Institute for Employment Research (Germany)

**Abstract** In 2008, a regulation of the European Parliament and of the Council introduced new Community-wide quarterly statistics on job vacancies. Most member countries are going to produce the required numbers on job vacancies by conducting a business survey.

In Germany, these statistics are provided by the Institute for Employment Research (IAB), which has conducted a national job vacancy survey for several years. Some information about job vacancies in Germany is also provided by the Federal Employment Agency, which collects reported job openings from business units. These "registered vacancies" cover less than 50% of all vacancies, however, since many businesses prefer not to cooperate with the Agency. Therefore, the IAB conducts a yearly job vacancy mail survey to estimate the total number of vacancies (accompanied by a quarterly CATI follow-up survey).

During the last decade, response rates have declined dramatically. Nonresponse rates in our survey are as high as 80%. Therefore, calibration to variables from external sources is important. Additionally, administrative microdata might be used for modeling (and adjusting for) nonresponse. In our paper, we compare different weighting estimators for the number of job vacancies in Germany that are calibrated to totals from administrative business data. Furthermore, we present our findings concerning the correlation between variables that are included in an administrative business register and the occurrence of nonresponse in the latest wave of our survey, and discuss how to exploit these correlations for nonresponse adjustment.

### 1.10.4 Evaluation of the redesign of the Structural Business Survey questionnaires

Deirdre Giesen, Statistics Netherlands (CBS) (Netherlands)

Mattijn Morren, Statistics Netherlands (CBS) (Netherlands)

Ger Sniijkers, Statistics Netherlands (CBS) (Netherlands)

**Abstract** The Structural Business Survey (SBS) measures a large number of indicators of the activity and performance of Dutch businesses. Each year SBS questionnaires are sent out to about 70,000 businesses.

After a thorough evaluation, the SBS questionnaires have undergone a complete redesign. This included a reduction of the number of variables measured, a new layout, new (shorter) explanatory texts, a different order of the questions and the introduction of an electronic version of the questionnaire for all respondents. The new design was implemented in 2007.

This paper describes the evaluation of the new design with respect to the perceived response burden and the quality of the data. For the evaluation both qualitative and quantitative data collection methods are used, including field visits to firms, focus groups with staff members of Statistics Netherlands and response analyses.

First results indicate that the new questionnaires are well received by the respondents. In a telephone survey SBS respondents reported a lower response burden for the new questionnaire and in open interviews during field visits they generally appreciated the content and design of the new form. The overall response rate of the new questionnaires is comparable to previous years and about 80% of the response was received electronically. Moreover, the response time has decreased by almost three weeks.

### 1.10.5 Marginal employment: linking individuals with industries by firms

Holger Alda, Federal Institute for Vocational Education and Training (BIBB) (Germany)

Daniela Rohrbach, Federal Institute for Vocational Education and Training (BIBB) (Germany)

**Abstract** Marginal employment in Germany is characterized by low earnings (maximum 400 ) and only marginal social security benefits. Between 2004 and 2008 the proportion of marginally employed workers rose from about seven to 15 percent of the total workforce, meaning that there are around seven million marginal workers in Germany today.

From an individual-level perspective marginal employment might be seen as a chance for lower skilled people to (re-)enter the labor market. However, since the quality of jobs with marginal employment is rather worse than in regular employment, and it also does not offer social security benefits - people usually should try to avoid marginal employment. From a firm or industry perspective hiring marginal employed workers is theoretically not clear as well. On the one hand marginally employed workers increase firms' (or industries') numerical flexibility. On the other hand it does not seem to be an appropriate tool for keeping or rising human capital.

Connecting individual-level data with firm and industry information from other sources, the paper investigates in workers and higher economic units (firms, industries) as determinants for marginal employment simultaneously, using multilevel analysis. Individual-level data are from the BIBB/BAuA-Survey of the Working Population on Qualifications and Working Conditions in Germany 2006 covering 20,000 individuals in the active workforce. To study industry-related effects on workers' chance to be marginally employed we rely on the IAB-Establishment-Panel (including about 16,000 German firms). Unlike administrative data, it includes a much greater set of relevant variables and allows for considering within-industry heterogeneity.

The analysis proceeds in two steps: First we run firm-level analyses trying to identify relevant predictors for marginal employment. Based on an aggregation of these factors at an industry-level (NACE), logistic multilevel analyses then allow for explaining variance in the risk to be marginally employed by individual as well as industry factors, simultaneously. First calculations demonstrate that in fact a considerable proportion of 25 percent of total variance comes from the industry level.



## 1.11 Macro Mechanisms and Macro Hypotheses

### Coordinator:

Heiner Meulemann, University of Köln (Germany)

### 1.11.1 Finding and Examining Hypotheses on Country Differences

Heiner Meulemann, University of Köln (Germany)

**Abstract** Cross-national research in how far country context determine attitudes and actions of person, independently form person level determinant. The problem is to find the mechanisms how country level characteristics determine mean individual attitudes or action of the population. To discuss this question, I will first explore what a country context may mean. Second, discuss how hypotheses of its effects on actions or attitudes of persons can be conceived of. And, third, show how such hypotheses must be examined.

### 1.11.2 What affects the demand for income redistribution? Macro level explanations for cross country variations.

Ursula Dallinger, University of Trier (Germany)

**Abstract** In research on public support for income redistribution by welfare states there is a vivid debate about the country level influences explaining the empirical cross national differences in the degree to which citizens vote for income redistribution.

According to the 'median voter-hypothesis' from the economists Meltzer/Richard the demand for income redistribution should rise with the degree of inequality in a country, because more voters expect advantages from redistribution and opt for governments with redistributive programmes. This direct link between the social facts, a certain level of inequality, and the distributive preferences suggested by the median voter hypothesis, was rejected by research. Because preferences for redistribution obviously are no demand reacting on the actual Gini, other explanations are sought for, especially those that introduce cultural factors like regime approaches. Again, the empirical support for that approaches is insufficient. So, research is still testing hypothesis about the influence of country conditions that make people prefer more or less equality and redistribution coming from different branches of research, like the effect of Gross Domestic Product or unemployment. Further institutionalist approaches assume an influence of different kind of welfare institutions (the universalistic vs. the selective welfare state), the varieties of capitalism approach argues similar. The overarching question is whether preferences follow a rational calculus of advantages or are driven by political cultures.

The paper will present this literature and explore the underlying mechanisms, by which country level effects are thought to influence behaviour or vice versa. Also aggregated individual preferences can have a collective effect. Then different hypothesis on the relation between country characteristics and demand for redistribution are analysed by multi level regression procedures based on the ESS and the ISSP data sets with additional country data.

### 1.11.3 Citizen's Critical Evaluation of the European Union: A Multilevel Model of Support for European Integration

Pascal Siegers, University of Köln (Germany)

**Abstract** Support for European integration in line with Eastons concept of political support is generally considered to depend on two classes of determinants: utilitarian evaluations (costs and benefits of European integration) and expressive evaluations (affective attachment to the European Union and the nation state) of the European Union.

The study of support for European integration differs fundamentally from studies of nation states. The member-states of the European Union possess their own historically grown political regimes and differ considerably with respect to their economic development. Consequently, they occupy different positions within the systems of solidarity and competition between European countries. Some countries, for example, are net-contributors to the budget while others are net-receivers. For this reason, research on support for European integration cannot pass over the impact that the national contexts have for evaluations of the European Union. Do the differences in national contexts explain the differences in support for European integration across countries?

Particularly, utilitarian evaluations of the European Union are expected to be affected by economic differences: (1) wealthier countries bear most of the direct costs of integration. Citizens might blame the EU for these transfers towards less wealthy countries and this could reduce the support for European integration (level-effect). (2) The costs of integration are also expected to decrease the effect that positive evaluations of EUs performance have on support for European integration (slope-effect) because evaluation takes place in an overall more skeptical context.

Expressive evaluations of the European Union, in contrast, are expected not to be systematically related to the economic context of evaluation.

The results from the hierarchical regression model reveal that the context of evaluation does indeed matter for support for European integration. Higher levels of social expenditure decrease support for European integration in the realm of welfare and education policies but contrary to the expectations contributor countries show higher level of support for European regulation policies. Even more: the effect of positive evaluations of EUs performance on support for political integration is stronger in net-contributor countries than in net-receiver countries. Hence, there are no traces for strong national egoisms.

#### 1.11.4 Preferences and Uses of Television and Newspapers in Cross-National Perspective

Heiner Meulemann, University of Köln (Germany)

**Abstract** It is examined whether preferences of persons and public control of the media determine informational rather than entertainment media use in European countries. Informational rather than entertainment media use is conceived of as two choices: newspapers rather than television, and political rather than other contents. Five dependent variables are constructed: (1) time devoted to television as percentage of the time devoted to television and newspapers; (2) time devoted to television; (3) time devoted to political contents as percentage of the time devoted to television; (4) time devoted to newspapers; (5) time devoted to political contents as percentage of the time devoted to newspapers. Of these, (1) and (2) indicate an entertainment media use; (3), (4), and (5) indicate an informational media use. On the person level, the hypotheses that preferences for information increase, and preferences for entertainment decrease, informational media use, if resources are controlled for, are tested. On the country level, the hypotheses that public control increases informational media use, even if media supply and wealth are controlled for, and that public control increases the effects of preferences for information as well as for entertainment on media use, are tested. Data are 69 samples from 29 countries of the European Social Survey 2002, 2004, and 2006.

On the person level, preferences for information increase, but preferences for entertainment do not decrease, informational media use. On the country level, the times devoted to political contents in each medium, (3) and (5), differ only marginally between countries. Therefore, only the (1) television time as a percentage of the time devoted to both media and the absolute times devoted to each of the two media, (2) and (4), are analyzed. A stronger public control does increase the informational media use in many tests, and strengthen the positive impact of preferences for information on the informational media use in even more tests. But the supply of newspaper increases informational, and decreases entertainment, media use as well.

#### 1.11.5 Bibliometric study of ESS empirical production

Claudia Abreu Lopes, London School of Economics (United Kingdom)

**Abstract** This bibliometric study examines and attempts to systematize the methodological approaches of crossnational studies based on ESS data. The objective is to identify tendencies and possible mismatches between theory, hypothesis, and statistical techniques.

The study analyses 150 journal articles selected from the ESS Bibliography Archive and bibliographic databases (ISI-Web of Knowledge, Science Direct, Scopus, and Proquest) that were published over a period of seven years, from 2003 to 2009. Only articles employing secondary analysis of ESS data for more than one country are considered, in line with Hantrais and Mangens (1996) definition of crossnational research as studies particular issues or phenomena in two or more countries with the express intention of comparing their manifestations in different socio-cultural settings, using the same research instruments (p.1).

Every article is classified according to:

- a the substantive domain covered (i.e., the corresponding ESS module);
- b the number of countries included in the analysis;

- c the typology of crossnational studies proposed by Van de Vijver and Leung (1997). Four types of crossnational studies are distinguished along two axis: they can be either exploratory or hypotheses testing and may or may not involve contextual information about the countries;
- d the nature of social mechanisms captured (Hedstrom & Swedberg, 1998): individual mechanisms (purely psychological), situational mechanisms (how macrolevel conditions affect the behaviour of individuals), or transformational mechanisms (how isolated individual actions generate collective outcomes);
- e whether hypotheses are level-oriented (focused on differences in magnitudes of average scores across cultures/nations) or effect-oriented (focused on the relationship between variables and the identification of similarities and differences in these relationships across cultures/nations); and
- f the statistical technique(s) employed (one-level or multi-level).

A multiple correspondence analysis elucidates the patterns of co-occurrence of types of social mechanisms, hypotheses, statistical techniques, and categories of crossnational research. Individual studies are described in order to illustrate particular points and methodological solutions. Limitations to the exploration of social mechanisms are discussed.

## 1.12 Measurement of Social Capital

### Coordinator:

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 1.12.1 Social Capital and the Social Statistical Database

Saskia te Riele, Statistics Netherlands (CBS) (Netherlands)

Hans Schmeets, Statistics Netherlands (CBS) (Netherlands)

**Abstract** There is a growing interest in social cohesion and social capital in the Netherlands. Issues like cultural assimilation, attitudes towards ethnic minority groups, the role of the Islam and its impact on public life are in the news almost every day. The Dutch government is willing to take a substantial number of measures to enhance social cohesion in Dutch society and has made social cohesion one of the six pillars of the government policy in 2007. Similarly, the subject of social cohesion and social capital is high on the international agenda. In line with this development, Statistics Netherlands (SN) aims to publish more statistical information on social cohesion.

Social cohesion is, however, an ambiguous and multidimensional concept of which no single universal definition is available. SN has therefore recently developed a framework for measuring social cohesion, consisting of three dimensions: (1) participation (social participation, participation in organisations and participation in politics), (2) trust and (3) integration. This framework is used to provide integrated statistical information based on both registers and large-scale survey data. The Population Register contains information on age, sex, ethnicity, place of birth, place of residence, marital status and other information for all registered persons living in the Netherlands. This registration has been available since 1995, and is updated monthly. Linking this longitudinal register information to data from the social surveys and other registers creates the Social Statistical Database (SSB). The SSB thus encompasses data on all Dutch citizens based on longitudinal register information as well as cross-sectional and longitudinal surveys, and contains both individual as well as context information.

In the presentation, the measurement of social cohesion by making use of various statistical sources in the SSB will be discussed. One example is volunteering work based on some 1 million records in the Labour Force Surveys, linked to the ethnic diversity in municipalities and neighbourhoods. A further example concerns social cohesion in the neighbourhood (e.g. contacts with neighbours) defined in the Security Monitor from 2005 onwards. Also, a new survey module on social cohesion to be included in the new design for the social surveys in 2010 at SN will be discussed.

### 1.12.2 Measurement of Social Participation and its place in Social Capital Theory

Laura Guillén, INSEAD (France)

Lluís Coromina Soler, University of Girona (Spain)

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

**Abstract** In social capital theory social participation plays an important role. A lot of questions concerning social participation have been asked in the first round of the European Social Survey. In this study we tried to develop an optimal measure for social participation based on these questions. However, in doing so we detected that a distinction has to be made between informal and formal participation because these two variables have different relationships with other variables like age, education, political action and happiness. It has also been found that these two participation variables had hardly any relationship with other important variables from the social capital theory such as social trust and political trust. The latter result does not devalue the validity of the developed indices for informal and formal participation but suggest that these important components of social capital should be seen as formative indicators for social capital.

### 1.12.3 Measuring social capital and its influence on voting behaviour at the aggregate regional level

Johanna Willmann, University of Vienna (Austria)

**Abstract** My approach to social capital stems from a behavioural research question: how to solve collective action problems. The answer I'm arguing for is: social capital helps in overcoming collective action problems.

I conceive social capital as a resource of the collective, not of the individual. My definition of social capital is: Social capital are those assets that increase the skills of a social unit. In concrete, I refer to social trust, norms and social contacts that constitute social capital. To attribute social capital to the collective and not to the individual makes it necessary to measure it at the collective the aggregate level. The choice for the regional level has practical reasons as this is the smallest possible unit in the ESS that covers at least 100 interviews within each unit to aggregate and it is the biggest possible unit to allow for the comparison of 140 aggregate units, which allows for better statistical comparison than the usual cross-20-country approach.

#### 1.12.4 A new approach to measure individual social capital

Gerhard Paulinger, University of Vienna (Austria)

**Abstract** Measuring a complex concept like Social Capital (SC) in a survey can be a confusing (and costly) task, especially when there is no sole definition, and accordingly a number of different possible operationalizations. It is inevitable to focus on a concise definition and formulate a compact instrument to make measurement feasible. In this paper we present a proposal for a possible operationalization of SC.

We consider Social Capital as individual asset, as useful resources, accessible through social networks that helps to achieve goals which would not be achievable otherwise. The individual value of the available SC depends on (1) the availability of network contacts, (2) the capability of these network contacts (the amount of resources), (3) the willingness of these network contacts to make their resources available and further (4) the situation and the goal to be achieved (matching of resources and goals). Further we distinguish between informal and formal SC, arising from interpersonal networks or respectively from formal organizations (state, associations, church, etc.).

We focus on Social Support as manifestation of, or "crystalized" Social Capital. We will show that it is plausible to measure SC by measuring the subjective expected certainty of support by network contacts in specific situations. This form of measurement integrates the main characteristics of individual social capital - availability, capability, willingness (and trust) into a single question. The question gets adapted and asked for a number of situations (or goals) and for formal and informal SC. The result is thus a relatively compact instrument that grows with the number of situations and goals.

The question is whether the assumed connection between the amount of SC and the subjective expected certainty of support holds true empirically and whether the complex cognitive task of evaluating the available SC overburdens the respondents or not.

## 1.13 Measuring the Quality of Democracy by means of Survey Data

### Coordinators:

Marc Bühlmann, University of Zürich (Switzerland)  
Hanspeter Kriesi, University of Zürich (Switzerland)

### 1.13.1 Independence and Legitimacy of Judicial Institutions. A Cross-Country- Investigation

Marc Bühlmann, University of Zürich (Switzerland)  
Ruth Kunz, Zentrum für Demokratie Aarau (ZDA) (Switzerland)

**Abstract** Political confidence can be seen as a precondition for a well-functioning democratic regime. Although some degree of scepticism towards the government and the political elite is considered healthy for democracies, the performance of democratic institutions heavily depends on the citizens' confidence. This is particularly important for judicial institutions: the principal capital of courts is institutional legitimacy, which can be equated with diffuse support or confidence. The effectiveness of the courts in protecting the rule of law and controlling the other powers two prerequisites for a high quality of democracy depends on its degree of legitimacy, thus on the citizens' confidence.

In our contribution we aim at finding possible determinants for public confidence in the courts. From a neo-institutionalist perspective, we argue that institutional settings affect individual trust in the judiciary. Thus, by conducting multilevel analyses, we show that judicial independence fosters an individual's confidence in the courts. Thus, to strengthen the legitimacy and the effectiveness of the judiciary, judicial independence is important. However, this conclusion on the macro level becomes only possible when we include the micro level, i.e. the view of citizens measured with survey data.

### 1.13.2 Election Observations Missions: How to Assess Democracy?

Hans Schmeets, Statistics Netherlands (CBS) (Netherlands)

**Abstract** Based on the 1990 Copenhagen Commitments, the Organisation for Security and Cooperation in Europe (OSCE) observes the elections of its 56 member states. For this purpose an Election Observation Mission (EOM) is on the ground to assess the various phases of an election: before, on and after election day. In order to assess the elections, the OSCE member states are requested to send international observers: long-term observers to cover the whole election period during 2 to 3 months, and short-term observers who stay for a week around election day. The OSCE Office for Democratic Institutions and Human Rights (ODIHR), based in Warsaw, provides the methodology and coordination framework for such an EOM.

Each year, the ODIHR deploys thousands of observers to monitor elections throughout the OSCE region in order to assess participating States compliance with OSCE election-related commitments. The number of short-term observers may vary of some 100 to 200 in small EOMs to 1,400 in the re-run of the second round of the presidential elections in the Ukraine on 26 December 2004. Observers, operating in pairs, visit approximately 12 polling stations during election day and produce their findings for each polling station in an observer report form. Apart from specific problems and irregularities outside and inside the polling station, observers are requested to provide an overall assessment about the voting and counting process. A vast majority of the reports are processed and analysed on election day and the findings are included in a preliminary statement issued at the press conference the day after. The ODIHR publishes a final election report with recommendations within approximately six weeks after the completion of the electoral process.

The presentation deals with the methodology used in an EOM, including the data-processing, analyses and interpretation concerning over 100,000 observer report forms in 75 EOMs, in the period 1996-2008. It will be demonstrated that:

- a The specific problems and irregularities are strongly correlated with the overall assessment;
- b The observers overall assessment varies considerably between the elections;
- c The counting process is more negatively assessed than the voting process;
- d The assessments on the voting and counting process are correlated; and

e Between 1996 and 2008 the overall assessment on the voting process has been stabilized, whereas the assessment on the counting process has deteriorated.

### 1.13.3 Understanding and Qualifying Democracy: The citizens' perspective

Mónica Ferrín Pereira, European University Institute (Italy)

**Abstract** Democratic regimes are today the most common political regime according to Freedom House reports and democracies are considered the most legitimate type of regimes. However, many voices are held that highlight a decline in public opinion towards the functioning of democracy and its institutions. Disaffection and apathy are the actual tendencies in the main consolidated democracies (Torcal and Montero, 2006). This is why, over the last years, scholars have been strongly concerned about the quality of democracy and the way democratic systems should be improved (Diamond and Morlino, 2005; Schmitter, 2003, etc.). However, there seem to be no consensus about how democracy should be enhanced. More discouraging, there still is no unique definition of democracy. Therefore, when trying to evaluate the quality of democracy (either from a descriptive perspective or from a normative point of view), scholars diverge in their results. Even critically, specialists have mainly addressed the topic of the quality of democracy without taking into account the opinions of citizens.

The classical question about satisfaction with democracy has been included in almost all surveys focusing on the performance of political systems. However, questions regarding the object of dissatisfaction are seldom included. As a result, we know that people are dissatisfied with democracy, but we rarely know what they really are dissatisfied about. The very first element to be acknowledged/ measured should be the meaning of democracy, so that we later know what is being evaluated by the individuals. Only then can we assess for the quality of democracy from the perspective of citizens. Therefore, the objective of this paper is to provide an overview of the possibilities for studying citizens preferences and expectations towards democracy through existing surveys. Both conceptual and measurement limitations will be accounted for, providing an in depth analysis of the possibilities of actual surveys for the study of the quality of democracy.

### 1.13.4 A proposal of measurement of democracy based on citizenship. The case of Spain

Irene Palacios Brihuega, Instituto de Estudios Sociales Avanzados de Andalucía (IESA) (Spain)

Braulio Gómez Fortes, Instituto de Estudios Sociales Avanzados de Andalucía (IESA) (Spain)

Laura Cabeza, Instituto de Estudios Sociales Avanzados de Andalucía (IESA) (Spain)

**Abstract** In 2008, democracy in Spain reached the thirty-year mark. During this time, Spain has achieved great social, political and economic development, and has become one of the most advanced democracies in the world. Nevertheless, there are clear faults and deficiencies in the system, as in other developed democracies in the region.

In recent years, empirical research has focused on studying what makes a democracy good or bad, and on creating indices that classify systems. Democracy in Spain, once maturity and consolidation is reached, must be subject to this analysis, and a good moment for this is the celebration of its thirtieth anniversary.

Hitherto, few studies in Spain have dealt with this topic. Most have focused on measuring specific aspects of the democracy, like the control of government (Maravall and Sanchez-Cuenca, 2007) or the quality of political representation (Martinez, 2006), but have not thoroughly studied the quality of the system in total.

This paper shows the outcome of the first project in Spain aiming to measure the quality of its democracy in a comprehensive way. This research, led by the Instituto de Estudios Sociales Avanzados (IESA), is based on a survey allowing citizens to evaluate the outcome of our democracy, following the line begun by O'Donnell (2003) in his studies of the Democratic Audit in Latin America.

In this survey, people can express their opinion about five dimensions of democracy: (1) legitimacy of democracy, (2) bureaucracy, Rule of law, and corruption, (3) parties and representation, (4) Government accountability, (5) and civil society.

One of the new features of this survey is its sample design, that includes some sub-samples of the independent communities that allow us to measure, in a comparative way, to what extent democracy in Spain has developed with the same intensity in all regions within the state.

## 1.14 Network Survey Measurement

### Coordinator:

Anuška Ferligoj, University of Ljubljana (Slovenia)

### 1.14.1 How (not) to measure social support networks? The name generator vs. the role relationship approach

Valentina Hlebec, University of Ljubljana (Slovenia)

Tina Kogovšek, University of Ljubljana (Slovenia)

**Abstract** Based on several experiments (e.g., Kogovsek and Hlebec, 2008) which compared the name generator and the role relationship approach to measure social support networks with regard to network composition (proportions of family, friends, neighbors and co-workers) some of the more important findings were as follows:

- there are larger differences in network composition with only one choice taken into account;
- there are larger differences for the most important relationships (especially partner and friend, but also close kin);
- there seems to be no systematic method order effect; the differences are much larger when comparing network composition estimated on one or two choices than comparing it with role relationship being administered in the first or in the second measurement;
- similar conclusions also hold for different support types separately.

However, in the previous studies both approaches differed in several methodological aspects (e.g., question wording, limitation of the number of named alters, ranking the named alters), which made the interpretation of results quite complex. Therefore, an experiment was designed in which all factors that were found to interact with network composition and the two approaches were controlled for. Only two choices were allowed for both the name generator and the role relationship approach, question wording was the same as well as ranking of the named alters was present in both approaches. Altogether seven questions measuring four theoretically defined social support types were measured in a split ballot experiment. Respondents were randomly assigned to the name generator and the role relationship approach. Based on previous studies several hypotheses were formulated and tested:

- as all the differences in measurement instruments from the previous studies were controlled for only minor differences in frequency distributions of network composition were expected;
- differences were expected only for roles that are not uniquely defined in the role relationship approach (e.g., friend, neighbor, co-worker);
- differences in frequency distributions would vary across different social support types and across different roles (friend, neighbor, co-worker) the more frequent they would be in the name generator approach the larger the difference between the two approaches;
- differences were expected in category no one as this category was explicit in the role relationship approach, but not in the name generator approach;
- dispersion of roles should be higher for the role relationship as weak ties were explicitly offered to respondent, however the overall effect should be small as only two ties were assessed.

Results will be presented and discussed in comparison with the previous findings.

### 1.14.2 Measuring Resource Richness: On the Quality of an Interval-Scaled Resource Generator

Julia Häuberer, Charles University of Prague (Czech Republic)



**Abstract** Networks contain resources denoted as social capital by Nan Lin (2001) that can be used to create returns, i.e., to attain goals (like finding a better job) after their mobilization. Van der Gaag and Snijders (2005) developed a great tool for their measurement The Resource Generator which is the starting point of the paper. In the frame of the telephone survey Social Relationships among Czech Citizens 2007/2008 we aimed to develop a Resource Generator suitable for the use in the Czech Republic. In contrast to the original version only 12 resources are included that seem most appropriate in the Czech context. It was controlled for tie-strength including access to the specific resource through family members, friends, and acquaintances known from the membership in an association. Though, the applied Resource Generator contains a crucial enhancement: It asks for a concrete number of family members, friends and acquaintances that could provide the resource in question. This allows accounting for multiple accesses to the resource from the three mentioned groups and therefore may be a proxy for (potential) resource richness of the respondent. Furthermore, this operationalization provides interval-scaled variables allowing us to apply basic statistical methods, like factor analysis.

However, the quality of this kind of measurement is questionable. Therefore, we conducted a test-retest study interviewing the respondents at two points in time. Using nonparametric correlations (because of a non-normal distribution of the variables) and structural equation modeling (using AMOS) we prove a high test-retest reliability of the variables. To assess the construct validity it is analyzed whether we can find a similar structure as Van der Gaag and Snijders did in the Dutch context. Our included resources allow to measure (only) three types of resources social support resources, prestige related resources and financial skills resources. The existence of these three factors is approved by confirmatory factor analysis (using AMOS).

### 1.14.3 The impact of the family network on fertility attitudes

Suzanne Noordhuizen, Tilburg University (Netherlands)

Paul M. de Graaf, Tilburg University (Netherlands)

Inge Sieben, Tilburg University (Netherlands)

**Abstract** In this paper, we study to what extent individual fertility attitudes are affected by the fertility attitudes of close family members: partners, siblings, parents and adult children. Two waves of the Netherlands Kinship Panel Study (NKPS) are used. This dataset contains multi-actor data: family members are interviewed separately. So far, it has often been claimed that similarity plays an important role in selecting friends. According to previous studies, individuals prefer others with similar ideas and values over those that are dissimilar to them. In addition, once friendships are established friends affect the attitudes and values of their network members. The question rises what happens when people cannot choose their network members themselves. The family is an important example of such a fixed network. The study described here advances our understanding of the influence of the family network with respect to fertility attitudes. Particular attention is given to interaction effects between family members fertility values and relationships characteristics, such as contact frequency and perceived relationship quality.

### 1.14.4 Appearance of trust in regional, co-operative organizational networks

Siru Korkala, University of Turku (Finland)

**Abstract** The prerequisite for successful social networking is trust, which may appear in many various ways. In this study, the concepts of Systemic trust and Personal trust by Adam Seligman are being scrutinized. Systemic trust refers to intrinsic norms and regulations of the organizational networks. On the other hand, Personal trust appears between autonomous individuals. Both these are present in any social network, and certain structural characteristics define which type is more emphasized. In this study, three co-operative, organizational networks are studied, and the main research method is Social Network Analysis (SNA). Interviews and questionnaires have been also performed in order to complete the results. Distinguished structural characteristics are density, degree of centralization, and number of cliques in networks. In addition, structural holes measures, namely constraints, effective size and hierarchy are calculated and verified.

The study demonstrates, how high density, low level of centralization and low number of structural holes tend to emphasize systemic trust. Especially the division of relative degrees of centrality and the division of constraint degrees between actors, are being scrutinized. If the relative degrees on centralities between actors distributes equally in network and if constraint degrees dont distinguish considerably this kind of network can be predicated as equal by its structure. In turn, an equal network facilitates appearance of trust on systemic level. Strong systemic trust facilitates the co-operation between actors and strengthens their commitment to the network. The most important qualitative indicators of appearance of trust are the base when network is being established and actors motivation level towards co-operation in general.

### 1.14.5 Understanding Interdisciplinarity using Social Network Analysis of Administrative Data

Mark Elliot, University of Manchester (United Kingdom)

Mark Tranmer, University of Manchester (United Kingdom)

**Abstract** This paper describes a research project using social network analysis and attitudinal measurement to investigate research collaboration within the school of social sciences at Manchester. The project is an interesting application of social network analysis, poses some important methodological questions particularly around the merging of different data and the weighting of temporal decay functions and might also prove a useful policy tool in terms of understanding research collegiality and collaboration within an academic institution.

Project objectives

- 1 To build, by merging several data sources, a current social network of the school of social sciences at Manchester University.
- 2 To investigate the combination of attitudinal and administrative data within a social networks framework.
- 3 To develop principles for weighting decay functions on network input data.
- 4 To develop measures of collaboration and interdisciplinarity and show how these may have change over time.

The paper describes work on these four objectives which is taking place during spring 2009.

### 1.14.6 Comparing two methods for estimating size of small population groups using network sampling

Volodymyr Paniotto, Kiev International Institute of Sociology (Ukraine)

Tatyana Petrenko, Kiev International Institute of Sociology (Ukraine)

**Abstract** It is a frequent need in sociology to estimate size of small subpopulations, such as not numerous ethnical groups, socio-cultural groups and people at high risk for HIV infection, particularly injecting drug users (IDUs), men and women involved in commercial sex and men having sex with men (MSM). Using customary representative surveys to resolve this task requires samples of too big size to cover such groups. To increase efficiency of sample surveys social network approach may be used in this case respondents are asked to report not only about themselves, but also about people they know (social network). In such a way fractions of target populations among the people respondents know are estimated, then the data are used to estimate subpopulations size in the population. Estimation of social network size is the key point in this method of size estimation. The two methods for estimation of network size in representative surveys have been developing in the last years: scale-up method and summation method. The scale-up method (Killworth et al.) uses questions to estimate numbers of persons whom a respondent knows among the subpopulations with known size (men named Pavlo, women who gave birth in 2008, etc.). Having these data, a researcher can estimate each respondents network size. The summation method (McCarty et al.) takes as a network size the sum of people with whom a respondent report to be in 16 types of relationships (relatives, neighbors, co-workers, etc.)

We conducted a nationally representative survey in Ukraine (sample size is about 11000 respondents) using scale-up method to estimate size of subpopulations at high risk for HIV infection. Besides, on a 400 respondents sample we conducted a survey using summation method. In our presentation we compare results we received using both methods and analyze merits and demerits of each method.

## 1.15 The study of left and right in political science

### Coordinators:

Javier Astudillo, Universitat Pompeu Fabra (Spain)

Ana Sofia Cardenal, Universitat Oberta de Catalunya (Spain)

### 1.15.1 Testing the Validity of Left-right Self-placement Question by Way of the Immediate Validation Technique

Jiri Vinopal, Institute of Sociology, Academy of Sciences of the Czech Republic (Czech Republic)

**Abstract** The paper demonstrates the application of immediate validation technique onto the classic standardized left-right self-placement question, often used in surveys.

Immediate validation is to some extent original technique for the verification of validity of survey question in a standardized questionnaire that has been developed by the author in theoretical and practical level. It draws upon the methodological stream of Cognitive Aspects of Survey Methodology (CASM) and concentrates on errors that grow out of the questionnaire and that emerge in the phase of interpretation of the question by respondent. It verifies the meaning of the target question/term by means of cognitive techniques and it conducts so right in the course of the main survey.

The possibilities of the usage of the technique will be demonstrated mainly via testing the validity of standard left-right political orientation self-placement question, which was conducted by the author in a series of methodological survey experiments. When applied to these terms the immediate validation method confirmed that these terms connote things from various different areas to different people, ranging from general categories or political parties to the finer aspects of individual civic freedoms and that such differences in the perception are then often related, for example, to the position on the left-right scale itself. Moreover, it will be shown, that roughly one-half of people perceive the terms leftwing and rightwing each from a different perspective that introduces the question of unidimensionality of the scale itself.

Important part of the presentation will be also the analysis of validity of the method itself that will be demonstrated upon the data of consequent methodological survey experiment. The conclusions suggest that the method is really able to map the content of short-term memory, which is the object of context effects, and does not merely investigate more stable long-term memory structures.

On the other hand, the paper does not intend to disqualify the question itself because of its questionable content validity, and concludes in the sense of constructive usage of deeper information obtained; that has to be put beside the fact, that construct validity of the question has been already approved for many times.

On the example of simple left-right question, immediate validation is described as a technically easy technique that can interestingly improve the methodological and informative value of social surveys. The possibility of including it in the main survey, which makes its results quite strong in a statistical sense, is emphasized as well as the possibilities of further enhancements by means of experimental designs of data collection process.

### 1.15.2 Ideological representation in Portugal: MPs-electors linkages in terms of left-right placement and substantive meaning

André Freire, Higher Institute for Labour and Business Studies (ISCTE) (Portugal)

Ana Maria Belchior, CIES-ISCTE (Portugal)

**Abstract** In the academic literature on substantive political representation the analysis of congruence between elites and masses is a central topic. Building on two recent surveys conducted in Portugal (2008), one based on a representative sample of the adult population (N=1350) and the other based on a sample of MPs (N=141 out of 230 MPs), the present study focus on the problem of congruence between elites and masses in terms of the left-right divide and, although only in a complementary fashion, also in terms of the libertarian-authoritarian divide.

The paper tries to answer several questions both by comparing the two samples overall and by comparing segments of the two samples (namely across party divides). Additionally, several alternative measures of ideological and/or policy congruence will be tested.

First, the paper analyses the level of congruence in terms of left-right self-placement (in an eleven point scale: 0-10) of both MPs and electors, overall and across party groups.

Secondly, using two sets of indicators of policy preferences that can be said to be organized in terms of the left-right (attitudes toward socioeconomic inequalities, the role of state intervention in economy and society, the role of the market, etc.) and the libertarian-authoritarian (attitudes toward hierarchies, authority, alternative life-styles, etc.) divides, the paper analyzes the level of congruence of both MPs and electors, overall and across party groups.

Third, relying on MPs and electors answers to two open-ended questions about the meaning of left and the meaning of right, the substantive meaning of left-right divide is compared across MPs and electors, both overall and across party divides.

### 1.15.3 Non-location on the left-right scale in Europe: a comparative analysis of its magnitude and factors

Lucia Medina, Institut de Ciències Polítiques i Socials (ICPS) (Spain)

**Abstract** Most Europeans prefer or identify with one position on the left-right scale, and these preferences or adscription feelings 1) constitute one more feature of their political identity; 2) contribute to construct their perception on political objects; 3) help to reduce the costs associated with acquisition of information about the different political proposals; 4) appear as the most widespread way of representing and articulating party competition; 5) and finally, mediate in the electoral choice anchoring the vote.

Analysing data from the first wave of the European Social Survey (2001-2002), this paper deals with the ability of European electors to locate themselves on the left-right scale. Are Europeans from some countries more capable of self-locating than others? In this case, what do these different levels of placement depend on?

I will proceed in three steps. Firstly, I will compare the level of non-location on the left-right scale across countries; non-location means here that the interviewed says he or she does not know where to place her or himself on the scale. Secondly, I will test three hypotheses, suggesting the existence of a negative relationship between, on the one hand, left-right scale and, on the other hand, longevity of democracy, cognitive mobilization and partisan ties. It will be shown that although cognitive mobilization is the most influential of these variables, partisan ties can also exert a relevant effect, compensating the lack of cognitive and motivational resources of some citizens. Finally, I will test the hypothesis that the level of ideological polarization in each country mediates the relationship between party loyalty and location on the ideological scale.

### 1.15.4 How do voters vote when they have no ideology?

Lluís Orriols Galve, University of Oxford (United Kingdom)

**Abstract** Although they make up a considerable share of the electorate, the literature has overlooked the voting behaviour of voters who report no ideology. Using Spanish electoral data from the 1979-2008 electoral period, we seek to identify which alternative cues these voters may use when they cast their ballot providing ideology is not available. We do not find that evaluations of the incumbents performance have a higher influence on non-ideological voters as the retrospective voting literature suggests. Nor we find that other shortcuts such as candidate evaluations or party identification are more used by this group when they vote. Instead, our results indicate that non-ideological voters have exchanged the traditional ideological shortcut for the simple pro-incumbent voting as a decision rule.

### 1.15.5 The relationship between issues and left-right ideology

Wiebke Weber, Universitat Pompeu Fabra (Spain)

**Abstract** The major body of the literature on the relationship between the value or issue-based component and the left-right self-orientation assumes that individuals values and attitudes towards different issues will determine their position in the left-right dimension. Regarding values, it is assumed that these are stable over (a long period of) time and hence, affect the left-right self-orientation. Yet regarding the impact of issues this seems to be less clear, in particular as it was found that issues differ over time and cross-nationally.

Given the dynamic of changing issues, it seems reasonable to expect them to have an effect on individuals left-right orientation but given the consistency of ideology, one can expect the opposite effect as well. Moreover, not all issues are equally important to different individuals and hence, the relationship between issues and left-right ideology might vary. This will be tested by using multigroup structural equation modelling with data from the first wave of the European Social Survey (2002/03) from the Netherlands.

## 1.16 Time use surveys: emerging issues (I)

### Coordinator:

Mario Cools, Hasselt University (Belgium)

### 1.16.1 The use of time use surveys to study activity participation: emerging issues

Mario Cools, Hasselt University (Belgium)

Elke Moons, Hasselt University (Belgium)

Geert Wets, Hasselt University (Belgium)

**Abstract** Time use surveys often are carried out to identify, classify and quantify social behaviour of people by focusing on the activities that people perform. Time use data, in the transportation field often referred to as activity-based data, can be used to study a variety of sociological, economic, and technological phenomena. Studying behaviour, social networks and transport patterns are a few of the topics that can be analysed based on these time use data.

This first introductory paper discusses some emerging issues in the collection of travel-related data such as automatic spatial information recording using cell phones and GPS, and survey design experiences. On the one hand new technology offers the opportunity to record at a relative low cost a wide variety of the data, while on the other hand new problems occur. One of such problems for instance is the problems of cold starts faced with GPS devices, clouding the first minutes of recording after each restart of the device after it was turned off completely. Notwithstanding, applications of new technologies offer the opportunity for performing detailed space-time analyses in different fields ranging from epidemiology to transportation science. Some of these applications are highlighted in this session.

Next to the advantages and potential risks of new technologies, this introductory paper also discusses the combination of different modes to analyze space-time behaviour. In particular, this study investigates potential mixed mode design effects, observed in a large activity-based travel survey, using a PDA application on the one hand, and traditional paper and pencil diaries on the other hand. The mixed-mode effects are analysed using heteroscedastic linear regression models, taking into account not only mode-effects, but also potential fatigue-effects. The results show that in this mixed-mode survey no attrition effects are present, and that the survey mode (PDA versus paper and pencil) has no direct impact on the quantities investigated (number of out-of-home activities reported and number of trips).

### 1.16.2 Comparing a 2 and 7 day registration diary in time-use research Evaluating the EUROSTAT guidelines

Ignace Glorieux, Vrije Universiteit Brussel (Belgium)

Joeri Minnen, Vrije Universiteit Brussel (Belgium)

**Abstract** Time budget surveys using diaries differ in the number of observation days. Since the 1990s EUROSTAT developed guidelines for EU and other countries to make time budget data from different countries comparable. The EUROSTAT guidelines recommend a two-day diary with both one weekday and one weekend day. Since a lot of activities follow a weekly rhythm, time use researchers often opt for a 7-day registration. A 7-day registration is more in line with the social relevant week cycle. With a two-day registration it is impossible to reconstruct the week cycle on an individual level, it is even impossible to reconstruct the weekend. It is well possible however to compose a synthetic week on an aggregated level.

In this paper we use the pooled Flemish time budget data from 1999 en 2004 (in total 3313 respondents) to analyse in full detail some of the advantages and disadvantages of a 2 and 7 day registration. The respondents the Flemish time use surveys all filled in diaries for 7 consecutive days. To simulate the 2-day registration, we randomly selected one weekday and one weekend day for each respondent. The 2 day selection was compared with the original 7 day registration. More particularly we investigate the differences between a synthetic week and a fully registrated week. The aim of this comparison is to inventory the advantages and disadvantages of the 2 day and 7 day registration method. To do that we compare different indicators, as time use per respondent and per participant, the standard deviation of the time use per respondent, the effect of different back ground variables, We also examine whether certain types of activities are more affected by the method of registrations

than others. Finally we examine whether a longer period of registration has effects on the quality of the data (less detail and less accurate).

### 1.16.3 Number of episodes as a measure for unobserved item nonresponse of diaries

Paavo Väisänen, Statistics Finland (Finland)

**Abstract** In sample surveys, non-response rates are generally reported as an indicator measuring data quality. Unit non-response and item non-response rates are reported for individual questions and study variables. In the diaries, an episode is defined as a time slot denoted by the same code. Missing episodes are difficult to observe because respondents record activities and its duration into diaries by themselves. Unobserved item non-response arises when a respondent forgets to record an activity in the diary, and this situation occurs when, for instance, a person travels home from work and stops for shopping. If the shopping stop is not recorded then the number of episodes is two episodes too low. The average episode numbers describes then some kind of filling intensity of diary keeping and missing episodes are unobserved item nonresponse. Diaries are used to collect data on the simultaneous activities or secondary activities of individuals, who may as well forget to record it in a diary. The total of secondary activities can also describe the filling intensity of diaries, and it forms the same kind of measure for item nonresponse. A large number of episodes and a high total time of secondary activities indicate valid diary data. Respondent's time use has an unwanted impact on this measure and, for example, persons with long working hours have usually fewer activities to report.

A quality indicator based on the number of episodes was studied using the diary data of the Finnish Time Use Survey, which was conducted using the guidelines of the Harmonised European Time Use Survey (HETUS). Around 20-25 activity episodes, on average, are usually regarded as a reasonable value for the indicator of well filled diaries. Comparisons can be made between the fourteen countries which made their time use surveys following the guide lines of the HETUS, but comparisons with other surveys, in which different diaries, days, instructions for diary keeping or coding are used, are not valid. In the countries included in the Harmonised Time Use Survey Database, the number of episodes was the highest in Sweden (26 episodes) and the lowest in France (19 episodes). The average total for secondary activities was 170-190 minutes, and slight correlation (from 0.11 to 0.34) was found between these quality measures. The amount for reported secondary activities was the highest in France (351 minutes) while the lowest total for simultaneous activities was reported in Spain (82 minutes). The number of episodes depended on gender, age and education level.

### 1.16.4 Intra-household correlation and the effect on time-use estimates

Joeri Minnen, Vrije Universiteit Brussel (Belgium)

Ignace Glorieux, Vrije Universiteit Brussel (Belgium)

**Abstract** Since the first time-use research project by Alexander Szalai in the 1960s a number of conventions with regard to the methodology of time budget research have been established. Subsequently, time budget research has evolved to be an established research method. Nevertheless, time budget surveys used different methodologies depending on the moment of recording the information, the number of recorded dates, the coding of the activities, the recording of secondary activities and the respondents that are approached.

EUROSTAT is developing initiatives on a European scale to streamline time budget surveys from various individual countries. With this objective in mind EUROSTAT established the HETUS-survey in the early 1990s. Statistics Belgium gathered time-use data in 1999 and in 2005 according to EUROSTAT guidelines.

In this paper we want to focus on the unit of analysis. Most time-budget data sets are organized on an individual level. However, some data sets, as the EUROSTAT micro data sets, include more than one individual of the same household. More specifically, all members of the household 12 years or above (in Belgium) asked to register their activities. Since it would not be surprising that members of the same household share the same values, behaviour and attitudes that could effect time-use, analyses on the level of households risk intra-household correlation. Also the fact that people of the same household spend time together could lead to biases in time-use estimates on the individual level. In this paper we want discuss this method of collecting time-use information and the use of it. We prepare some simulations using a multilevel technique to make the consequences of this methodological option visible.

## 1.17 Time use surveys: emerging issues (II)

### Coordinator:

Mario Cools, Hasselt University (Belgium)

### 1.17.1 Timing of life course in emerged adults

František Baumgartner, Academy of Sciences of the Czech Republic (Czech Republic)

**Abstract** Emerging adulthood is a transitive period of personality development with a focus on ages 18-29. The typical features are the tendency to change and explorations of different life possibilities in variety of areas, especially love and work. This exploration offers the best opportunity for clarifying identities of young people. Our broad empirical research was devoted to different psychological characteristics of emerged adults. The part of it came under the context of life course study.

The life course approach is the multidisciplinary paradigm representing views of different disciplines to study of individual human life in sociocultural perspective. The life course is understood as a sequence of social defined events and roles.

The data was collected in last year. Research sample was comprised by students of Masaryk university in Brno. The contribution concentrate how respondents structure their life, how they evaluate the importance of some events for attain of adulthood and what their future plans are. The findings are compared with results of ESS project (Round 3).

### 1.17.2 Measurement of need satisfaction in the realm of time use

Krzysztof Zagorski, Kozminski University (Poland)

**Abstract** Dominant paradigm in time use research consists of the time budget analyses, i.e. investigating the length of time devoted to various activities. The paper will present methodology and substantive results of a quite different approach. Instead of time span devoted to particular activities, subjective feelings have been investigated, whether the respondents devote too much time, as much time as they want or less time than they would like to spend for each activity. People who do not spend time on particular activity are asked, whether they have no possibility but would like to spend the time on it, or whether they do not want it. Thus, the proposed paradigm consists of the analysis of need satisfaction in the realm of use of time rather than reconstructing the time budget composition. Seventeen categories of various activities are distinguished. The data are from a very big survey on a representative sample of the whole Poland (38 thousand respondents) plus separate 64 small samples of populations of selected gminas, which are smallest Polish administrative units (another 25 thousand respondents). The survey has concerned broadly understood living conditions and contained a section of time use questions. The satisfaction of needs to spend time on various activities will be analyzed in their relations to socio-demographic characteristics of the respondents and their households, their life satisfaction and the characteristics of their regions of residence.

### 1.17.3 Roundtable discussion concerning identification and prioritizing of key issues in time use surveys

Ignace Glorieux, Vrije Universiteit Brussel (Belgium)

## 1.18 Trust and Trustworthiness (I)

### Coordinator:

Patrick Sturgis, University of Surrey (United Kingdom)

### 1.18.1 A Genetic Basis for Social Trust?

Patrick Sturgis, University of Surrey (United Kingdom)

Sanna Read, University of Surrey (United Kingdom)

Peter K. Hatemi, University of Iowa (United States)

Gu Zhu, Queensland Institute of Medical Research (Australia)

Tim Trull, University of Missouri (United States)

Margaret J. Wright, Queensland Institute of Medical Research (Australia)

Nicholas G. Martin, Queensland Institute of Medical Research (Australia)

**Abstract** A propensity to believe that fellow citizens will not act against our interests in social and economic transactions has been identified as key to a host of favourable outcomes for both individuals and society. Yet the causes of this type of generalized or social trust are far from clear. To date, researchers within the social sciences have focused almost exclusively on social-developmental and political/institutional features of individuals and societies as the primary causal influences. In this paper we investigate the intriguing possibility that social trust might have a genetic, as well as an environmental basis. We use data collected from samples of monozygotic and dizygotic twins to estimate the additive genetic, shared environmental, and non-shared environmental components of trust. Our results show that the majority of the variance in a multi-item trust scale is accounted for by an additive genetic factor. On the other hand, the environmental influences experienced in common by sibling pairs have no discernable effect; the only environmental influences appear to be those that are unique to the individual. Our findings problematise the widely held view that the development of social trust occurs through a process of familial socialization at an early stage of the life course.

### 1.18.2 Generalized Trust and the country context - does it matter and why?

Ursula Dallinger, University of Trier (Germany)

**Abstract** In recent research generalised trust is much discussed as a resource that facilitates economic exchange, political participation and government, and even fosters the welfare states viability. Countries with different levels of generalised trust are quite systematically spread across Europe. High generalised trust can be found in the Scandinavian, low levels in the eastern and middle European countries, other countries somewhere in between. However, what exactly are the reasons for this international differences in this important social resource, what are the country level characteristics?

The paper will explore the causes of varying levels of generalised trust adopting above all country level explanations, because the degree to which the variation of the dependent variable (generalised trust) can be explained on the country level, is (according to IKK) higher than regarding other research topics/ dependent variables. Especially the impact of political institutions and of the inclusiveness of welfare states usually are seen as the correlates of high generalised trust. The paper will test hypothesis on both country and individual level by multilevel regression using the European Social Survey data (2006) with additional country variables.

### 1.18.3 How generalized is generalized social trust? A cross-national estimation of trust radii

Jan Delhey, Social Science Research Centre Berlin (WZB) (Germany)

**Abstract** It is commonly agreed that generalized interpersonal trust is doing a lot of good things to communities and societies. Much of the research on trust is based on the standard question: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" It is assumed that countries in which many citizens choose "trust" are characterized by a high level of trust in a wide circle of people generalized social trust. However, actually we have very little information about who "most people"



are for the respondents. In other words: we do not really know to what extent our standard measure actually measures generalized trust, and whether it is working better in some countries than in others. It is perfectly possible that people in some of the high-trust countries think primarily of kin and family as "most people", hence their high level of trust could be confined to a narrow circle of familiar and similar others only.

With data from the most recent World Value Surveys wave 5, I analyse how wide the radius of trust is in 45 countries. With the help of a new item battery specifying targets of trust such as "family" or "people of other religion", it is possible to measure the impact thick (particularized) trust and thin (generalized) trust have on the answers to the standard trust question. Thereby we can indirectly approach which kind of people respondents primarily have in mind when rating the trustworthiness of "most people". The analysis reveals (1) that in most countries, trust as measured with the standard question depends more on thin trust than on thick trust; and (2) that the radius of trust varies considerably between societies. The latter means that with the standard question we tend to overestimate the level of generalized trust in some countries (e.g. in China and South Korea), while underestimating it in others (e.g. in Italy and Poland). It is further shown that the radius of trust varies systematically across countries, with more modern and individualistic societies having wider radii of trust.

### 1.18.4 Measuring Trust: Experiments and Surveys in Contrast and Combination

Michal Naef, Royal Holloway, University of London (United Kingdom)

Jürgen Schupp, DIW Berlin (Germany)

**Abstract** Trust is a concept which has gained much importance in economic theory and research within the last two decades. It has been used in a number of contexts and has been researched both as an explanatory as well as as dependent variable. In view of this relevance, we explore the questions of what exactly is measured by the diverse survey-derived scales and experiments claiming to measure trust, and how these different measures are related. Using nationally representative data, we tested a commonly used experimental measure of trust for robustness to a number of interferences, finding it to be mostly unsusceptible to stake size, the extent of strategy space, the use of the strategy method, the framing of the instructions, and the characteristics of the experimenters. Inspired by criticism of the widespread trust question used in many surveys, we created a new, improved survey trust scale consisting of three short statements. We show that the dimension of this scale is distinct from trust in institutions and trust in known others. Our new scale is a valid and reliable measure of trust in strangers. The scale is valid in the sense that it correlates with trusting behaviour in the experiment. We demonstrate that the test-retest reliability of six weeks is high. The experimental measure of trust is, on the other hand, not significantly correlated with trust in institutions nor with trust in known others. We therefore conclude that the experimental measure of trust refers not to trust in a general sense, but specifically to trust in strangers.

## 1.19 Trust and Trustworthiness (II)

### Coordinator:

Patrick Sturgis, University of Surrey (United Kingdom)

### 1.19.1 Different Predictors of Neighbourhood Trust: Measuring social cohesion in Amsterdam

Lonneke van Oirschot, Dienst Onderzoek en Statistiek, Amsterdam (Netherlands)

Jeroen Slot, Dienst Onderzoek en Statistiek, Amsterdam (Netherlands)

**Abstract** This research investigates whether Robert Putnams (2007) renowned findings on the negative influence of ethnic diversity on social cohesion hold in Amsterdam. We expressed social cohesion in measures of neighbourhood trust.

Multilevel analysis of the data Living in Amsterdam (Wonen in Amsterdam, 2007) has pointed out that diversity is an important predictor of neighbourhood trust. Other neighbourhood aspects (education level, violent and non-violent crimes) but especially individual aspects (number of years respondent lives in the neighbourhood, home ownership, income, age, ethnic background) are additional significant predictors of neighbourhood trust. When replacing the diversity measure by the percentage of first generation non-western immigrants, this percentage proved to be an even better predictor of neighbourhood trust. However, the most important predictor of neighbourhood trust was contact with the neighbours.

Additionally, we tested whether Putnams findings (interpreted such that we expected an increase of ethnic diversity to result in a decrease of neighbourhood trust and vice versa) hold for all Amsterdam neighbourhoods. The results show a mixed picture, as the relationship between diversity and neighbourhood trust is not in the expected direction in some neighbourhoods. We will investigate the possible causes of the deviations found, by use of the important predictors of neighbourhood trust and other possible explanations (e.g. urban development).

### 1.19.2 Social Participation and Trust: Disentangling Causes and Effects

Rene Bekkers, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** In debates on social capital, participation in voluntary associations is often believed to make citizens more trusting of others. This paper reports longitudinal analyses of different forms of participation - ranging from leisure activities to volunteering - and generalized social trust using panel data from the Netherlands and the UK. Analyses of data from three waves of the Giving in the Netherlands Panel Study spanning four years (2002-2006; n=692) shows that trust is relatively stable over a four year period (.78). Fixed effects regression analyses reveal that changes in volunteering are not related to changes in trust. Trust is higher among volunteers mainly because of selective attrition: persons with low trust are more likely to quit volunteering. Analyses of 10 years of data from the British Household Panel Study show that trust is somewhat less stable in the UK than in the Netherlands. Also the UK data reveal significant effects of various forms of social participation on trust.

### 1.19.3 Is generalized trust explaining political support in Europe?

Eduarne Bartolomé Peral, University of Deusto (Spain)

**Abstract** The main goal of this proposal and the paper to be presented is to answer the research question of to which extent political support in Europe is being explained by generalized trust in European societies. Is trust explaining political support equally in all European societies? The theoretical approach for this analysis is Eastons theory of political support and will analyse the effect of trust in the support for the regime, specially the democratic performance and institutions, and in the support for the authorities. It is expected that the effect of trust will be different depending on the diffuse-specific type of support and its also expected that trust will have a stronger explanatory capacity in older European democracies than in newer democracies. The suggested method to answer this question is the multilevel analysis and data of European Social survey will be used to answer the research question.



## **Chapter 2**

# **Sampling and Nonresponse**

## 2.1 New Challenges in Sampling (I)

### Coordinators:

Siegfried Gabler, GESIS - Leibniz Institute for the Social Sciences (Germany)  
Seppo Laaksonen, University of Helsinki (Finland)

### 2.1.1 New Challenges in Sampling: Introduction

Seppo Laaksonen, University of Helsinki (Finland)

**Abstract** Quantitative survey research requires micro data. If the target population is large, sampling is necessarily the only option. There exists a general acceptance that sampling should be probability based. The available sampling frame may however be inadequate and non-response may be high and selective. These nuisances are becoming more common for many reasons. An important reason is the survey mode that is more often a telephone or a web since these are cheaper and more comfortable. However, the frames for such surveys are more restricted than in the case of face-to-face interviewing. Non-response has been generally increasing over recent decades. Moreover, multinational surveys have been becoming more common, and the success in sampling and data collection will have a crucial role in order to get comparable estimates. Also, longitudinal approaches are often desired to follow individuals over time. The session has invited contributors who share these challenges and who have new experiences in these areas. We have received a good number of papers and several above questions are considered in these papers. The introduction first gives an overview about challenges and nuisances in today and future survey research from the point of view of sampling, and then briefly summarises the papers presented in the following three sessions.

### 2.1.2 Impact of J. Neyman and W.E. Deming on Sample Survey Practice in Poland

Jan Kordos, Warsaw School of Economics (Poland)

**Abstract** The author begins with a synthetic presentation of Neymans contribution to the theory and practice of sampling. First, Neymans initial contribution to the theory of sampling in the 1920s is emphasized, and second, Neymans famous 1934 paper (Neyman, 1934) is discussed in detail. It is stressed that despite many contributions of others to the technical development of sampling in the period from 1900 to 1926, Neymans 1934 paper played a crucial role in turning pure mathematics into real sampling plans for actual randomly selected large scale surveys. The Neymans 1934 paper played a catalytic role in the actual use of randomization in the utilization of methods for random sampling (Fienberg, S.E. and Tanur, 1996).

Neyman originally prepared the paper in 1932 in Polish (with an English summary) as a booklet growing out of his practical experience, and published in 1933 (Neyman, 1933). Summary of this paper is given and the contribution of J. Piekalkiewicz, another Polish statistician, to practical aspects of the research is stressed (Piekalkiewicz, 1934).

Special attention is paid to the co-operation of J. Neyman, after the Second World War, with Polish statisticians, and mainly with the Mathematical Commission of the Central Statistical Office of Poland (Fish, 1950; Zasepa, 1958). The author was one of the Polish statisticians who participated in that cooperation in 1958. Some personal reflections of this cooperation are mentioned.

Next the author considers the Demings (1950) approach to survey design, and discusses its impact on sample survey practice in Poland. He considers planning and preparatory work of survey design: defining major objectives, preparing first operation plans, method of data collection (questionnaire design), preliminary tabulation and analysis plans, frames and sample design, pre-test and pilot studies, preparing final survey documents; survey implementation (field work), data processing; analysis of survey data, publication and results dissemination. The author points out that the Mathematical Commission of GUS accepted the Demings approach that the statistician should keep in mind the root-mean-square error (MSE), which is the resultant of the two parts sampling error, and the biases, and strive to reduce it. When the biases are large, it is uneconomical and ineffective to spend a lot of money to keep the sampling error small. As a matter of fact, often the most effective way to shorten the MSE is to decrease the bias while increasing the sampling error (Kordos, 1975,1985,). At the end some conclusions are drawn for future development of sample surveys in Poland and Demings 14 points are reminded (Deming, 1986).

## References:

- DEMING, W. E. (1950), *Some Theory of Sampling*, New York: Wiley.
- DEMING, W. E. (1986), *Out of the Crisis*, Massachusetts Inst Technology.
- FISZ, M. (1950), Consultation with Prof. Neyman and conclusions. *Studia i Prace Statystyczne*, nr 3-4. (In Polish).
- FIENBERG, S.E. and TANUR, J.M. Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling, *International Statistical Review*, vol. 64, pp. 237-253.
- KORDOS, J. (1975), 25 years of activities of the CSO Mathematical Commission, *Przegląd Statystyczny*, nr. 1. (in Polish).
- KORDOS, J. (1985), Towards an Integrated System of Household Surveys in Poland, "Bulletin of the International Statistical Institute", (invited paper), vol. 51, Amsterdam, Book 2, pp. 1.3.1 -18.
- KRUSKAL, W.H. and MOSTELLER, F. (1980), Representative sampling, IV. The history of the concepts in statistics, 1895-1939, *International Statistical Review*, Vol. 48, pp.169-195.
- NEYMAN, J. (1925), Contributions to the Theory of Small Samples Drawn from a Finite Population, "Biometrika", nr 17, pp.472-479.
- NEYMAN, J. (1933), An Outline of the Theory and Practice of Representative Method, Applied in Social Research, Institute of Social Problems, Warszawa (In Polish with an English summary),123 pp.
- NEYMAN, J. (1934), On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, "Journal of the Royal Statistical Society", nr 97, pp. 558-606.
- NEYMAN, J. (1938), Contribution to the Theory of Sampling Human Population, "Journal of American Statistical Association", nr 33, p.101-116.
- PIEKAKIEWICZ, J. (1934), Sprawozdanie z badan skladu ludnosci robotniczej w Polsce metoda reprezentacyjna (Report on Investigation of Composition of Workers' Population in Poland by Sampling Method). Instytut Spraw Spolecznych, Warszawa.
- ZASEPA, R. (1958): Problems of Sampling Surveys of CSO in the light of consultations with Prof. J. Neyman), *Wiadomosci Statystyczne*, nr 6, pp. 7-12 (in Polish).

### 2.1.3 Design effect of randomized systematic sampling

Siegfried Gabler, GESIS - Leibniz Institute for the Social Sciences (Germany)  
Horst Stenger, University of Mannheim (Germany)

**Abstract** In statistical practice systematic sampling is used in many modifications due to its simple handling. In addition, systematic sampling may provide efficiency gains if it is well adjusted to the structure of the population under study. However, if systematic sampling is based on an inappropriate picture of the population a high decrease of efficiency, i.e. a high increase in variance may result by changing from simple random sampling to systematic sampling. In the context of 2-stage designs systematic sampling so far seems often in use for subsampling within the primary units. In contrast to this practice we propose to randomize the order of the primary units to select thereafter systematically a number of primary units and, then, to draw secondary units by simple random sampling without replacement within the primary units selected. To compute the design effect the inclusion probabilities of second order must be known. First findings concerning the design effect will be presented.

### 2.1.4 The best of two worlds: Model-based and design-based approaches to estimation of design effects

Matthias Ganninger, GESIS - Leibniz Institute for the Social Sciences (Germany)

Siegfried Gabler, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Due to the lack of an appropriate frame, many surveys in the social sciences apply a cluster sample design. This has some merits (e.g. reduced fieldwork costs) but also considerable drawbacks. One of the most prominent ones is an increase in variance of estimators when, at the estimation stage, the design is ignored and survey data is treated as having arisen from a simple random sample. The factor by which the variance of an estimator is increased as compared to the variance of the same estimator under simple random sampling is the design effect.

Design effects basically arise from two sources: clustering (deff\_c) and b) weighting (deff\_p). As in most fields of survey statistics, estimators of the design effect have been proposed motivated both from the design-based and the model-based perspective. Estimators derived under either approach have their merits and drawbacks in terms of quality and extensibility. The design-based approach considers both sources of variation directly as it is based on a direct comparison of an estimator's variance under the complex and under a (hypothetical) simple random sample design. The model-based approach decomposes deff\_c and deff\_p explicitly and estimators of either component of the design effect have been proposed. Under this approach, the estimator's quality depends mainly on the quality of an appropriate estimator of  $\rho$ , the intra class correlation coefficient.

In this presentation the results of a simulation study on the quality of a number of estimators of both approaches is presented. The talk will summarize important findings and gives recommendations that can guide the researchers choice of an estimator.

### 2.1.5 Advantages of cluster sampling

Artur Pokropek, Polish Academy of Sciences (Poland)

**Abstract** Most of the survey research focuses on individuals, as separate and independent units. The samples are random, seldom include people from the same neighbourhood or particular small social groups. Almost never sampled people know each other, and seldom share the same problems. As J.S. Coleman wrote fifty years ago, which is also true today, individual in many researches is treated as "an atomistic entity" torn out of its social context. This must sound paradoxically in the field of sociology. This problem arises due to usual sample design (random sampling and stratified random sampling) and classical statistical methods that are used in many surveys. Is there any alternative? Large scale educational research (like PISA, TIMS, PIRLS) are some option here. Those researches use some classical survey methods like questionnaire. Children and parents are asked about classical issues of sociology (social status, attitudes, etc.). The important difference is that whole school or whole classes are sampled. Parents are linked and set in particular context, because of children's schools. Individuals are not sampled independently, what is considered as an advantage here. Also methods of analysis are somewhat different multilevel modelling is a main method of analyzing data. Not only relation but variations of relations in different context are taken into account. My aim is to show on some particular examples (from PISA research) how we can see that kind of research as an interesting alternative to classical surveys based on some kind of simple random sample.

## 2.2 New Challenges in Sampling (II)

### Coordinators:

Siegfried Gabler, GESIS - Leibniz Institute for the Social Sciences (Germany)  
 Seppo Laaksonen, University of Helsinki (Finland)

### 2.2.1 Sampling strategies dependent on order statistic of auxiliary variable

Janusz Wywiał, University of Economics, Katowice (Poland)

**Abstract** The sampling designs dependent on sample moments of auxiliary variables are well known. Lahiri (1951) considered a sampling design proportionate to a sample mean of an auxiliary variable. Sing and Srivastava (1980) proposed the sampling design proportionate to a sample variance while Wywiał (1999, 1999a) a sampling design proportionate to a sample generalized variance of auxiliary variables. Some other sampling designs dependent on moments of an auxiliary variable were considered e.g. by Brewer and Hanif (1983), Tille (2006) and Wywiał (2000, 2003). These sampling designs cannot be useful in the case when there are some censored observations of the auxiliary variable. Moreover, they can be too sensitive to outli observations. In these cases the sampling design proportionate to the order statistic of an auxiliary variable can be more useful. That is why such an unequal probability sampling design is proposed here. Let a fixed population be of the size  $N$ . An observation of a variable under study (an auxiliary variable) attached to the  $i$ -th population element will be denoted by  $y_i$  ( $x_i > 0$ ),  $i=1, \dots, N$ . The sample of size  $n$ , drawn without replacement from the population, will be denoted by  $s$ . The sampling design is denoted by  $P(s)$ . Let  $X(r)$  be the  $r$ -th order statistic from a simple sample drawn without replacement. Let  $G(i,r)=s: X(r)=x_i$  be the set of all samples  $s$  whose  $r$ -th order statistic of the auxiliary variable is equal to  $x_i$  where  $r \leq N-n+r$ .

Wywiał (2008) proposed the following sampling design proportional to the  $x_i$  value of the  $X(r)$  statistic.

$$P_2(s|r) = \frac{x_i}{\sum_{j=r}^{N-n+r} \binom{j-1}{r-1} \binom{N-j}{n-r} x_j}$$

for  $s$  in  $G(i,r)$ ,  $i=r, \dots, N-n+r$ .

Particularly, if  $x_i = c > 0$  for all  $i=1, \dots, N$ , the above sampling design is reduced to the  $P_0(s)$  simple sampling design.

The particular cases of the sampling design as well as its conditional version are considered, too. The sampling scheme implementing this sampling design is proposed. The inclusion probabilities of the first and second orders were evaluated. The well known Horvitz-Thompson estimator is taken into account. A ratio estimator dependent on an order statistic is constructed. It is similar to the well known ratio estimator based on the population and sample means. Moreover, it is an unbiased estimator of the population mean when the sample is drawn according to the proposed sampling design dependent on the appropriate order statistic.

Wywiał (2007) deals with an analysis of the accuracy of the strategies for estimating the mean as well as the total value of a variable under study. Three strategies called quantile types are compared with a simple sample mean, with an order ratio estimator from the simple sample mean as well as with the order ratio estimator from a sample drawn according to the sampling design proportional to the sample mean of an auxiliary variable. The comparison of the strategies accuracy has been based on a computer simulation. In the case of a small size population, the mean square errors have been evaluated on the basis of all possible samples which can be selected. In the case of a larger population, the samples have been drawn according to the considered sampling schemes. Finally, appropriate conclusions have been formulated. The results of the comparison of the introduced quantile type strategies with the well known moment-based ratio strategies are equivocal. Generally, the quantile type strategies can be more precise than the simple sample means when the degree of the order statistic is large and the sample size is small. Hence, it can be eventually useful as sampling design on the second stage of two-stage sampling design.

Key words: sampling design, order statistic, auxiliary variable, sampling scheme, estimation, strategy, accuracy comparison, relative efficiency

### References:

Brewer K. R.W., Hanif M. (1983). Sampling with unequal probabilities. Springer Verlag, New York-Heidelberg-Berlin 1983.



- Lahiri G. W. (1951): A method for sample selection providing unbiased ratio estimator. Bulletin of the International Statistical Institute, vol. 33, s. 133-140.
- Singh P., Srivastava A.K. (1980). Sampling schemes providing unbiased regression estimators. Biometrika vol. 67, 1, pp. 205-9.
- Tille Y.(2006). Sampling algorithms. Springer.
- Wywiał J. L. (1999). Sampling designs dependent on the sample generalised variance of auxiliary variables. Journal of the Indian Statistical Association vol. 37, pp. 73-87.
- Wywiał J. L. (1999a). Generalisation of Singh and Srivastavas schemes providing unbiased regression estimators. Statistics in Transition vol. 4, nr 2, pp. 259-281.
- Wywiał J. (2003). Some Contributions to Multivariate Methods in Survey Sampling. Katowice University of Economics, Katowice.
- Wywiał J. (2000). On precision of Horvitz-Thompson strategies. Statistics in Transition vol. 4, nr 5, s. 779-798.
- Wywiał J. L. Simulation analysis of accuracy estimation of population mean on the basis of strategy dependent on sampling design proportionate to the order statistic of an auxiliary variable. Statistics in Transition-new series vol. 8, no. 1, 2007, pp. 125-137.
- Wywiał J. L. (2008). Sampling design proportional to order statistic of auxiliary variable. Statistical Papers vol. 49, Nr. 2/April, 2008, s. 277-289.

## 2.2.2 The use of control samples in sampling theory

Paola Maddalena Chiodini, University of Milan (Italy)

Donata Marasini, University of Milan (Italy)

Piero Quatto, University of Milan (Italy)

**Abstract** In sampling theory from finite population sometimes is necessary to introduce some techniques that are customary use in observational studies of epidemiological type. The reason for introducing these techniques in sampling theory is that instead of estimating the population parameters of reference, we may be interested in measuring the association between a variable object of interest for the study (effect) and other variables which may be causes for it. One of the techniques used in epidemiological studies is the so-called control sample. There are also other variables, named confounding in the epidemiological literature, which could be thought as causes but which in fact are kept under control and used to divide "people" into groups of reference. So the aim of this paper is to provide a sampling plan to identify effect samples and control samples and then identify appropriate measures of association. Given that the effect can be seen dichotomy without loss of generality, indicate with  $U$  the population, with  $E$  and  $C$  the two subsets, the first formed by the units with the variable of attention  $e$  and the second where  $e$  is absent. In this context the construction of the two types of samples is done by splitting the  $N$  units of  $E$  and  $M$  units of  $C$  in  $H$  groups using confounding variables independent by  $e$ . The  $N_h$  units of the  $E$  subset are the  $h$ -th effect group and the  $M_h$  units of the  $C$  subset are the corresponding control group. The  $N_h$  units are similar to each other compared with confounding variables and also happens to  $M_h$  units of the control group which, in turn, are similar to units of the effect group. The confounding variables used to divide people into groups should not be thought as auxiliary variables with which stratify the population or as naturally existing variables with which build clusters designed as at first stage. A first problem, of not trivial solution, is to establish a basis on which set the sample sizes  $m_h$  in each of the  $H$  groups; another, even more difficult, is to identify an appropriate measure of association which takes into account the measures in each of the groups. These are the two objectives addressed in this work.

### 2.2.3 Sampling Frame Coverage and Domain Adjustment Procedures for Internet Surveys

Zerrin Asan, University of Anadou (Turkey)

Öztaş Ayhan, Middle East Technical University (Turkey)

**Abstract** The objective of this work is to study, define and compare alternative sampling frames (population lists) for the representative population coverage as a base for sample selection in internet surveys. In addition, the study aims to provide a methodology for the domain adjustment procedures which enables to use the latest available information. There are several types of sampling frames such as target population frame and available list frame. The target population frames are the ideal frame, which cannot easily be obtained. On the other hand, total population list frames can be either household or individual unit based. The quality of the existing sampling frames can be different. The sampling frames may be; undefined, ill defined, partially defined, or perfectly defined.

The target population of the sample information collected from internet surveys (e-mail surveys & web surveys) faces many difficulties in the identification of the possible sampling frame. For voluntary participation, one has to identify the coverage of the target population. Within the available data sources, special adjustments are proposed for the small domains. Some basic variables can be proposed for this purpose. Adjustments can be made for gender breakdown, age groups, and education groups.

In terms of sample selection, the type of access to internet surveys can be based on several limitations. Early attempts was based on the restricted access designs for e-mail survey, or e-mail message followed with a web survey. As an alternative, voluntary participation designs became common for web surveys. In this type of design, observations are obtained through haphazard entry to the web survey questionnaire. Therefore, a probability sample cannot be obtained, due to unknown selection probabilities. Alternative data adjustment procedures for the case of voluntary participation is proposed in this study. Several types of weighted location estimators are proposed on the basis of these designs. For surveys having complex sample designs, a combined ratio mean (proportion, mean, or ratio) or separate ratio mean alternatives can be proposed.

The overall selection probabilities for households can be determined by obtaining household based information from the population information for each domain. This information can be obtained by dividing the total population of each domain by the corresponding average household size for this domain. The overall selection probability of the domains will be their overall sampling fractions. From the previous experience with the variability of the average household sizes in different domains, an alternative approach is also proposed. Due to the unavailability of the census data as Urban and Rural, it was decided to evaluate the available version as City (province centre & district centre) versus Village (sub-district centre & village).

In this study, alternative sampling frames are compared for their population coverage and representation for sample selection in internet surveys. The work also aims to provide a methodology for the domain adjustment procedures which also enables to eliminate the sample selection bias on the basis of the corrected information.

### 2.2.4 What to do if Probability Sampling is Impossible in a Web Survey?

Efi Markou, Institut National d'Etudes Démographiques (France)

Nicolas Razafindratsima, Institut National d'Etudes Démographiques (France)

Bernard de Cledat, Institut National d'Etudes Démographiques (France)

Pernelle Issenhuth, Institut National d'Etudes Démographiques (France)

Raphael Laurent, Institut National d'Etudes Démographiques (France)

**Abstract** Quantitative research requires probability sampling: that is a commonly accepted principle. Must we, however, always set aside the idea of a survey without sampling? This paper presents a survey that seeks to address this challenge.

ELVIRE is a survey conducted on 2006-2008 by the French National Demographic Institute, in order to investigate the languages used by the researchers working in the French public research Institutes and Universities. The survey has two parts, each with its own self-administered web questionnaire. The first part (called Unit) aimed at surveying exhaustively the 4000 heads of the public research units; the second part (called Individual) targeted all researchers and PhD students working in these units.

The starting point of the sampling was a non-updated and incomplete database of the research units, including the name of their heads and some other information. Therefore the crucial task was update the units database, necessary for the first part of the survey; and to find the lists of their members, necessary for the sampling of the second part. The first point was successfully completed and 54% of the heads filled out the questionnaire

available on-line between December 2007 and March 2008. The second part, aimed at establishing an exhaustive list or a representative sample of the target population of the second survey, turned out to be impossible.

The paper will describe the strategies implemented to obtain a database or a representative sample of the research units members: requests to Universities and research Institutes to send us lists; requests to units heads to do so or to indicate a small number of persons according to objective criteria, etc.

The different efforts were not successful but they provided us with one third of the member lists even if not always completed and statistically representative. Thus, the Individual part of ELVIRE was conducted as an open, voluntarily jointed, survey based on a positive starting point of two databases: the list of units (with fitting information) and the members lists already collected. Finally, almost 9000 questionnaires were filled out between November and December 2008.

The paper will also present the setting up of this open survey in order to maximize its diffusion; contacting people as different as possible; estimating the distance between our respondents and a representative sample. So, a special effort was made to diffuse the information to all the research units using the heads database. At the same time, we informed all the researchers listed in the second database (almost 35 000 persons) and diffused the information across the Institutes, universities, professional associations, scientific libraries etc. Some questions of the questionnaire were designed to measure the profile and the degree of interest of the respondents. Finally, it will try to compare the respondents profile with some already known external data on the composition of the research community.

## 2.2.5 Self-Selected Samples in Customer Satisfaction Surveys

Giovanna Nicolini, University of Milan (Italy)

Luciana Dalla Valle, University of Milan (Italy)

**Abstract** We have a self-selected sample when the units of the frame population are not chosen by the researcher, but these units choose to participate in the survey process by self-decision. We can find self-selected samples very often in web surveys, but the reasons of self selection are very different. In fact, suppose you have a mailing list from a frame population i.e. the students of a university course and suppose you send the questionnaire to all students, our intention is therefore to plan a census, however only a part of the students replies to the questionnaire. Or suppose you post a questionnaire in a website of a firm and ask the hypothetical customers to fill in the questionnaire. In both cases we will have a self selected sample. However the survey backgrounds are very different. In the first case we know the frame population and we can know auxiliary variables for each unit. On the contrary, in the second case we do not know the frame therefore we cannot have auxiliary information we do not know the population size. Moreover, some web surfers that fill in the questionnaire could not be customers and many customers are not web surfers.

It is well known that the self selected sample is not probabilistic, therefore we would not have estimators according to the design-based approach.

Heckman (1979) proposed a method to estimate the values of the study variable for the unobserved units, in order to overcome the problem of sample selection bias. The method is performed in two steps and requires the use of some auxiliary variables. Then it can be employed when the frame population is known. The contribution of this paper is to use the Heckman method in a customer satisfaction survey where the study variable is not quantitative but categorical.

## 2.2.6 Building fresh annual sample frames for household surveys, based on the new annual French Census

Marc Christine, INSEE (France)

Sébastien Faivre, INSEE (France)

**Abstract** For more than forty years, the sample frames for national household surveys carried out by INSEE (except the Labour Force Survey) have been built from the lists of dwellings established by the Census.

A new rotative Census has been taking place in France since January 2004. Instead of an exhaustive counting of the whole population, it consists of yearly surveys on a part of the territory based on samples of municipalities or addresses (called rotation group, five of them are built over a five years cycle), one of them being covered by the Census each year. This very new technique implies to redefine the sample designs of all household surveys, the Census becoming the first phase of the new sampling design.

One of the most important principles is to use as a frame for the surveys of a current year the lists of dwellings covered by the Census of the previous year. But, since most of the surveys are face to face, the dwellings which

must be surveyed have to be concentrated in some areas and not spread all over the territory, in order to reduce the costs. It implies to build primary units (PU) and to employ a network of interviewers located not far from those PU.

The first difficult work was to build fixed PU, under constraints of size and conditionally to the Census sample of addresses or municipalities. An automatic process of construction of PU has been developed for that purpose.

A major methodological issue was then to establish the sample frame of the PU. A sample of 525 PU was drawn, using the balancing technique developed by Deville and Tille. A methodological work of simulation has led to establish the final PU sample design.

The next step was to check if the annual sample frames formed by the censused parts of the drawn PU are representative of the whole country. As some results tended to show that some specific populations were underestimated or overestimated by the drawn PU, a solution of calibration of the PU was established. It was then checked that this calibration procedure leads to have a more representative Master Sample.

The number of dwellings to be drawn in each PU is then calculated in order to minimize the variations of the final weights of dwellings in the sample (derived from the different steps of sampling), so that final weights for all the dwellings drawn will be as close as possible, under additional constraints of minimum and maximum number of dwellings drawn in each PU. We give here some details about the algorithm and some results on the variation of the final weights. In each drawn PU, dwellings are then drawn with equal probabilities with a systematic random sampling procedure.

## 2.3 New Challenges in Sampling (III)

### Coordinators:

Siegfried Gabler, GESIS - Leibniz Institute for the Social Sciences (Germany)

Seppo Laaksonen, University of Helsinki (Finland)

### 2.3.1 Population Projections and Adjustment Methodologies for Probability Selection in Household Surveys

Öztaş Ayhan, Middle East Technical University (Turkey)

Turgay Ünalán, UNICEF (United States)

**Abstract** The aim of this study is to provide a methodology for the population projections which enables to update information for the future population representation. The techniques used to make population projections can be classified as mathematical models and cohort component projection models. Mathematical models are useful when we only need to project the total population. Therefore, they use total population figures from the past to project future population levels. Among the possible alternative curve fitting methods like linear, geometric, exponential and logistic, the most widely used mathematical model for populations is exponential.

In order to use the exponential model for the population projections, it is essential to obtain and use annual population growth rates. In practice, the recent past population growth rates are determined from the two latest population censuses (if available) for the related domains. For simplicity, we can call these as data source (k) and (k+1). The previous annual population growth rates for the corresponding domains can be obtained by using the population counts of the available data sources. Using the same domain dimensions, the future projected population can be labeled as data source (k+2), for convenience. Generally, the related population domains (or strata) will correspond to the segregated class boundaries in survey sampling methodology.

Information provided for data sources (1) and (2) corresponds to the two previous population census results of the related domains. Therefore, domain totals and overall total information will be identical. On the other hand, information which is based on population projections for data source (3), will not have the same desired properties. In other words, the sum of domains will be equal to the existing domain total, but this will not be equal to the total projection estimate of the overall population. In order to balance the relationship between the domain totals and the overall total, an adjustment is required. This can be called as the total population projection adjustment. This adjustment will correct the difference between the sum of the projected domain populations and total population projections. Total population refers to the household population (members of household), institutional population (armed forces, dormitory, hospital, prison, etc.) and mobile population (homeless, nomadic tribes, etc). For a representative household based sample surveys, the target population considered will be the household population. Therefore, total population has to be adjusted again for the household population. When information is available on the amount of institutional population and mobile population, these amounts has to be subtracted from the adjusted total population in order to obtain the household population for each domain.

Overall sample selection probabilities for households can be determined by obtaining household based information from the population information for each domain. This information can be obtained by dividing the total population of each domain by the corresponding average household size for this domain. The overall selection probability of the domains will be equal to the overall sampling fractions. In this study, overall selection probabilities are compared for projected and un-projected populations. The results have indicated, under representation in the probabilities of selection for unprojected populations. This is naturally considered to be the main source of sample selection bias and consequently the sample estimates will also be biased.

### 2.3.2 Migration Estimation based on the Labour Force Survey: An EU-15 Perspective

Mónica Martí, Universidad de Alicante (Spain)

Carmen Ródenas, Universidad de Alicante (Spain)

**Abstract** It can be observed in the research that the European Union Labour Force Survey (EU LFS) only allows a satisfactory estimation of the stocks of non-nationals or those born abroad in some countries, whereas it proves to be less than adequate in most of them with regard to migration flows. We believe that this very limited success is due to a two-fold statistical problem of imprecision and bias, which is intensified by the embarrassing

question of answer impossible. These difficulties exist among the Member States to a greater or lesser degree, depending on the characteristics of the migratory domain and the particular features that the EU LFS acquires in each country.

### 2.3.3 The “independent” statistical indices as applied to estimation of the quality of sampling

Andrey Veykher, Higher School of Economics (Russia)

**Abstract** The increase of non-response rates during last years limits the adequacy of the rules used for mathematically random sample in the evaluation of real samples. Non-response rate higher than 30% makes the researchers to bring in special control of the structure of non-respondents group. Even the simplest control of the age and gender of the non-respondents shows that representatives of different age-gender groups agree to participate in surveys are different. Reasons for the refusal to participate get from part of non-respondents show that they differ in their living conditions and social practices from the respondents of the same survey. Theoretically we have every reason to suppose that the act of non-response itself is a behavioral manifestation of special social types with specific cultural stereotypes, norms etc. The hypothesis of random non-response, dependant on unpredictable situations, can explain only part of such refusals. Therefore, it is reasonable that respondents grouping into divisions of the indices, which are the aim of the survey, can be different from such distributions of non-respondents, i.e. they are not representative for the general population. Rate of divergence of the values of these indices in sampling and population can be used as a quality assessment instrument for the realization of theoretical sample.

For the practical testing of such hypotheses the author suggest the following method. It was developed and tried out in 2005-2008 during the representative surveys of adult population of St-Petersburg, Russia.

In the independent statistics (governmental or municipal statistics from control accounting and reporting better) we choose the indices, reflecting the elementary events in peoples life, related to the subject of our initial survey. In specific surveys we used:

- 1 studying usage of public transport number of citizens, whove got the social transport pass for the current month from authorities (source Committee for social services of SPb Government),
- 2 studying informal wages formal/legal wages before taxes average in the city during the month prior to the survey date (source The State statistics service)
- 3 studying influence of grandmothers participation in babysitting on the need in day care groups (kindergartens) for children number of children under 3 years old and from 3 to 7 years in the family (source The State statistics service and SPb Government Public Health Committee).

Every questionnaire included relative questions. Based on the responses we calculated the indices, analogues to those received from independent information source. Deviations of calculated indices from independent were interpreted as a measure of difference of the sampling pattern and the population pattern regarding given aspect of the citizens life. Deviation less than 5% was considered insignificant, allowing to use other responses regarding the same aspect of life as representative.

### 2.3.4 Telephone Survey and political behaviour estimates in 22 European countries: Evaluating the need for a dual frame.

Volker Hüfken, University of Düsseldorf (Germany)

**Abstract** Objectives. We sought to determine whether the exclusion of adults without mobile telephone access may bias estimates derived from political behaviour -related telephone surveys.

Methods. We took data from the European Social Survey (ESS) 2006/07 and used logistic regression to compare the odds of political related behaviour for adults with fixed phone access to those for adults with mobile telephone access and those without telephone access.

Results. When interviewed, adults in households did not have telephones it varied from 0 in Sweden to 23,1 % in Russian Federation and the rate of all households have only mobile telephone access varied from 1% in Switzerland to 50 % Finland. Relative to all adults, adults with mobile telephone access had lower odds of vote in the last national election and had lower odds of political interest, watching news and politics in TV and lower odds in reading newspaper about politics or current affairs on average weekday Conclusion. As people substitute mobile telephones for fixed telephones (the percentage is currently for eight countries under 12 % low) which

minimizes the bias resulting from their exclusion from telephone surveys. In twelve countries the percentage is currently more than 20 % high. Their exclusion from telephone surveys could increase the bias. Bias between ten through twenty percentage point is expected for estimates of voting last national election, for about 20 countries.

### 2.3.5 Presentation of WEBSURVNET

Pablo de Pedraza, Universidad de Salamanca (Spain)

Stephanie Steinmetz, Erasmus University (Netherlands)

Kea Tijdens, Erasmus University (Netherlands)

**Abstract** The Social Scientist community needs a multidisciplinary network to study web surveys methodology and its scientific validity.

The network aims to bring together survey methodologist, web surveys experts, social scientist and official survey institutes; synergize the knowledge of survey methodologists and web survey experts; develop guidelines for official bodies and statistical institutes; provide tools to take advance of technological changes; foster the development of new cross-national research proposals and a young researches network.

The network will contribute to the development of theoretical foundations of web survey methodology and innovative ideas. It will be able to offer advise to statistical offices and institutions aiming to equip with Internet for their data collection.

The network will organize academic conferences; establish working groups according to the scientific programme; enhance exchanges among researchers, publish findings and disseminate activities.

## 2.4 Nonresponse

### Coordinator:

Elena von der Lippe, Robert Koch Institute (Germany)

### 2.4.1 Sending advance letters to participants in a Health Telephone Survey as an incentive of reducing non response. A report of a methodological study

Elena von der Lippe, Robert Koch Institute (Germany)

Patrick Schmich, Robert Koch Institute (Germany)

Cornelia Lange, Robert Koch Institute (Germany)

**Abstract** Background The Robert Koch Institute conducts regular telephone health surveys as part of nationwide health monitoring (Kohler und Ziese 2004; Kohler et al. 2005; Ellert et al. 2006). One of the methodological problems in carrying out telephone surveys is the low response rate. Similar problems are facing all the scientists in social research and the topic is highly discussed in the scientific literature. Different experiments are done in testing ways of heightening the response rates in telephone surveys (Singer et al, 2000; Link and Mokdad 2005; Goldstein and Jennings, 2002, etc). Most of the studies report that rising response rates does not automatically lead to improving the sample bias. The general conclusion is that research should focus on reducing nonresponse error and improving data quality rather than simply on raising response rates (Groves and Couper 1998; Lessler and Kalsbeek 1992).

**Objective** We carry out a methodological study on the basis of the ongoing Health Interview Survey GEDA . Our aim is to analyse how the response rate changes and how the sample bias when we send advance letters to the people we want to interview on the telephone. We want to determine what effects have the advance letters on the sample and how the costs change for conducting the interviews.

**Methods** For our purpose, we are going to compare the response willingness of three groups. The first is the group of people for which the addresses are not available. Our second group is the so-called control group it is the group of people for which we have the addresses, but they are do not receive an advance letter. And the third group is with available addresses and receiving advance letters. The comparison of these groups will show us the effect of sending an advance letter over the nonresponse rates and will also show us how the groups differ demographically and how this affects our sample design.

Our whole sample size is planned to consist of 4000 telephone numbers. Our expectations are that for about 2500 2700 households will be possible to recruit the addresses. Half of them (randomly selected) will receive a prenotification letter and the other half will be directly asked to participate in the telephone interview

**Expected results** We suppose that the overall response rate will be significantly higher when an advance letter is sent. We expect that the study gives answers to the questions if the data quality is improving and how the sample changes. Other studies report sample bias towards highly educated and older people (for the case of US, Link and Mokdad, 2005), therefore we are also anticipating similar bias for our study. However, conclusions we can not make before our scientific experiment has taken place. The field work will be done in March and in April-May the results will be scrutinized.

### References:

- Ellert, U., Wirz, J., & Ziese, T. (2006). Telefonischer Gesundheitssurvey des Robert Koch-Instituts (2. Welle). Deskriptiver Ergebnisbericht. Berlin: Robert Koch-Institut.
- Goldstein, K. M., & Jennings, M. K. (2002). The effect of advance letters on cooperation in a list sample telephone survey. *Public Opinion Quarterly*, 66, 608-617.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Kohler, M., Rieck, A., Borch, S., & Ziese, T. (2005). Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts - Methodische Beiträge. Berlin: Robert Koch-Institut.
- Kohler, M., & Ziese, T. (2004). Telefonischer Gesundheitssurvey des Robert Koch-Instituts zu chronischen Krankheiten und ihren Bedingungen. Deskriptiver Ergebnisbericht. Berlin: Robert Koch-Institut.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York: Wiley.
- Link, M. W., & Mokdad, A. (2005). Advance Letters as a mean of improving respondent cooperation in random digit dial studies. *Public Opinion Quarterly*, 69(4), 572-587.



Singer, E., van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, 64, 171-188.

### 2.4.2 Health Interview Surveys in practice: do we miss the sick?

Stefaan Demarest, Scientific Institute of Public Health (Belgium)

Jean Tafforeau, Scientific Institute of Public Health (Belgium)

Johan Van der Heyden, Scientific Institute of Public Health (Belgium)

Lydia Gisle, Scientific Institute of Public Health (Belgium)

Sabine Driekens, Scientific Institute of Public Health (Belgium)

**Abstract** Background Unit non response is a major threat for the external validity of survey results. Especially in health interview surveys it is important to know to which extent the participation rate is associated with the health status.

**Methods** Data were obtained for all households that were selected to participate in the first wave of the Belgian Health Interview 2001 (N = 6,105). The participation status of these households was linked with health data from the households reference person derived from the National Census 2001 in which a limited number of health related questions was included. While participation to the Health Survey was voluntarily (response rate: 48.5%), it was compulsory for the Census (response rate: 96%). The participation status for the Health Survey was assessed in terms of refusal rate and the percentage of households that were not contactable. The health status, derived from Census data, was estimated by measuring the perceived health, the presence of longstanding health problems and the possible limitations due to health problems. Other covariates used in the analysis were age group, gender and educational level. Different logistic regression models were fitted with, respectively, participation versus refusal and participation versus non contactable as the outcome variables.

**Results** For all reference persons selected to participate in the Health Survey, Census data could be traced (corresponding first name, date of birth and gender). The results show that only the presence of longstanding health problems positively affects the refusal rate (OR: 1.23; 95% CI: 1.09 1.39) but does not affect the probability to be able to contact the household. No significant association could be established between perceived health nor limitations due to health problems and the participation status. Given the relatively high levels of item non response in the some Census questions (longstanding health problems: 14.1%, limitations due to health problems: 15.8%), a sensitivity analysis was performed with alternating imputed values for item non response. Yet this analysis did not affect the main findings.

**Conclusion** Having a longstanding health problem is a barrier to participate in health surveys and possibly biases the survey results. Specific efforts in terms of information, mode of data collection, mode of contact, length of the questionnaire should be made to improve the willingness to participate for this specific group.

### 2.4.3 Getting value for money in longitudinal study design: larger samples vs. higher retention

Margaret Ely, University of Otago, Dunedin (New Zealand)

Richie Poulton, University of Otago, Dunedin (New Zealand)

**Abstract** Background The Dunedin Cohort study is world renowned for its high retention rate. Information on difficulty of contact is used to assess the likely impact of attrition on the estimation of the prevalence of risk factors and associations of these risk factors with poor health outcomes.

**Methods** 972 (96% of living) Dunedin Study members who presented at phase 32 were scored for difficulty of contact. Bias in the prevalence of risk factors and associations with poor health outcomes when those who were most difficult to contact (hard-to-get) are excluded (quasi attrition) are assessed by simulation (using R). Where associations are not biased, we estimate the increment in sample size which would be needed without the hard-to-get, in order to achieve the same efficiency as the sample with the hard-to-get. The relative value (RV) of a study member who is hard-to-get is defined as the ratio of the additional sample size needed to the number of the hard-to-get recruited.

**Results** Work in progress has demonstrated that Dunedin Study members who were hard-to-get were more likely than the easy-to-get group to have higher rates of substance use and dependence, and to have behavioural disorders. Hence the prevalence of risk factors for poor health for the sample without the hard-to-get were negatively biased. Associations between exposures and poor health outcomes may not be biased, for example

that between cannabis consumption and periodontal disease. However, excluding those who are hard-to-get, a greater than expected loss of efficiency is observed ( $RV > 1$ ). Generalised results for the dependence of the RV on prevalence of exposure amongst the hard-to-get are derived.

Conclusions Because those who are hard-to-get tend to be at the extremes of distributions of risk factors for poor health outcomes, or to have greater prevalence of conditions that affect only small proportions of the population, they contribute more information about associations between such predictors and conditions than do those who generally participate in studies. Efforts in retention can therefore represent greater value for money than increases in sample size.

#### 2.4.4 Moving up: Does an increase in response mean different results? First evidence from an experimental study in the German General Social Survey 2008

Michael Blohm, GESIS - Leibniz Institute for the Social Sciences (Germany)

Achim Koch, GESIS - Leibniz Institute for the Social Sciences (Germany)

Hanna Kaspar, University of Mainz (Germany)

**Abstract** In the past decades response rates have been declining for a large variety of surveys in a lot of countries all over the world (deLeeuw & deHeer 2001). Also the German General Social Survey (ALLBUS) has been facing an increase in nonresponse in the last years. The ALLBUS is a face-to-face survey of the adult population in Germany, which has been fielded every second year since 1980. Between 1996 and 2006, e.g., the response rate of ALLBUS has decreased from an already rather low level of 54 % in 1996 to 43 % in 2006. The basic question is whether this increase in nonresponse does also mean a reduction in data quality, in particular a higher nonresponse bias. Recent research has questioned, whether empirically a close relationship between the nonresponse rate and the size of the nonresponse bias can be observed (e.g. Groves 2006).

In order to find out, whether an increase in response rates in ALLBUS would also mean different (and probably better, i.e. less biased) results, an experiment was implemented in 2008. The general idea was to compare the results of the regular ALLBUS 2008 (with an expected response rate of 45%) with the results of an additional sample (called ALLBUS +) in which intensive fieldwork efforts would (hopefully) result in a considerably higher response rate (targeted response rate of 70%). Amongst other things, the additional efforts to achieve a higher response rate included the deployment of a special group of interviewers, the personal briefing of interviewers, better payment of interviewers, and the use of respondent incentives. To decrease the respondent burden, also the interview length was reduced in the experimental sample from approximately 70 minutes to approximately 25 minutes on average. Both samples were fielded concurrently between March 2008 and July 2008.

In our presentation we will describe the design and implementation of the study and present first results. We will provide an overview on the comparison of distributions between the regular ALLBUS and the ALLBUS + sample, distinguishing between different types of variables. In addition, the stability or change of results over the fieldwork period will be analysed. In doing this, the emergence of differences in results between ALLBUS and ALLBUS + can be highlighted.

#### 2.4.5 Follow Up Survey – What Is It Good for? Implications from an empirical study

Frode Berglund, Statistics Norway (Norway)

Øyvind Kleven, Statistics Norway (Norway)

Kristen Ringdal, Norwegian University of Science and Technology (NTNU) (Norway)

**Abstract** Non response rates in surveys are increasing in most countries. As non response is a potential source of bias, it is a serious potential threat to the quality of survey data. In surveys such as the European Social Survey, substantial resources are used to keep up the response rates by various means such as incentives to respondents and interviewers and refusal conversion. Are these efforts worthwhile, do they actually reduce bias? This is the point of departure for this paper based on round 3 of the ESS and a follow-up survey of non respondents and respondents.

A well known, but not so frequently used, method to explore and reduce non response error is to perform a follow-up survey. In ESS R3, a short self-completion questionnaire was distributed to a subsample of the original Norwegian ESS sample after the main survey. The questionnaire contained a few key background and interest

variables, and was distributed to all non respondents and a random sample of respondents from the main survey. Of course, the follow up survey also suffered from non response the response rate was only 30 per cent among non respondents, while corresponding figures among respondents was 60 per cent.

The purpose of our paper is to compare estimates on selected variables from the response group and the non respondent group in order to elaborate on non response errors in the main survey. On the background of our empirical results, we will discuss the value of follow up surveys.

## 2.5 Selection Bias in Panel Research (I)

### Coordinator:

Gerty Lensvelt-Mulders, University for Humanistics (Netherlands)

### 2.5.1 Selection bias in Internet panels: challenge or dead blow?

Gerty Lensvelt-Mulders, University for Humanistics (Netherlands)

**Abstract** Currently it is not possible to draw random samples from the general Internet population because there is no high quality sampling frame available. As a result most access panels consist of volunteer-opt-in samples. This can seriously hamper the results of surveys based on these samples. When under coverage of certain groups results in non-random missingness (non-MAR) the validity of the results could be seriously corrupted and selection bias will become apparent. It is not yet known, how to deal with selection bias in non-random samples, but since access panels are here to stay and are used more and more in research for policies and practices scientists have the obligation to study the phenomenon and search for solutions that will prevent against or correct for selection bias in access panels. Solutions can be methodological, i.e. aiming at the prevention of selection bias or statistical, i.e. aiming at adjustment techniques like weighting. This presentation will give a short overview of current streams of thinking on this problem, as an introduction to the presenters of the session on selection bias in panel research.

### 2.5.2 The Civic Duty Weight Factor

Rene Bekkers, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** This paper introduces the civic duty weight factor. The weight factor is the result of a procedure designed to correct self-selection of individuals with a higher sense of civic duty into survey respondent pools. The procedure weighs observations from survey respondents based on self-reported civic behaviors for which population values are known. Examples of such behaviors are voting, filing tax forms, blood donation, and registration with the authorities. Illustrations of the civic duty weight factor are given using data from various surveys on charitable giving and volunteering in the Netherlands. Unweighted estimates are found to be higher than estimates weighted by the civic duty weight factor. The factor should be used in addition to ordinary weight factors. It is concluded that the factor is a useful tool in survey research aiming to estimate the prevalence of norm-guided behaviors for which population values are unknown.

### 2.5.3 Comparing different weighting procedures for volunteer online panels - Lessons to be learned from German, Dutch and Spanish Wage indicator data

Stephanie Steinmetz, Erasmus University (Netherlands)

Kea Tijdens, Erasmus University (Netherlands)

Pablo de Pedraza, Universidad de Salamanca (Spain)

**Abstract** The strengths and weaknesses of web surveys have been widely described in the literature. Of particular interest is the question of the quality and reliability of web surveys for scientific use - more concretely, to which degree can the obtained results be generalised for the whole population? As respondents are not selected at random and the target population rather forms a convenience than a probability sample, particularly volunteer web surveys are subject to selection bias.

To deal with this problem, weighting adjustment, like post-stratification and propensity score weighing, have been seen as a possible solution to reduce the biases in web surveys. However, particularly the post-stratification weighting, aiming to adjust for demographic differences between the sample and the population under consideration, seems to be limited. As some variables of interest often do not show a sufficiently strong relationship with the demographic weighting variables, this method can correct for proportionality but not necessarily for representativeness. As a consequence, another weighting technique called Propensity Score Adjustment (PSA) has been suggested as an alternative for statistically surmounting inherent problems in web survey data. It aims to correct for differences caused by the varying inclinations of individuals to participate in web surveys. It adjusts

for selection bias due to observed covariates which are demographic as well as webographic(lifestyle) variables. These variables measure general attitudes or behaviour that are hypothesised to differ between the online and the general population. In the scientific community, however, this method has traditionally not been applied in the field of surveys, and there has been a minimal amount of evidence for its applicability and performance, and the implications are not conclusive. Moreover, the statistical theory behind this approach is not well developed and the effectiveness and implications, particularly for survey methodology, still need to be better studied

Against this background, the present paper attempts to explore various statistical weighting procedures for volunteer web surveys and evaluate their effectiveness in adjusting biases arising from non-randomised sample selection. In order to achieve the goal, three methods are compared in more detail: post-stratification, correlations and finally PSA. A first essential step for exploring the existing selection bias within the existing data-sets will be a detailed bias description. It is particularly needed for the application of PSA searching for the variables that are going to be included in the calculation of the Propensity Score weights. The efficiency of different weights will then be tested by comparing unweighted and weighted results from the German, Dutch and Spanish sample of the Wage-indicator Survey 2006 with those that could be found using data from the German Socio-economic panel, the OSA Labour Supply Panel and the Spanish Structure Earnings Survey for the same year. In the framework of these examinations, analytical graphics and formal tests of significance will be used. Furthermore, the sensitivity of the results, particularly to changes in the specification of the propensity score, will also be addressed.

## 2.5.4 Attrition of French electoral panels, 1958-2007: a “political mood” explanation?

Viviane Le Hay, Sciences Po (France)

**Abstract** Attrition is a regular challenge and feature of panel survey. Belong the inference problem attrition raises substantial question about the evolutions of ordinary citizens regarding electoral choice. The aim of the study is to analyze the process of attrition of panel data, by comparing four French electoral surveys between 1958 and 2007 (conducted by phone and face to face over national elections in France):

- the electoral panel of 1958: referendum and parliamentary elections (Association française de science politique, 1960);
- the French National Election Panel Study, 1967-1969: parliamentary election, referendum, presidential election (Converse, Pierce, 1986);
- The Panel electoral français (2002) CEVIPOF-Ministère de l'Intérieur : presidential and parliamentary elections (Cautres, Mayer, 2004) ;
- The Panel electoral français (2007) CEVIPOF-Ministère de l'Intérieur : presidential and parliamentary elections (Cautres, Muxel, 2009).

This comparison is unique and helps to assess the impact of elections types and survey design on panel attrition, but also the impact of the transformation of the general public in term of level of education and generational renewal. The comparison of these four surveys shows two important results which underlines the dialectics of changes in permanence:

- 1 The permanence of the explanation by socio-political characteristics between 1958 and 2007 in terms of political interest, age, level of education. Our results demonstrate that the usual suspect of political awareness explain very well those who refuse to integrate the panel (the Refusers). But we also discover that level of political sophistication does not play a role for the Leavers (i.e. the respondents quitting the panel).
- 2 The scope of attrition varies a lot between surveys: for example, Leavers are much more numerous in 2002 than in 2007. The political climate of the surveys seems a strong hypothesis: invariably since 1958, a current favorable political mood is accompanied by a lower attrition of panel surveys.

Our comparison draws major results and open new solutions for lessening attrition. Along with context and survey design effects, the 2007 survey allows a assessment on the interview effect on recruitment of panelist.

## 2.6 Selection Bias in Panel Research (II)

### Coordinator:

Gerty Lensvelt-Mulders, University for Humanistics (Netherlands)

### 2.6.1 The Internet sample

Elzbieta Getka-Wilczynska, Warsaw School of Economics (Poland)

**Abstract** In this presentation Poisson processes and basic methods of the reliability theory are proposed to interpretation, definition and analysis some stochastic properties of process of Internet data collection. At first, the notion of uncontrolled sample is introduced and random size of it is defined as a counting process. At the second, the process of Internet data collection is considered as a life test of the population surveyed. The events which appear in Internet survey are interpreted as a lifetime, arrival, death of the element of the population and the basic characteristics of reliability of the length of the population lifetime are described, calculated and estimated by using the notions and methods of the reliability theory.

### 2.6.2 How to cover the general public by Internet interviewing

Marcel Das, CentERdata (Netherlands)

**Abstract** The world of survey research is rapidly changing, moving from face-to-face (CAPI) and telephone interviews (CATI) to Internet interviewing. Several methods are used to attempt to survey the general public by Internet, including the use of volunteer and RDD samples. However, these methods may not fulfil the demands on coverage, sampling and response posed by scientific researchers. People in volunteer samples are often a rather selective part of the general public. In addition, because of a vast increase of mobile-only households and of declining response rates, RDD samples do not cover the population anymore.

A newly established Internet panel in the Netherlands uses a different design. A true probability sample of households drawn from a population register by Statistics Netherlands is contacted with a telephone or in-person interview, asking respondents to join the panel. The panel provides a computer and internet connection to those households that cannot otherwise participate (about 85% of the households have Internet access).

In the presentation an overview of this new panel concept is given. An experiment was carried out to determine the optimal recruitment strategy for the panel. The factors to be optimised were: contact mode, incentive amount, timing of the incentive, content of the information letter, and timing of the panel participation request. The experimental design took into account the naturally varying factor of whether or not the households involved had a known fixed landline. The highest response rate was found with an incentive of 10 euro. The incentives of 20 euro and 50 euro did not substantially increase response rates beyond those seen at the 10 euro level. All incentives were found to have much stronger effects on response rates when they were enclosed in the announcement letter than when they were paid later. Another important factor was the contact mode that was used. Contact was made either by CATI or CAPI. While the contact rate was somewhat higher with CATI, the CAPI interviewers were more successful in obtaining panel participation. The contact rate was substantially lower in the subpopulation of households without a known fixed landline, even when controlling for the effect of contact mode. The response rates were not affected by the content of the information letter. Response rates were also independent of whether the nature of the panel study was explained before or after the recruitment-interview was completed.

The panel design is evaluated by comparing the composition of the recruited panel to population statistics, to traditional face-to-face scientific studies and to commercial access panels.

### 2.6.3 Motivating different groups: questionnaire topic and participation rates

Miquelle Marchand, CentERdata (Netherlands)

**Abstract** There are many challenges to data collection in survey interviews. In contacting respondents and gaining cooperation we have the challenge of (unit) non response; how can respondents best be approached to convince them to cooperate and how can we keep respondents motivated? To try to answer this latter question, our research is conducted in the CentERdata LISS-panel, a panel consisting of about 5000 households representative of the Dutch speaking population. The design of the LISS-panel combines a probability sample

and traditional recruitment procedure with online interviewing. A specialty of this panel is that people without Internet access are provided with the necessary equipment so that they are also able to participate in the panel. Respondents receive questionnaires each month, this way providing data on many different topics.

More specifically, we will look at reasons given for not wanting to join the panel (we will both look at reasons falling in predefined categories such as no time or not interested as well as code other reasons given, i.e., open answers). For those respondents who did join the panel we will investigate whether questionnaire topic influences participation rate: both taking into account demographics and values people may hold. Does questionnaire topic motivate certain groups of people to participate (e.g., elderly, immigrants, or people with certain values)? In addition, we will look whether some groups evaluate different questionnaires (i.e., topics) differently when they answered the survey. Evaluation consists of questions about liking and difficulty of the questionnaires.

## 2.7 Strategies for Nonresponse Adjustments (I)

### Coordinator:

Vasja Vehovar, University of Ljubljana (Slovenia)

### 2.7.1 Imputation of annual income in panel surveys with partially non-responding households

Joachim R. Frick, DIW Berlin (Germany)

Markus M. Grabka, DIW Berlin (Germany)

Olaf Groh-Samberg, German Institute for Economic Research (DIW) (Germany)

**Abstract** Household panel surveys by definition and for good reason collect information from all members of a survey household and trace these individuals over time. Typically, all adult household members are interviewed personally (with the only exception of the US Panel Study of Income Dynamics (PSID), where the household head is asked for proxy information). This design, however, creates the risk of a certain type of non-response if one member (or more) of an otherwise interviewed household rejects to give a personal interview (partial unit non-response, PUNR).

In most household panel surveys around the world, an increasing share of households is affected by PUNR. This leads to severe problems for any kind of analysis that is concerned with the aggregation of resources across household members or with the interaction of individuals and their choices and behavior within the household context (e.g. labour supply decisions). The relevance of these phenomena is intensified by the fact that partial unit-non-response is most likely not at random. In this paper, we focus on the problems of income analysis and poverty.

The standard approach in inequality research assumes that all household members pool and share all available resources (i.e., income). Applying an equivalence scale to control for differences in household size and composition yields a welfare measure which is the same for all persons living in that household. However, in cases of partial non-responding households, the aggregation of income falls just short of at least one individual's income, thus understating the full household income.

Using data on 24 waves of the German Socio-Economic Panel Study (SOEP) we evaluate four different strategies to deal with this phenomenon. (a) Ignoring the fact that a household member and its income information is missing, thus assuming the missing individual's income to be Zero. (b) Adjusting the equivalence scale by ignoring the person's contribution to household income as well as household needs, thus assuming the person does not exist at all. (c) Eliminating all households observed with PUNR from the analysis population, thus assuming that these households are missing at random. (d) Imputing the missing income components taking advantage of the longitudinal information, if at all available, and cross-sectional regressions otherwise.

Results on income inequality and even more so on relative poverty may crucially depend on the choice of the aforementioned option, because this decision will affect the income of all individuals living in households which happen to be concerned by PUNR, as well as the relative poverty line to be derived from the national mean or median. In our paper we (a) describe incidence and time trends of PUNR, (b) analyse the selectivity of PUNR, (c) carry out a three-stage longitudinal imputation of missing income components due to PUNR at the individual level, and (d) provide sensitivity analyses showing the variation in the results for income inequality and poverty using alternative imputation strategies.

### 2.7.2 Estimation of population parameters under non-independent nonresponse models

Wojciech Gamrot, Karol Adamiecki University of Economics (Poland)

**Abstract** The phenomenon of missing data in sample surveys is often modelled as a random event under the assumption that all respondents are characterised by some unknown response probability. Another popular assumption states that response events are independent and hence the decision of a particular sampled unit to participate in the survey (or not) does not impact decisions of other units. However, there may be several practical reasons to question this assumption. This paper focuses on the situations where it is not satisfied. The impact of potential dependencies between behavior of population units on estimator properties are investigated. Simulation results are presented.



### 2.7.3 The influence of selective nonresponse in the analysis of levels of annoyance and sleep disturbance due to aircraft noise exposure

Oscar Breugelmans, National Institute of Public Health and the Environment (Netherlands)

**Abstract** With dwindling participation rates, selective unit nonresponse has become a major issue in the analyses of mail, telephone and web surveys. This is especially true when the main research question of the survey is closely related to situations that can have a direct influence on the well being of the population.

The GES study is an ongoing monitoring program that studies the impact of Amsterdam Schiphol Airport on the health and well being of the population in the surrounding areas. Over the last decades the situation around the airport has changed considerably. While the number of aircraft has increased dramatically, the noise emitted by each aircraft has decreased due to the phasing out of noisier aircraft. The growth of the airport has led to the installation of a new runway in 2002. Over the years, the growth of the airport has sparked strong opposition from residents and the airport is almost constantly on the policy agenda. In this volatile situation, three postal questionnaire surveys have been carried out since 1996 to assess, among other topics, annoyance and sleep disturbance due to aircraft noise exposure. With participation rates of approximately 50% there was ample possibility for selective nonresponse bias. To assess this influence, a subsample of nonresponders was contacted by telephone or visit to the house to answer a subset of critical questions from the main questionnaire. In 2002 we were able to reach 83% of the nonresponders in the subsample. The paper will discuss the influence of selective unit nonresponse on the different study outcomes and the methods that were used to correct for this influence.

### 2.7.4 Individual Follow-up of the Target Population: the Plural Strategies of a Web Survey

Efi Markou, Institut National d'Etudes Démographiques (France)

Bernard de Cledat, Institut National d'Etudes Démographiques (France)

Nicolas Razafindratsima, Institut National d'Etudes Démographiques (France)

Raphael Laurent, Institut National d'Etudes Démographiques (France)

Pernelle Issenhuth, Institut National d'Etudes Démographiques (France)

**Abstract** ELVIRE is a survey in two parts, each with its own self-administered web questionnaire. Conducted by the French National Demographic Institute, it aimed to investigate the languages used by the researchers working in the French public research Institutes and Universities. The first part of the survey (called Unit) was sent to about 4000 heads of all the public research units. The questionnaire was available on-line for four months (December 2007 to March 2008). The second part (called Individual) was an open survey (without previous sampling) addressed to all researchers and PhD students working in these units. The questionnaire was available between November and December 2008.

The purpose of this paper is to describe the strategies implemented in order to reduce non-response to the Unit part and to measure the impact of non-response bias. As a self-administered web survey, a high non-response rate was expected, all the more because of the duration of the questionnaire (more than 1 hour on average) and the respondents' limited availability (overworked unit heads). However, 54% of the heads filled out the questionnaire. Among other strategies described by the paper, a special effort was made to contact the heads individually several times, alternating a variety of means. After sending one letter and if no answer or no completed questionnaire two emails to each of them, a follow-up phone call was made one month later. The interviewers had to contact the potential respondents and persuade them to fill out the questionnaire or else obtain a reason for their refusal. All information was coded in a data file. This telephone contact not only increased the response rate by 24 percent points (from 30 to 54%) but also offered a source of information on respondents reactions and the reasons for their refusal. To better understand these reactions, two other sources, more qualitative in nature, will be mobilized. First, the e-mails exchanged between the survey team and the respondents. Second, the field of the questionnaire especially dedicated to respondents comments on the quality of the questionnaire.

Furthermore, the comparison with the Individual part (almost 9000 responses) may give additional clues about the impact of some factors, such as the questionnaire design (more functional and attractive in the second part); its length (20 min instead of 1 hour); or the respondents' status.

## 2.8 Strategies for Nonresponse Adjustments (II)

### Coordinator:

Vasja Vehovar, University of Ljubljana (Slovenia)

### 2.8.1 Influence of nonresponse in business tendency surveys on properties of expectations

Barbara Kowalczyk, Warsaw School of Economics (Poland)

Emilia Tomczyk, Warsaw School of Economics (Poland)

**Abstract** Some of the most valuable surveys conducted in order to obtain information about economic situation are business surveys. In qualitative business surveys, also known as tendency surveys, respondents are asked to specify direction of changes recently observed and expected in the future. Survey questions typically focus on such economic variables as general business conditions, prices, unemployment, and production; in most cases, answers are limited to options increase/improvement, no change, decrease/deterioration. Qualitative business surveys, although do not provide as many possibilities of formal analysis as quantitative business surveys, are often regarded as more reliable because respondents are more likely to correctly identify direction of changes than to provide a precise point assessment.

Expectations concerning key economic variables constitute a particularly interesting field of analysis as they certainly influence decisions undertaken by managers among them selection of investment and production levels. Since the assumption of rationality of economic agents forms the basis of the standard neoclassical economic theory, question of whether expectations of enterprises are indeed formed rationally deserves careful attention.

As in almost every survey, in qualitative business tendency surveys the problem of nonresponse appears. It has been observed that the nonresponse rates are much higher for questions about expectations than analogous ones about realizations. In light of these facts, we propose to analyse selected properties of expectations expressed in business tendency surveys taking into account various nonresponse mechanisms.

The empirical part of the paper is based on business tendency surveys conducted by the Research Institute for Economic Development (RIED) of the Warsaw School of Economics.

### 2.8.2 Characteristics of "No Shows" in The Netherland Study of Depression and Anxiety (NESDA)

Johannes H. Smit, Vrije Universiteit Amsterdam (Netherlands)

Stasja Draisma, Vrije Universiteit Amsterdam (Netherlands)

T. Zandbelt, Vrije Universiteit Amsterdam (Netherlands)

W.A. Nolen, University of Groningen (Netherlands)

E. de Beurs, University of Leiden (Netherlands)

**Abstract** The present study investigates characteristics of those who appeared versus those who did not appear for a baseline assessment in the Netherland Study of Depression and Anxiety. In NESDA 2981 persons are included in a 8 year follow up study on the course of anxiety and depression disorders. Within the NESDA screening data are collected on demographics, physical and mental conditions of potential participants. At the end of the screening it is registered whether the person meets the inclusion criteria and is willing to participate in NESDA. Within the inclusion procedures for the Nesda study about 30% of the persons who passed the screening and initially agreed to participate in the baseline assessment did not show up for the interview. Because of the large numbers of persons screened the NESDA study has enough statistical power to investigate differences between characteristics of shows and no shows.

We will use data from the inclusion phase basic demographics as sex, age, current employment status, and education and data on physical and mental functioning -collected with questions from the CIDI interview- will act as determinants. We will answer the following two questions.

- 1 To describe the characteristics of persons who initially agreed to participate in the NESDA study but failed to show up at the actual research assessment appointment.
- 2 To compare these characteristics with the persons who initially agreed to participate and did show up and thus investigate which variables are associated with not showing up in spite of initial consent to participate.

### 2.8.3 Declining selectivity of foreigners, immigrants and persons with migration background by using questionnaires in their first language

Michael Ruland, Institute for Employment Research (Germany)

Britta Matthes, Institute for Employment Research (Germany)

**Abstract** The readiness to participate in voluntary interviews is selective. Particularly foreigners, immigrants and persons with migration background take part in interviews less frequently. At first it is necessary to analyse the effect of well-known procedures for the diminution of participation failures in order to be able to develop specific strategies to increase the readiness to participate of these persons. More than 10.000 persons of the birth cohorts 1956-1988 were interviewed on the basis of a registration office sample with the help of computer-assisted telephonic interviews in the context of the ALWA study. To be able to formulate as representative statements as possible on the base of this sample, special values concerning the avoidance of failure quotas regarding foreigners, immigrants and persons with migration background were ascribed. To avoid increased failure quotas with non-German-speaking persons foreign-language interviews conducted by foreign-language interviewers were additionally offered. Furthermore was tried to recognize a bias by non-attendance. If no phone numbers of selected people were found, a short questionnaire would be sent to them by mail.

In our presentation we explain which of the strategies used in the ALWA study have influenced the foreigners, immigrants and persons with migration background readiness to participate the most. To what extent can we increase the readiness of these persons to participate through interviews held in their first languages at voluntary elevations? Is it possible to reduce the selectivity of the willingness to participate with this strategy? Is it possible to formulate general statements concerning selectivity by using non-response questionnaires regarding the non-attendance of foreigners, immigrants and persons with migration background?

### 2.8.4 When Normality Fails: a Multiple Imputation Technique Based on the Cumulative Distribution

Bryce Weaver, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

**Abstract** Objectives. One of the drawbacks of multiple imputation is that assumptions are generally made on the form of the distribution e.g. normality. These assumptions may often be invalid. In our work, we develop a technique of multiple imputation that makes no a priori assumptions on the form of the distribution. This tool is useful for the imputation of data under the standard assumption that the values are missing at random (being missing may depend on other variables, but not on the value of the variable to be imputed). It is designed to reproduce the cumulative distribution of any variable continuous or dichotomous. The conceptual framework is transformed into a simple tool in SAS. Empirically, we use the Swiss Household Panel (SHP) wave 8 (2007) to apply this new method and compare the obtained cumulative distribution with the ones from the standard PROC MI procedures in SAS.

Method. The entire sample (missing and non-missing) is separated into imputation classes based on up to five categorical variables (there can be grouping of classes in order to maintain a sufficient number of non-missing observations). These variables are selected by the researcher as being important indicators of the values of the variable(s) to be imputed. First, within each imputation class, a cumulative distribution for the variable to be imputed is built from the non-missing cases. This cumulative distribution is obtained by linearly interpolating between the observed points (non-missing) to allow for any form of the distribution. Then, for each missing observation, the imputed value is obtained by randomly selecting a number between zero and one from a uniform distribution and by taking the value that corresponds to the number selected on the cumulative distribution. This last procedure is repeated multiple times. The SAS program is flexible as it allows the user to choose (i) the variables to be imputed, (ii) the categorical variables used to create the imputation classes, (iii) the number of imputed files to be produced, and (iv) whether the variable(s) to be imputed are continuous or discrete variables.

Results. Using the SHP wave 8 (treated as cross sectional), which has 5794 observations, income is assigned a missing value for nearly one third of non-missing cases ( $n=1815$ ). These are then imputed using this procedure and some PROC MI procedures. The obtained distributions are then compared to the actual distribution of the observations that had been assigned missing. The above technique performs better at reproducing the actual form of the distribution on missing data than the PROC MI procedure.

Discussion. Some of the advantages of the technique are that it works for any type of missing categorical or continuous, it can reproduce any form of a distribution, and it provides relatively accurate standard descriptive statistics. Future work could consist in developing some more theoretical basis to produce smoother distributions as opposed to the non-smooth distribution based on linear interpolation.

## 2.9 Understanding nonresponse and attrition: Research from the UK Survey Design and Measurement Initiative

### Coordinator:

Patrick Sturgis, University of Surrey (United Kingdom)

### 2.9.1 Results from an experiment to prevent refusals in a continuing longitudinal study

Sosthenes Ketende, Centre for Longitudinal Studies, Institute of Education, University of London (United Kingdom)

Lisa Calderwood, Centre for Longitudinal Studies, Institute of Education, University of London (United Kingdom)

Ian Plewis, Social Statistics, University of Manchester (United Kingdom)

Rebecca Taylor, National Centre for Social Research - NatCen (United Kingdom)

**Abstract** Strategies for converting refusals into productive cases have received some attention in a cross-sectional context but the literature on converting refusals in ongoing longitudinal studies is sparse. Refusers at wave  $t$  ( $t > 1$ ) will have participated at wave one and so the reasons for refusing can be hypothesised to be circumstantial rather than being based on a deep-seated antipathy to social investigations. In this paper, we report results from a randomised experiment that tests the efficacy of two different kinds of intervention: (i) a letter to initial refusers at wave  $t$  addressing reasons for refusing based on data from earlier waves, and (ii) reissuing all refusals to a different interviewer. These treatments were crossed in a  $2 \times 2$  design with standard fieldwork procedure as the control group. The intervention was applied in wave four of the Millennium Cohort Study, the fourth in the renowned series of UK birth cohort studies.

### 2.9.2 Solving the problem of attrition in longitudinal surveys

John Bynner, Longview (United Kingdom)

Harvey Goldstein, University of Bristol (United Kingdom)

Gabriele Durrant, University of Southampton (United Kingdom)

**Abstract** The paper reports work conducted in the first stage of a project directed at reducing attrition in longitudinal surveys. The first stage comprises an investigation of the correlates of non-response in longitudinal surveys through statistical modelling of attrition processes using data relating to attrition in established UK longitudinal surveys: British Household Panel Study (BHPS) 1958 and 1970 British birth cohort studies, Family and Childrens Survey (FACS). The second stage is to formulate, in the light of the results of the modelling, optimum strategy for encouraging continued participation as a basis for improved field strategy to reduce attrition. The third stage is to design and implement an exemplar field experiment to test the new strategy using a survey from a series of repeated cross-sectional ('Omnibus') surveys to supply the first wave of a longitudinal survey in the second wave of which respondents are allocated at random to an experimental and control condition. The results of the secondary analysis identify amongst other factors interviewer attributes and experience as factors in attrition and provide the basis for the manipulation of standard interviewer allocation procedures in the Omnibus experiment

### 2.9.3 An Investigation of Interviewer Effects on Household Nonresponse Using a Multilevel Modelling Approach

Gabriele Durrant, University of Southampton (United Kingdom)

Robert M. Groves, University of Michigan (United States)

Laura Staetsky, University of Southampton (United Kingdom)

Fiona Steele, University of Bristol (United Kingdom)

**Abstract** Nonresponse is a major problem facing research in the social sciences. Response rates in many surveys have been falling in many countries. To improve response rates and to decrease non-response bias, it is necessary to better understand the nonresponse process and why and how sample members reply to a survey. In face-to-face surveys the interviewer plays a vital role in contacting sample members and achieving their cooperation.

This paper investigates the influence of interviewers on unit-nonresponse across a number of UK government surveys. The analysis focuses on the effects of socio-demographic characteristics of the interviewer, interviewer experience, interviewing strategies and interviewer behaviours and attitudes. Of particular interest is the interaction between household and interviewer characteristics. Since several surveys are combined in one study, survey specific and survey independent interviewer effects will also be investigated.

The study makes use of a unique dataset, which links the survey outcome of several surveys to a rich set of auxiliary variables for both responding and nonresponding households, including socio-demographic information from the 2001 UK census, detailed information about the interviewer and information about the geographical area. The data available have a hierarchical structure with households nested within interviewers and interviewers being cross-classified with geographical areas. The paper explores the use of multilevel cross-classified models to analyse the effects of interviewers taking account of household level and area level characteristics. Implications for survey practice for the improvement of interviewer training and allocation will be discussed.

### 2.9.4 Methods to Reduce Attrition on Longitudinal Surveys

Peter Lynn, University of Essex (United Kingdom)

Laura Fumagalli, University of Essex (United Kingdom)

**Abstract** We report on large-scale randomised experiments carried out to study possible methods of reducing panel attrition. Specifically, the experiments address issues associated with a) ability to locate sample members and b) willingness of sample members to continue to participate. The experiments were carried out in 2008 between waves 17 and 18 of the British Household Panel Survey, involving a sample of around 12,500 persons.

The first experiment extends ideas in Couper and Ofstedal (2009) concerning the process of tracking sample members. It involved change-of-address notification cards, with four treatment groups. One group received no card, the other three groups received a card with different types of incentive/reward for returning it. The outcomes at the subsequent survey wave were systematically recorded: not just response outcome but also the costs of field tracing resulting from un-notified changes of address. We will assess relative costs of alternative procedures as well as impacts on non-response bias. We will consider also the possibility of developing tailored approaches based on previous wave tracing history and other indicators.

### 2.9.5 The Use of Scanner Technology for Collecting Expenditure Data

Andrew Leicester, Institute for Fiscal Studies (United Kingdom)

Zoë Oldfield, Institute for Fiscal Studies (United Kingdom)

**Abstract** Household expenditure data is vital for many areas of analysis, from the analysis of long-run poverty and inequality to the construction of national price indices. Specialist expenditure surveys tend to be cross-sectional and record detailed expenditures using a short-term diary to minimise respondent burden, but which lack any panel element. Expenditure questions in more general surveys generally ask respondents to recall expenditures on different broad categories, meaning much of the detail is lost and making the figures prone to recall biases or rounding.

An alternative source of expenditure data comes from market research companies. Barcode scanners are installed in the homes of participants which allows the detailed items purchased to be quickly and accurately recorded. This scanner data has three key advantages: it is a panel expenditure data source which allows household responses to socioeconomic events to be tracked; it provides extremely detailed and disaggregate expenditure data; and it combines price and quantity data in one source. However, not much is known about the implications of this mode of data collection in terms of response, attrition and representativeness. There may be concerns that some household types will find it particularly hard to use the technology (such as the very elderly) or that households with little time may quickly tire of scanning all their purchases and attrit. Using data from market research company TNS on the food and grocery purchases of tens of thousands of households over 6 years, our research explores these issues by comparing the expenditure and demographic details of the data to existing data sources more traditionally used in social science research, including the EFS, the Census and the British Household Panel Survey. We aim to document and explore issues around response and representativeness and to assess how far we can attribute differences across the surveys to the mode of data collection.

## **Chapter 3**

# **Data Archive Teaching**

## 3.1 Access to Survey Data on the Internet (I)

### Coordinator:

Kirstine Kolsrud, Norwegian Social Science Data Services (NSD) (Norway)

### 3.1.1 Access to Survey Data on the Internet

Kirstine Kolsrud, Norwegian Social Science Data Services (NSD) (Norway)

**Abstract** Access to survey data on the internet has over the last half decade become increasingly more common. One might even say the internet has become the primary means of dissemination of survey data. Large cross national surveys like the International Social Survey Programme (ISSP), the European Social Survey (ESS), the Comparative Study of Electoral Systems (CSES), the European Value Study (EVS) and the World Value Survey (WVS) all make their data available over the Internet.

Software systems like Survey Documentation & Analysis (SDA), Nesstar, JD-systems, The Dataverse Network and others are increasingly used by data providers, data archives and others to offer not only access to survey data but also online analysis and metadata.

For the social science data archives these software systems offers great opportunities to make data (not only survey data) directly accessible for the research community. At the same time it is also true that that dissemination of data over the internet poses some challenges on the distributors. For one the need for high quality and readily available documentation becomes increasingly important when the users are no longer in direct communication with the data providers.

With the rapid developments in the information technologies and the increasing number of possibilities the internet offers to communicate, cooperate and share work, it is also more challenging (but also more fun) to plan for the future. How do data providers meet these challenges? What do the users want?

In this session we will first have a closer look at two of the online distribution and analysis tools, the SDA and Nesstar and what they can offer. Secondly some of the experiences GESIS in Cologne has had with offering online access to survey data will be presented.

Looking more to the future, the activities that are being undertaken to upgrade the research infrastructure of the Council of European Social Science Data Archives (CESSDA) will be also be presented.

In addition, presentations on more specialised tools and resources will cover topics like post harmonisation, contextual data, and automated disclosure protection.

### 3.1.2 Making Use of Online Survey Documentation & Analysis

Michael Terwey, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Several sites providing online access to data and/or metadata have come into practice (e.g. ESS/NSD, IQSS Dataverse, ASEP/JDC data bank, ICPSR, CCESD, badason, ZACAT, Statista). In this presentation the online service of SDA (Survey Documentation & Analysis - <http://sda.berkeley.edu>) is focused. This set of programs provides documentation and web-based analysis of survey data. It is maintained by the Computer-assisted Survey Methods Program (CSM) at the University of California, Berkeley.

At this time, SDA is considered from two perspectives, those of a data service manager and of an empirical sociologist. My attention was drawn this site, because it offered straightforward access to some important data as GSS, American National Election Study, and Census Microdata. The facilities were used for various purposes as investigating metadata, data download, statistical analysis, and finally for the pilot implementation of special surveys into the system (ISSP and ALLBUS). Comparisons across time were done concerning religious beliefs and social pessimism.

Starting major tests with the integration of ALLBUS 1998 into SDA the facilities of looking into metadata were already rather satisfying, but later on some important improvements could be added (especially text search within a single study and across studies). This improved the relative position SDA formerly had in comparison with some similar systems.

Analysis of data has always been exceptionally quick. The access to the basic statistical procedures is pretty easy even for newcomers with only basic statistical and/or computer knowledge. Helpful documentation on statistics is online. Moreover, interested users are not restricted to the given original data, but may create and even store their own user-defined variables. New highlights at SDA concern safeguarding against disclosure risks and offering complex standard errors in some statistical analysis, which are two quite relevant implications for elaborate statistical analyses.

Unrestricted download of even large files in various formats is provided. Download of customized metadata and data sets turn out to be straightforward. On the other hand, there are also possibilities of controlling the access, if such protection is needed.

Of course, there would be challenges for future work (e.g. implementing more exploratory statistical procedures (e.g. PCA, LCA)). Nonetheless, as data manager and as researcher I do confirm that SDA has really proven to be useful online tool.

### 3.1.3 Nice portal! But where is the data ...? - Experiences of a data archive with offering online access to survey data

Reiner Mauer, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** In 2006 the department Data Archive and Data Analysis of GESIS (formerly known as Central Archive for Empirical Social Research, ZA) introduced its NESSTAR based data portal ZACAT (<http://zacat.gesis.org>). Maintaining a large archive for social science mainly survey data, GESIS disseminates a large number of data sets and related materials through different channels to the international scientific community. Since the implementation of its first web based download system in 1999 for selected data sets of the German General Social Survey (ALLBUS), dissemination of data through the internet has become increasingly important, both for data users and distributors.

Establishing ZACAT marks a significant step forward in our efforts to offer major parts of our holdings via internet. After a quick online registration procedure users have direct (and easy?) access to the latest versions of data sets and documentation from study series like Eurobarometer, ISSP, European Values Study etc.

After a brief introduction to contents of ZACAT and some special features supporting comparative research, this talk will focus on our experiences with distributing survey data via internet. Particular emphasis will be given on using NESSTAR as a distribution channel. Who uses such a portal and for what purposes? How does NESSTAR meet the needs of data users when it comes to identifying appropriate data and accessing this data online? Which issues are involved for data publishers when distributing data and metadata through the internet? Finally, I will try to give a short outlook on potential directions of our future strategy with regard to online dissemination of data.

### 3.1.4 Using the Available On-line Secondary Data in Education and Research Practice

Jolanta Perek-Bialas, Warsaw School of Economics (Poland)

**Abstract** Nowadays the secondary data sources (especially surveys with free on-line access to raw data like in Poland: Social Diagnosis, but as well on European level: ESS, SHARE or international one like VWS) - they are used not only in teaching (ie. for students of social sciences disciplines like sociology and economics) but also in various areas of research. The specific similarities and differences of some available secondary data will be shown. But the main aim of the presentation is to answer the question how effective for survey methodologists is to use available secondary data in teaching and research practice? It leads also to the discussion how users could via available portals of existed sources to check in standardized way their quality (i.e. based on additional technical documentation like about sampling procedure/weights). So, there will be shown advantages and disadvantages of some solutions chosen by providers of these secondary data sources available on line but also with suggestions of some possible changes which could be helpful in better using them by different actors in future. This whole evaluation and assessment will be based on my teaching experience in this field along with students opinions and recommendations what could be changed/improved to use available secondary data better and correct in education and as well also in research practice. At the end, also a few interesting examples of using available secondary data will be presented (ie. in the special on-line project for the Ministry of Labour and Social Policy and in students papers).



## 3.2 Access to Survey Data on the Internet (II)

### Coordinator:

Kirstine Kolsrud, Norwegian Social Science Data Services (NSD) (Norway)

### 3.2.1 The CESSDA Research Infrastructure: new developments in data and metadata dissemination, standardisation and harmonisation

Hillary Beedham, University of Essex (United Kingdom)

**Abstract** This presentation will outline the activities that are being undertaken to upgrade the CESSDA Research Infrastructure (RI) for the benefit of researchers in the European Research Area, focussing on the need for standards and harmonisation to achieve many of its goals. The undertaking has been possible following a major award totalling 2.7 million euros and is a direct result of the CESSDA RI being identified by the ESFRI Roadmap [3] as an existing pan-European RI recommended for a major upgrade.

Collectively, CESSDA members hold over 25,000 data collections which are made available via online services to researchers in the social science and humanities. The amount of datasets available will increase as the CESSDA RI develops and engages with data producers, (for example Inter-governmental organisations, National Statistical Institutes and other organisations who commission publicly funded research). Our goal is to allow easy data discovery and access from the CESSDA portal; a 'one-stop-shop' built on tools and underlying systems that will permit seamless access to data for researchers regardless of the location of either data or researcher. Central to this goal will be the need for harmonisation and the adoption of common standards by all member organisations. The enhanced CESSDA be responsible for: identifying appropriate and necessary standards for its membership; assisting prospective members to meet these; and monitoring ongoing adherence by its member organisations.

### 3.2.2 Beehives and Harmonisation Routines, or: How to Make the Unpublishable Public. The Approach of the CESSDA Survey Data Harmonisation Platform

Markus Quandt, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Social science has come to regard empirical comparisons across space and time as a matter of course. A host of methodological work is focused on evaluating the validity of such comparisons across data from different sources. In contrast, much less attention is being paid to the practical task of making existing data sets more comparable. Practical harmonisation work is already an issue for programmes such as ESS or ISSP, which design their surveys for comparability from the outset. But beyond that, vast amounts of data not collected with comparability in mind can be highly relevant for comparative research. When such data are manipulated to improve comparability, researchers often fall back on ad hoc considerations, using a diversity of tools selected mostly by random availability, and usually creating no or little documentation. Harmonisation work in itself is usually not regarded to be worth publishing. The result is a lack of supply of proper harmonisation work, and a lack of transparency in what is available.

The European data archives (as represented by CESSDA) have therefore designated one work package of a larger project to developing an open web platform for systematic ex post-harmonisation work. This platform shall offer researchers the tools to actually perform proper data harmonisation work online, and it shall at the same time support the publication of finalised harmonisation routines. The platform will be open to different types of contributors, individual and institutional, for inputting or editing harmonisation routines. Unrestricted and citable publication of a harmonisation routine is expected to form an incentive for contributors. Publication of the routines enhances the intellectual property rights of the originators in the same way that publication of research results earns credit to the authors. Further, it creates transparency on the actual coding of harmonised data and thus may also foster scientific discussion about particular routines and thus increase the quality of harmonisation work at large. The presentation will inform about how the platform intends to realise these goals.

### 3.2.3 Linking survey data with their context - the ESS contextual data repository

Gyrid Havag Bergseth, Norwegian Social Science Data Services (NSD) (Norway)

**Abstract** Survey data is increasingly used to understand how individuals think and act, and governments, industry and academia are becoming more and more dependent on this kind of data. Much of the reason for this is that survey data are manageable, relatively easy to collect and can sustain complex analyses. The European Social Survey (ESS), the International Social Survey Program (ISSP), the General Social Survey (GSS), the World- and European Value Survey (WVS and EVS), and the Eurobarometer are all examples of widely known and used surveys.

However, when analysing survey data, it has become more and more common to focus on placing the individuals concerned in a society with history, geography, social relationships or social and institutional structures. By including contextual data in the analysis scientists are able to get a more detailed picture, and a better understanding of individual behaviour and attitudes.

Especially when it comes to cross national surveys this element is important. Not only on a national, but also on regional level, individuals are influenced by the society in which they live. But even though contextual data are widely available both online and through other channels, they are not always easily accessible, and getting access and preparing the data for analysis can be time consuming and thus a hindrance for many potential users. This is especially the case for data at sub-national (regional) levels.

As a response to this, the ESS is developing a contextual data repository where contextual data are directly linked to the ESS survey data. The goal is to ease the process of finding and linking contextual data to the ESS survey data. Key elements in the development of the ESS contextual data repository have been the harmonisation of data and metadata, functional browsing and search, development of functional basic analytical tools adapted to data on multiple levels, the possibilities for cartographic presentation and easy download of data/variables and metadata.

This presentation will describe some of the challenges in the ESS contextual data repository project. This includes both the functionality accommodating the repository editor, and the needs of the end user. On the one hand linking contextual data to the survey data has not been straightforward. For example, the regional levels in a survey organisations sampling plan are not always the same as the statistical units. On the other hand, it is vital that data are presented in a way that will suit the end user. The wants and needs of the research community are highly heterogeneous, both regarding functionality and data.

### 3.2.4 Online Analysis and Programmed Disclosure Risk Protection: New Access to Restricted-use Microdata

JoAnne McFarland O'Rourke, University of Michigan (United States)

Sarah H. Rush, University of Michigan (United States)

Christopher Maxwell, University of Michigan (United States)

**Abstract** The challenges of protecting subjects in public-use microdata while maximizing the analytic utility of these data have been discussed in the disclosure literature. This paper describes a new version of Survey Documentation and Analysis (SDA), an online data analysis program that performs bivariate cross-tabulation, comparison of means, correlation and regression analyses. Offering the same analytic capabilities as SDA, Secure SDA (SSDA) \* is designed to provide a safe, reliable way to publicly distribute restricted-use data, thereby democratizing access to data previously unavailable or that required special procedures to obtain. SSDA's development focused on three areas: enhancing analytic capability, programming options to restrict both access and analytic output and determining how to provide maximum electronic security of SSDA data. SSDA has improved its analytic capacity with the proper calculation of variances to adjust for complex sample designs in means and tables analyses. To provide maximum flexibility and utility, SSDA automates several disclosure protections that prevent the use of organization-defined high-risk variables, singly or in combination, as well as output (e.g. unweighted n's). When the organization-defined rules are violated by an attempted analysis, the resulting output is fully or partially suppressed. The Web server structure was configured to provide optimal system security. These protections along with SSDA's features and analytic capabilities are illustrated using data from the National Survey on Drug Use and Health (NSDUH). Future enhancements to the SSDA system are also discussed.

### 3.3 How to facilitate in order to compare - infrastructures for data cooperation

#### Coordinators:

Brigitte Hausstein, GESIS - Leibniz Institute for the Social Sciences (Germany)

Agnieszka Wenninger, GESIS - Leibniz Institute for the Social Sciences (Germany)

#### 3.3.1 National Standardization on the example of the German Demographic Standards versus International Harmonization

Jürgen H.P. Hoffmeyer-Zlotnik, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** The German Demographic Standards for Surveys and Polls are a joint recommendation by Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute, and the Federal Statistical Office.

The initiative started in the early 1980s. All five years a revision is done, so that authors can control whether social change or new laws demand new response categories or a reformulation of a question. The next issue is announced for 2009.

The goal behind Demographic Standards: In the past each project is formulated its own questions asking about the socio-demographic and socio-economic variables. Comparison of data between projects or with census or micro census was not possible. Today the German Demographic Standards offer a basic questionnaire for those set of questions widely used by projects in social or market research are using.

The German Demographic Standards present the operationalization of the variables: gender, age, citizenship, marital status and partnership, education, occupation and occupational status, size of household, personal income and household income.

In the presentation first, the composition of the questionnaire is explained as well as the planned revision; second, the use of the questionnaire is explained a researcher is not forced to use all questions or all categories of one question; third, the benefit for using the Standards is explained; fourth, single items like education, occupation, household and income are discussed under the aspect of harmonization for international comparison.

#### 3.3.2 The Relevance of 'private Household' for Cross-cultural Survey Comparison

Uwe Warner, Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques (CEPS / INSTEAD) (Luxembourg)

**Abstract** Different cultures and states use their national definition of private household. In the EU nearly each country has an own definition of household. These definitions correspond to the cultural and national structures of social life. The differences result in diverse household compositions and unequal sizes across European nations. Comparing household measures over countries survey analysts face several inconveniences.

The composition of the surveyed household has direct impact on the respondents answer about the household size. Therefore an effect on total net household income is visible. The composition of the private household plays a role with regard to household income since the latter can vary considerably depending on the number of persons who contribute to it and the composition of household income sources in question.

Different membership rules constitute also differences in the socio economic status of the individual household members when the person in the private household with the highest status determines the status of the group as a whole.

With regard to the sociological variables "total household income" and "SES", the composition of the household and, therefore, the definition by means of which this composition is determined, is of central importance. In a first step we summarize definitions of household used in national surveys across Europe. Same dwelling, sharing economic resources, common housekeeping and family ties are the main and mostly used criteria. In a second step we discuss the possible combinations of these elements and the strategies of operationalization in social surveys. The third part illustrates the findings. We use ESS, ECHP and administrative micro data from official statistics. The country differences become obvious.

Our conclusion is a revised fieldwork instrument measuring household in social surveys that increases data comparability across cultures and countries. However, harmonization cannot mean that the respondents from all

participant cultures must adopt the definition used by one selected culture. Because, even in a national context, it is likely that, in the case of a tacitly understood private household concept, researchers, interviewers and respondents nonetheless have different definitions and, accordingly, different perceptions of household composition. Indeed, even among respondents, there is probably no consensus as to what the term "private household" means.

### 3.3.3 Standardization and Harmonization of the Data from the Polish Panel Survey, 1988-2008: Twenty Years of Experience

Kazimierz M. Slomczynski, Ohio State University (United States)

**Abstract** The data for this paper come from the Polish Panel Survey, POLPAN, conducted in 1988, 1993, 1998, 2003, and 2008, containing rich information about all respondents jobs, changes in the family structure, personal history of political activities and other elements of individual biographies, including attitudes and opinions. POLPAN is the only panel survey in Europe in which the same men and women, constituting a national representative random sample at the beginning of the study, were interviewed over a twenty year period. Since the original panel sample is naturally aging, sub-samples of younger respondents have been added to the core of the panel sample. The sub-samples consist of persons aged 21-25 years old in 1989-1993, 1994-1998, 1999-2003, and 2004-2008. At present, GESIS has archived the first four waves. In the future, the last wave will be archived together with separate files for panelists only. In this paper the standards of constructing variables are discussed, separately for (a) the basic socio-demographic and social-stratification variables, and (b) selected variables pertaining to attitudes and opinions. In the case of socio-demographic and social-stratification variables a comparison with standards of cross-national academic and commercial research is provided. The properties of the POLPAN variables pertaining to attitudes and opinions are assessed and evaluated together with the analogous variables in the World Value Survey, the International Social Survey Programme, and the European Social Survey. The paper ends with a discussion of methodological implications of having standardized universal and country-specific indicators for various domains of social life.

## 3.4 Quality Enhancement in Repeated Surveys

### Coordinators:

Eric Harrison, City University London (United Kingdom)  
Anouk Zabal, GESIS - Leibniz Institute for the Social Sciences (Germany)

### 3.4.1 Building a Quality Enhancement Network for a Cross-national Survey

Eric Harrison, City University London (United Kingdom)  
Anouk Zabal, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Repeated surveys face a number of challenges which can threaten to erode their quality over time. These include shortfalls in funding, the decline in salience of their original content, turnover of central personnel, changes of survey agency, and alterations in the survey environment such as the availability and quality of sampling frames. This paper describes a strategic initiative mounted on the European Social Survey to maintain and enhance quality over a five year period through the construction of a Quality Network. Three types of activity are outlined: thrice-yearly gatherings of National Co-ordinators to discuss data collection and methodological refinements; a symposium of field directors from countries participating in ESS; and a series of three Quality Enhancement seminars where experts in particular fields evaluate current practice and recommend changes in practice to maintain high standards in all aspects of the survey. The paper evaluates the effectiveness of the overall initiative to date and discusses how such a network could be better integrated into the European Social Survey in future rounds.

### 3.4.2 The View from the front lines – a contractor view of quality enhancement

Nick Moon, GfK NOP (United Kingdom)  
Sarah McHugh, GfK NOP (United Kingdom)

**Abstract** The purpose of this paper is to give a contractor's eye view of the kind of quality issues that arise out of conducting a very high profile longitudinal study over many years. Many of the quality requirements were already part of our standard procedures, but the particular requirements of the survey led us internally to improve procedures in other areas, while other changes were made after external requests from the client, and in both these cases the changes made have led to improvements that have been carried through to other surveys.

### 3.4.3 Refusals in longitudinal mental health studies: causes and consequences

Johannes H. Smit, Vrije Universiteit Amsterdam (Netherlands)  
Adriaan W. Hoogendoorn, Vrije Universiteit Amsterdam (Netherlands)  
Chantal Hoekstra, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** Results of mental health studies could potentially be biased by non-participation of respondents. Non-contact, refusal, ineligibility or death are main causes of non participation and especially in longitudinal (mental) health studies biased results can occur because non-random attrition is to be expected at each wave. In order to understand this bias it is important to determine the differences between those respondents who will continue participating in the study and those who do not. Also, such samples may have reasons for refusal which are specific to groups with mental health problems, such as lack of general motivation and concentration problems.

In studying non-response, refusal is of particular interest because it is amenable to change. The more we know about those who have been contacted, but refuse to take part, the more we are able to tailor our approach and hopefully lower our refusal rates. Especially in longitudinal studies much more is known about those who do not participate in second and subsequent waves and therefore longitudinal studies are prone to study characteristics of respondents and determinants of those who stop participating in the study.

The aim of this paper was to examine the relationship of refusals in the second and third wave, to the demographic, physical, psychological and social state of participants at subsequent waves in The Netherlands Study of Depression (NESDA). A longitudinal mental health survey among 2981 persons at baseline (2006). Reasons for refusal are discussed as well as possible solutions to reduce these refusal rates. Also consequences for longitudinal data-analysis and imputation methods are introduced.

### 3.4.4 Improving quality over 25 years of the British Social Attitudes Survey

Alison Park, National Centre for Social Research - NatCen (United Kingdom)

Miranda Phillips, National Centre for Social Research - NatCen (United Kingdom)

**Abstract** The British Social Attitudes (BSA) survey is NatCen's annual survey of the British general public, which aims to monitor and analyse consistency and change over time in attitudes and values. This talk will examine the ways in which the quality of the survey has improved since it began, and discuss how concerns about quality improvement have to be balanced against the importance of ensuring that data remains comparable from one year to the next. A range of topics will be addressed, from the overall survey design (sampling and weighting; interview mode), to procedures for question testing and data production.



## **Chapter 4**

# **Design and Quality of Survey Questions**



## 4.1 Can multiple question testing and evaluation methods improve survey questions or predict sources of measurement error?

### Coordinators:

Margaret Blake, National Centre for Social Research - NatCen (United Kingdom)

Debbie Collins, National Centre for Social Research - NatCen (United Kingdom)

Michelle Gray, National Centre for Social Research - NatCen (United Kingdom)

### 4.1.1 Using qualitative and quantitative methods to develop ethnic identity questions in the UK

Alita Nandi, University of Essex (United Kingdom)

Lucinda Platt, University of Essex (United Kingdom)

**Abstract** The measurement of ethnicity and ethnic identity in survey research is a vexed question. While, the problems with single, uni-response categorical measures have been much discussed and it is widely recognised that they conflate disparate concepts and only capture certain "dimensions" of ethnicity, their ease of use and ability to distinguish between populations has nevertheless rendered them widely used in UK surveys and there have been few attempts at alternatives or supplements in national large-scale representative surveys. At the same time there has been burgeoning interest in questions of ethnicity and identity within the UK and a large, predominantly qualitative, literature addressing these issues. The inauguration of a new UK panel survey, Understanding Society, the UK Household Longitudinal Study, with a large sample size, a strong "ethnicity strand" and a separate "Innovation Panel" for advance methodological testing has provided the opportunity to develop more comprehensive measures of ethnicity, using both qualitative and quantitative approaches.

Our approach to question development has involved placing new questions in the Understanding Society Innovation Panel to evaluate their acceptability (as measured by non-response), and how responses (in particular "don't know" or non-response) may vary with relevant known characteristics collected in an earlier wave, such as country of birth, categorical ethnic group, age and sex. It also enables us to examine differences in responses across modes (telephone and face-to-face). We have complemented this by using focus groups and individual interviews to build up information on how people talk about ethnicity and ethnic group, what terms and concepts make sense to them and how they understand them generally and personally. A particular issue, here, is how "ethnicity" is understood by the majority, where dominance tends to normalise and thereby render invisible (and potentially inexpressible) their own ethnicity. An earlier phase of cognitive testing had identified a preference for multiple response measures in ethnic group questions, and in the interviews we explicitly tested forms of wording for a suite of multi-dimensional questions allowing multiple response. The integrated results from both these phases will determine the wording and form of questions illuminating ethnic identity, which will themselves be subject to a further phase of qualitative (cognitive testing) and quantitative (piloting) evaluation.

In this paper, we discuss the conceptual and measurement issues facing survey measurement of ethnicity, and common approaches in both the UK and other countries. We then go on to describe our approach to developing new questions using both qualitative and quantitative methods. We also highlight the particular advantages of being able to develop questionnaire content in the context of a longitudinal survey. Finally we conclude with a discussion of the ethnic identity questions and the ways in which they were shaped by the different elements and phases of the quantitative and qualitative development.

### 4.1.2 The use of behaviour coding and cognitive interviewing methods to evaluate a family function questionnaire: Are the evidence comparable?

Miguel Castillo, University of Granada (Spain)

Isabel Benítez, University of Granada (Spain)

Jose Luis Padilla, University of Granada (Spain)

**Abstract** Cognitive interviewing and behaviour coding are methods frequently used in pre-testing questionnaires. Both methods can identify problematic questions and sources of measurement error investigating the question-and-answer process. Evidence provided by identifying sources or measurement errors allow survey researchers to improve the validity of survey data. The aim of the present study was to identify possible problems in the question-and-answer process to a family function questionnaire items by means of behaviour coding and cognitive interviewing. The tested questionnaire intends to measure five components of a family function construct: adaptability, partnership, growth, affection and resolving capabilities. The questionnaire consists of 5 items in a rating scale format with 3 response options. The function family questionnaire is included in a health survey questionnaire. 21 cognitive interviews were performed together with a behaviour coding analysis. Data from behaviour coding were analysed using Sequence Viewer Program (Dijkstra, version 4.4.a. 2008), and data from cognitive interviewing by AQUAD Program (Huber, version 6.8.1.1, 2008). In general terms, cognitive interviewing identified comprehension problems with respect to the family concept. For example, 78.95% of responses do not fit the intended family concept. Furthermore, problems with the response options were flagged, founding a high percentage of mismatch answers when applying coding behaviour analysis. How cognitive interviewing and behaviour coding can be complementary methods is also illustrated. Finally, we also discuss the conditions which must be taken into account to improve the comparability of evidence provided by both cognitive pre-test methods.

### 4.1.3 Combining qualitative and quantitative methods in survey development: An example from the Life Opportunities Survey

Joanna Bulman, Office for National Statistics (United Kingdom)

**Abstract** In Great Britain there has been a growing need for information about the dynamics of disability over time and the factors associated with transitions and onsets of disability. A new survey due to be launched by the Office for National Statistics will monitor the experiences and barriers that disabled people face from early childhood through their life course, how these relate to the type of impairment and how they change over time.

In recognising the conceptual shift from a medical to a social model definition of disability, the survey called for new measures of disability to be developed. These new measures, which focus on how social and environmental barriers interact with impairments to construct participation restrictions, were developed and tested through a combination of quantitative and qualitative methodologies.

The qualitative work began with a deliberative consultation exercise with disabled people, who were meaningfully involved from the start of the development process. A reference network was convened, comprised of 60 people with a range of impairments who took part in focus groups, conferences and online discussions. Findings from this consultation fed directly into the design of the questionnaire, which was qualitatively tested using cognitive interviewing techniques. Key aims of this phase of testing were to explore how respondents reacted to the proposed questions, how they understood them and if they were able and willing to answer them.

The initial version of the questionnaire went through major changes as a direct result of the qualitative testing. This revised version was then tested by means of traditional quantitative testing, however since the questionnaire had been thoroughly cognitively tested, the quantitative phase was less concerned with testing the wording of the questions, and more about reviewing field systems and practices. A feasibility trial was carried out which involved interviews at 118 addresses, purposively sampled to include a high proportion of disabled people. This was followed by a larger scale pilot, which focused on providing an indicative response rate and interview length, as well as a final opportunity to test the questionnaire and field processes.

### 4.1.4 Direction of response scale format and response behaviour

Dagmar Krebs, University of Giessen (Germany)

Yaacov G. Bachner, Ben Gurion University of the Negev (Israel)

**Abstract** Formatting an answer as one step in the cognitive process of answering a question means to map the selected answer to the categories on a response scale. In this process response scales, as one part of questions, serve as orientation for respondents in mapping their answers to response categories. Therefore it can be expected that the direction of a response scale (either starting with the positive or the negative response category on the left side of the scale) provides information for respondents influencing their response behaviour. If left means top, then starting a response scale requesting approval of a behaviour with, for example, completely approve or with completely disapprove might result in different subjective hypotheses that respondents form about the

most frequent or most accepted behaviour. This would imply that direction of reading is important in reacting to response scales.

If this principle applies in general, then it should be observable in response behaviour of respondents in different contexts: first in a context, where reading direction and accordingly response scales run from left to right, and second, in contexts where reading direction runs from right to left. In the latter context, the right side of the scale would carry the meaning of top and response behaviour can be expected to vary accordingly comparable to response behaviour on scales running from left to right.

This paper tests the hypothesis that first, scales beginning with the positive pole bring about different responses than scales starting with the negative pole. Second, it is tested that reading direction is important resulting in comparable response behaviour under the condition of different reading direction.

The condition direction of reading has been varied by using a German and an Israeli student sample. Within each sample 7-point scales with different direction (negative-positive as well as positive-negative) and with labels only at the extreme points have been applied. Question content was, besides others, extrinsic as well as intrinsic job motivation.

A panel design was combined with split-ballot forms of the questionnaire. Therefore it is possible to test the postulated effects of response scale direction on response behaviour by comparing responses across split groups within each sample as well as across samples within each split group. Split groups of respondents did only participate once, panel members participated either twice or three times. Results of this step reveal if there is an effect of scale direction and if this depends on reading direction.

Additionally, by analysing an MTMM model based on the repeated measures with different response scale direction, it is possible to uncover the effect of the scale direction as a method on the reliability as well as on the validity of indicators and constructs. This can be done for the German and Israeli group separately (to assess the method effect of scale direction within each group) or in case results are comparable across the two groups combining the two groups to assess the strength of an overall method effect of response scale direction. Results of the MTMM model complete and validate results of the comparison across split groups.

#### 4.1.5 Facilitating triangulation of pretesting techniques through improved documentation of (cross-national) questionnaire development

Rory Fitzgerald, City University London (United Kingdom)

**Abstract** Despite increases in the transparency of elements of survey design in recent years, the process of questionnaire development has somehow bucked this trend. And the detailed decision making process for handling pilot data often remains a mystery. In the European Social Survey (ESS), we have tried to rectify such omissions both to improve transparency and to improve communication and decision making. We have therefore developed a coordinated approach to the post pilot decision making process, triangulating different sources of evidence on the performance of survey measures.

Half of the ESS questionnaire consists of rotating modules on specific topics awarded to multinational teams. These modules go through a protracted period of design and development which aim to achieve reliability and validity, as well as attaining equivalence across countries and languages. This involves a wide range of quantitative and qualitative methods. Based on all of this input and the advice of substantive and methodological experts, items or scales are then adopted, re-designed or discarded.

The work on the rotating modules is carried out jointly by a selected team of academics in collaboration with a subset of experts from the Central Coordinating Team (CCT) of the ESS. During Round 4 we improved documentation of the questionnaire design process by asking the Question Design Teams and the CCT to document the process via a pre-designed template. Following a two nation pilot, a series of different types of evidence were entered into the template then all considered simultaneously. This included evidence from a range of methods including cognitive interviewing, feedback from the translation process, factor analysis to test expected relationships and anticipated scales in the data, non response analysis, split ballot analysis, other quantitative analysis to examine distributions in the data as well as evidence from interviewer debrief feedback.

This paper provides a methodological investigation into the use of these different evaluation and testing methods. In particular the paper will look for evidence of whether the data obtained using these different methods was comparable. Key questions include:

- 1 What evidence was there that different pre-testing techniques found the same errors?
- 2 What kinds of data did the different pretesting methods provide and how useful were these to the question designers?
- 3 How beneficial was it to present all of the evidence in a single template to facilitate multi method informed decision making?

References:

Converse J and Presser S (1986), *Survey Questions. Handcrafting the Standardized Questionnaire*, Sage: Beverly Hills.

#### 4.1.6 A cognitive map and Formal Concept Analysis approach for building and testing alternative models

Ahmet K. Suerdem, Istanbul Bilgi University (Turkey)

**Abstract** This paper presents a methodology aiming to bridge exploratory and confirmatory data analysis. First step involves an exploratory factor analysis of the data. In the second step, cognitive maps are elicited through qualitative interviews. Cognitive maps are cause-effect networks, with nodes representing concepts articulated by individuals, and directional linkages capturing causal dependencies. A group of social science experts are asked to draw the causal relations between the concepts extracted from the exploratory factor analysis in terms of nodes and directed arrows. Elicited cognitive maps are translated into concept matrices. In the third step causal relation dyads obtained from cognitive mapping enter as attributes and experts as objects to Formal Concept Analysis. Formal Concept Analysis is a classification system which clusters attributes and objects within a conceptual structure. This structure is presented in terms of a Gallois lattice. In the fourth step, clusters of experts and causal relation dyads are used for determining competing theoretical frameworks. The cognitive maps of the experts located within the same cluster sets are aggregated and concept co-occurrence matrices representing alternative theoretical frameworks are obtained. These matrices are used to present the group cognitive maps in terms of network diagrams. As a final step, concept network diagrams representing alternative theories are translated into Structural Equation Models. These alternative models are tested in terms of their fitness to the data used in the exploratory factor analysis. The constructs obtained from the best fitting model can be used for the design of questionnaires involving measurement scales with higher reliability and validity.

The data set used in this research is composed of seventeen attitudinal indicators which operationalise attitudes towards social, economic and political issues. The sample is representative of Turkish population (n=1500). The recovered factors are named as moral conservatism, position towards privatization, position towards parliamentary democracy, dogmatism, position towards redistribution and solidarity vs. individualism orientation. Besides the concepts representing the exploratory factors, theoretical concepts such as left-right, conservatism and authoritarianism are also involved during the elicitation of cognitive maps.

## 4.2 Implementation Issues in Internet-based Surveys

### Coordinator:

Anja Göritz, University of Würzburg (Germany)

### 4.2.1 Twisting Rating Scales: Horizontal versus Vertical Visual Analogue Scales versus Categorical Scales in Web-Based Research

Frederik Funke, University of Tübingen (Germany)

Ulf-Dietrich Reips, University of Zürich (Switzerland)

**Abstract** Changing the alignment or type of rating scales is a simple way of expanding the possibilities for designing questionnaires. Flexibility in design is appreciated in Web questionnaires, for example to be able to optimize questionnaires according to the respondent's screen format (especially in mobile Web surveys to avoid bias that occurs when respondents have to scroll to see all response options). But there is a danger that changes in response scales can impair data quality.

We manipulated the available rating scale in a 2 (alignment) x 2 (type of scale) factorial between-subjects design. Alignment was either horizontal or vertical. For type of scale we used discrete five-point categorical scales and continuous visual analogue scales (VASs). VASs were implemented as plain lines with 250 possible values and both ends verbally anchored. VASs are appreciated, e.g. in medical research, because small differences can be detected within- or between-subjects. However, they are rarely used in the social and behavioral sciences, most probably because the analysis of data from non-computerized VASs is burdensome and because of a considerable lack of methodological research.

#### Method and Results

392 students participated in a Web experiment with a questionnaire consisting of 40 items on the Big Five dimensions of personality. Neither type of scale nor alignment mattered regarding dropout and lurking (i.e. participants rating not a single item at all). Missing data occurred more often with 5-point scales than with VASs,  $\chi^2(2, N = 391) = 5.24, p < .05$ , alignment did not matter. Neither alignment nor scale type had a statistically significant effect on means of the five dimensions of personality.

Thanks to computerized data collection we were able to collect two paradata: response time and frequency of modified answers. Neither type of scale nor alignment showed an influence on response time. With VASs ratings were modified more often than with 5-point scales,  $F(1, 390) = 49.83, p < .001, \eta^2 = .114$ . Alignment had no influence on frequency of modified answers. There were no interactions.

Five items (one from every dimension of personality) were repeated at the end of the questionnaire. We found test-retest reliability (Cronbach's alpha) to be a little higher with 5-point scales ( $M = .92$ ) than with VASs ( $M = .90$ ). Horizontal alignment led to higher scores ( $M = .92$ ) than vertical alignment ( $M = .90$ ). Consequently, horizontal 5-point scales led to the highest reliability scores ( $M = .93$ ) and vertical VASs performed worse ( $M = .88$ ).

Overall, alignment can apparently be used as a means of design with five-point categorical scales as well as with VASs without running the risk of seriously biasing data. However, it should be kept in mind that scales of short to medium length were used in this study. Further research with longer scales is needed.

### 4.2.2 The influence of the field time on data quality in list-based Web surveys

Anja Göritz, University of Würzburg (Germany)

Stefan Stieger, University of Vienna (Austria)

**Abstract** A short field time is an often-cited benefit of Web-based surveys that rely on pre-recruited people. However, it has never been examined how different field times as implemented through different deadlines for participation influence response behaviour. Four experiments were conducted in which the deadline for taking part in the study was varied across several days, and there was a control group who were not told any deadline. We examined the impact of both stating a deadline versus not stating a deadline and the length of the deadline on the response rate, the retention rate, and response completeness. It was found that response rises with the number of days a study is in the field. There is tentative evidence that the more generous the deadline the smaller the retention rate and clear evidence that response completeness is lower.

### 4.2.3 Effectiveness of incentives in mixed-mode systems: An evaluation of errors & costs

Katja Lozar Manfreda, University of Ljubljana (Slovenia)

Nejc Berzelak, University of Ljubljana (Slovenia)

Vasja Vehovar, University of Ljubljana (Slovenia)

**Abstract** Declining response rates force survey researchers to implement surveying strategies that would motivate target persons to cooperate. Incentives have often proved to be a viable approach to address this issue. They have been increasingly used also in web surveys, which are relatively cheaper, compared to other modes but at the same time suffer from generally lower response rates. Since the problems like noncoverage and lack of sampling frames still make web surveys largely inappropriate for surveying general populations, the possibility of their integration into mixed-mode designs has become one of the central questions of survey methodology. This introduced a new area also for methodological research on incentives.

While incentives can substantially increase response rates, they can also significantly heighten the research costs. Furthermore, the impact on survey responses often remains uncontrolled and is largely unclear in mixed-mode designs. In this paper, we study the problem of using incentives in mixed-mode surveys of the general population by comparing response rates, errors and costs among different experimental mixed-mode designs.

The experiment was conducted on a sample of the general Slovene population. Participants were randomly assigned into one of experimental groups, manipulating two different combinations of modes (web/mail and web/telephone) and three different types of incentives (no incentive, wallet as a gift and 5 Euros in cash). We first present the impact of incentives on response rates in each of the groups and compare errors using the Mean Square Error approach. The results show that the response rates are not a good predictor of the data quality, since the monetary incentives significantly boosted cooperation but at the same time produced a high bias. This finding is then elaborated by a detailed analysis of the response patterns that might have contributed to the bias. Possible adjustments of survey design, that may provide a higher data quality, are discussed as well. Finally, the cost aspects are taken into account and evaluated by a simulation of different initial sample sizes and their effect on the costs-errors optimization. The findings of the study suggest that the use of incentives in mixed-mode surveys should be methodologically well elaborated, not only from the aspect of response rates, but also with a careful consideration of survey errors and costs.

### 4.2.4 Increasing Confidence in Survey Estimates with Visual Analogue Scales

Frederik Funke, University of Tübingen (Germany)

Ulf-Dietrich Reips, University of Zürich (Switzerland)

Randall K. Thomas, ICF International (United States)

**Abstract** Survey data are unavoidably contaminated with measurement error. We focus on formatting error, the increase in confidential intervals of survey estimates, making it more difficult to detect existing differences. Formatting error happens when respondents do not find an option on a rating scale that perfectly reflects their true value. The difference between the true value and the chosen response option is formatting error.

From a theoretical point of view, continuous visual analogue scales (VASs) have on the individual as well as on the aggregated level an expected formatting error of zero, because there is a perfectly fitting option for every graduation of the true value. An empirical determination of formatting error with VASs is pending and it is unclear, if populations with a low formal education are able to use VASs in a meaningful way.

Formatting error with categorical scale is different for single individual variables and for aggregate data. In the first case, it depends on the number of categories only. In the second case it is additionally influenced by the actual distribution of values in the sample. We simulated differently distributed data (e.g. from uniform distributions, narrow and wide normal distributions, chi-square and exponential distributions) to determine the expected formatting error with categorical scales consisting of 3 to 21 categories.

#### Study Design

To determine formatting error we chose numeric values as a construct with a linear mental representation that should be understood by every respondent in the same way. In a Web experiment participants had to locate 15 percentages (ranging from 5% to 95%, presented in random order) on VASs anchored with 00%0 and 0100%0. In a between-subjects design respondents were randomly assigned to VASs made of plain horizontal lines or to VASs with a number of vertical equidistant markers (ranging from one marker indicating the middle of the rating scale to 100 equally spaced markers).

#### Results & Conclusion

Overall ( $N = 1909$ ), we found a very low mean empirical formatting error of  $M = -1.24$  percentage points ( $SD = 3.05$ ). Ratings on plain VASs without any marker ( $n = 167$ ) were worse ( $M = -2.48$ ,  $SD = 2.64$ ) and formal education (below college:  $M = -2.87$ ,  $SD = 2.93$ ; at least college:  $M = -1.71$ ,  $SD = 1.61$ ) made a statistically significant difference:  $F(1, 166) = 7.56$ ,  $p < .01$ ,  $\eta^2 = .04$ . VASs with ten markers ( $n = 181$ ) lead to the smallest formatting error ( $M = -0.41$ ,  $SD = 1.55$ ) for respondents with a low education ( $M = -0.47$ ,  $SD = 1.63$ ) and respondents with a high formal education ( $M = -0.30$ ,  $SD = 1.39$ ),  $F(1, 180) < 1$ .

For individual variables, the empirical formatting error for VASs with ten markers is even with respondents with a low formal education lower than the expected formatting error for categorical scales up to 50 options. Overall, the authors strongly recommend considering VASs in computer-based self-administered questionnaires for the sake of more confident survey estimates.

#### 4.2.5 Presentation of WEBSURVNET

Pablo de Pedraza, Universidad de Salamanca (Spain)

Kea Tijdens, Erasmus University (Netherlands)

Stephanie Steinmetz, Erasmus University (Netherlands)

**Abstract** The Social Scientist community needs a multidisciplinary network to study web surveys methodology and its scientific validity.

The network aims to bring together survey methodologist, web surveys experts, social scientist and official survey institutes; synergize the knowledge of survey methodologists and web survey experts; develop guidelines for official bodies and statistical institutes; provide tools to take advance of technological changes; foster the development of new cross-national research proposals and a young researches network.

The network will contribute to the development of theoretical foundations of web survey methodology and innovative ideas. It will be able to offer advise to statistical offices and institutions aiming to equip with Internet for their data collection.

The network will organize academic conferences; establish working groups according to the scientific programme; enhance exchanges among researchers, publish findings and disseminate activities.

## 4.3 Interviewer and Respondent Behaviour in Survey Interviews

### Coordinators:

Wil Dijkstra, Vrije Universiteit Amsterdam (Netherlands)  
Yfke P. Ongena, University of Twente (Netherlands)

### 4.3.1 Analyzing interviewer behaviour in semi-standardized face-to-face interviews on suicidal ideation

Johannes van der Zouwen, Vrije Universiteit Amsterdam (Netherlands)  
Stasja Draisma, Vrije Universiteit Amsterdam (Netherlands)  
Merijn Eikelenboom, Vrije Universiteit Amsterdam (Netherlands)  
Johannes H. Smit, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** In a large-scale survey on mental health issues (N=2981), suicidal ideation has been measured with the five screening items from the Beck Suicide Ideation Scale.

Questioning about suicide ideation is not an easy task for an interviewer. Reporting about suicidal thoughts may be painful and threatening to respondents if such ideations are present. But it may also work in another way: questioning after suicide ideation may hamper rapport between interviewer and respondent when it becomes very clear, after asking the first screening questions, that suicide ideation is not present and the interviewer continues asking questions on this topic.

In order to ease this difficult interviewer task, interviewers were allowed to skip some questions, if they were convinced that they already had enough information to come up with a positive or negative score on the presence of suicidal ideation. Moreover, interviewers had some freedom to adjust the wording of the questions to what the respondent had already reported, implying loosing the standardization of interviewer behaviour so much striven for in survey designs.

In analyzing the response distributions of suicidal ideation we found proof of highly significant interviewer effects.

To investigate what particular type of interviewer behaviour contributed to these interviewer effects, transcripts of more than 500 interviews were analysed with a, for this purpose developed, coding scheme.

Coding was done with the aid of the Sequence Viewer software by trained and supervised coders. The inter-coder reliability was quite sufficient.

Using the search procedures of Sequence Viewer, i.e., looking for the occurrence of specific patterns of codes in the sequences, enabled the exploration of differences between interviewers in questioning techniques used. This detailed observation of the question- answer process made it possible to gain more insight in interviewer and respondent behaviour when involved in questioning such loaded topics.

### 4.3.2 What makes children tense in interviews? An Interactional Analysis of Semi-Standardised Interviews

Susanne Vogl, Katholische Universität Eichstätt-Ingolstadt (Germany)

**Abstract** Children are increasingly seen as competent informants when their activities, views and needs are concerned. But methodological aspects of researching children are rarely discussed. Therefore the possible adjustment of research instruments or whether such an undertaking is feasible at all needs full consideration. We lack competence on cognitive, interactive and verbal capacities of children in interview situations. Surveys in particular require this kind of previous knowledge so that interviewer behaviour is adjusted accordingly and we construct a questionnaire that can measure valid data. The interviewer and respondent behaviour with child interviewees is a very specific one. Not only there is an authoritarian relation but also skills of both parties are different. Tension and insecurity are signs of problems either in this interaction or with the content or research instrument. An in depth analysis of those occasions therefore promises the option of improved interviewer behaviour and of better research instruments.

For studying the interviewer and respondent behaviour 112 qualitative semi-structured interviews with 56 children aged 5 to 11 were analysed. Every child was interviewed twice, once face-to-face and the other time via telephone. All interactions in the interviews have been coded with the Interaction Process Analysis (IPA; Bales, 1950). By using the IPA instances where tension evolves on behalf of the child respondent can be easily spotted.



Also the interactional context can be qualified. This quantitative analysis is accompanied by a more qualitative and content related analysis that connects interaction with the specific task or content of a question within an interview. Causes and consequences of tension within the interview progress can be analysed. Interviewer behaviour, question format and content are taken into account for causes. Furthermore, how do children cope with insecurity and tension? Does it have an impact on the interview progress afterwards? Beyond this, age and interview mode are seen as factors that influence not only the total amount of interactions showing tension and insecurity but also causes, expression and coping-strategies.

This qualitative study is used as a vehicle for better understanding of answers in more standardised procedures. The added value in enhanced understanding is the option of an enhanced quality of survey methodology.

### 4.3.3 Interactive feedback can improve accuracy of responses in web surveys

Frederick G. Conrad, University of Michigan (United States)

Mick Couper, University of Michigan (United States)

Roger Tourangeau, University of Maryland (United States)

Mirta Galesic, Max Planck Institute for Human Development (Germany)

Ting Yan, University of Chicago (United States)

**Abstract** Because of its interactive character the web may promote more accurate survey data than do other modes. One way that interactivity can increase data quality is by providing feedback to respondents about their answers. For example, questions known as tally or constant sum items require that multiple answers sum to a fixed total, e.g. 24 hours or 100 percent; for these items a web questionnaire can inform the respondent about whether or not their individual answers add-up to the target value. Designers can provide this feedback when the respondent submits the entire set of responses or in a running tally that is incremented as the respondent enters each component answer. We report three experiments that investigate how this kind of feedback affects the quality of answers for tally items. In experiment 1 we found that if the computer displays a running tally this increased the number of totals equal to the target sum (100% of a respondents internet usage) relative to feedback that is delayed until after the total is submitted, which in turn led to more totals equaling the target than no feedback. While promising, this result really concerns the number of well-formed answers (i.e., those equal to the target) which is not necessarily the same as the number of accurate answers because some respondents may produce a well-formed answer by adjusting one or two component answers to make the total add to the target without improving the accuracy of the components. So in experiments 2 and 3 we asked whether feedback about the total also led to more accurate data. In experiment 2, respondents reported how long it had taken them to complete each section of the questionnaire they were just finishing and we compared this to their actual times; the difference between their self-reports and the server-recorded durations was smaller when a running tally was displayed than when it was not. In experiment 3 we asked respondents to report how much time they had spent on each of several activities in the previous day and compared their time use estimates to published population estimates. Again, the running tally led to smaller differences between the self- and published reports than when this feedback was not provided. Taken together, the studies suggest that the interactivity of the web can improve both the form and the accuracy of answers for this type of item. It is not possible to provide feedback like this with self-administered paper questionnaires and it is not easy to do so in interviews so this may be a case where web administration adds clear value over other modes of data collection.

### 4.3.4 Determinants and Consequences of Respondents' Strategy Choices when Answering Proxy Questions

Tobias Stark, University of Groningen (Netherlands)

Volker Stocké, University of Bamberg (Germany)

**Abstract** In recent years, it has become more common to ask survey respondents questions about characteristics of their peer group. The use of these so-called proxy questions increased due to the progress in ego-centric network analysis where the influence of a respondents peer group is investigated. However, validation studies have shown that answers to proxy questions are very often incorrect. Even though the use of proxy question has become a standard approach, the reasons for respondents mistakes have hardly been understood.

In this study, we develop hypotheses about the conditions under which incorrect answers to proxy questions are given based on Krosnick's theory of survey satisficing (Krosnick 1991, 1999). 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 the respondents motivation, their (cognitive) abilities and

the task difficulty. We empirically show that those determinants indeed explain which respondents give incorrect proxy answers. Furthermore, in contradiction to a common hypothesis, we show that the determinants do not moderate each others effects but rather are independent. This can have important consequences for research with proxy questions since it shows that it is not necessary for researchers to improve all three determinants in order to receive correct answers from respondents.

We test our hypotheses in a 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 correctly predicts the likelihood of correct survey responses. Respondents that are highly motivated (operationalized as personal relevance of the survey topic and importance of accurate answers), those with high abilities (closeness to alter, duration of relationship with alter, and discussion of the survey topic with alter), and those with high cognitive skills (highly educated and had to spend less effort to answer a survey) are more likely to give correct answers, while an increase in the task difficulty (questions about alters socio-demographic characteristics vs. questions about their attitudes) decreases this likelihood. Furthermore, insignificant interaction effects suggest that the determinants do not moderate each others effect. Instead, the determinants seem to be additively connected and can substitute each other. An additive index of the determinants is indeed positively related to the likelihood of correct proxy answers and negatively related to the use of the dont know option. We discuss to what degree the results can be translated to response generating processes in general.

#### 4.3.5 The use of behaviour coding for the convergence evaluation of proxies and direct informants responses in a disability survey.

Isabel Benítez, University of Granada (Spain)

Jose Luis Padilla, University of Granada (Spain)

Yfke P. Ongena, University of Twente (Netherlands)

**Abstract** The use of proxy informants in household surveys has increased the interest in analysing the agreement of their answers with the direct informants answers. Behaviour coding can provide useful additional information in the analysis of the proxy informants behaviour when answering to survey questions.

The aim of the present study was to evaluate the agreement between direct informants and proxies with additional information based on behaviour coding. The evaluation of the convergence between both types of informants was realized during the pretest of a questionnaire used in a disability survey. This questionnaire includes questions related to different limitations about daily life activities. 16 direct informants and 16 proxies answered to 11 "target questions about the same content for both types of informants. The agreement level was evaluated between the information provided for both types of informants. The influence of informants characteristics and questions characteristics on the convergence levels was also evaluated. The analyses were carried out using the program Sequence Viewer (Wil Dijkstra, version 4.4.a. 2008).

The results showed a 52% of agreement in the answers of the informants, being these answers different in 19% of the cases. The remaining 30 % concern other situations in which the answers of one of the informants or of both were inadequate (as detected by means of behaviour coding) and therefore the comparison between them was not possible. Differences were also found in the percentage of agreement between the different couples and between the questions of the questionnaire.

A possible explanation might be the difference in the level of salience for type of informant at the moment of answering, or the quantity of available information, which can facilitate or inhibit the work of translating their response in an alternative. Future research will be focused on the study of other factors which can influence the convergence between proxies and direct informants, as well as on the definition of the "ideal profile of the Proxy".

## 4.4 Methods and approaches for coding textual survey variables

### Coordinator:

Stefania Macchia, Statistics Italy (Italy)

### 4.4.1 Inconsistencies in reported job characteristics among employed stayers: An analysis on two-wave panels from the Italian Labour Force Survey, 1993-2003

Francesca Bassi, University of Padua (Italy)

Ugo Trivellato, University of Padua (Italy)

**Abstract** In recent years, labour markets in industrialised countries have shown quite a high degree of mobility. Extensive literature on the micro-dynamics of the labour market focuses on job-to-job flows, only a few papers explore the kind of job changes that workers experienced. The complexity of job mobility also demands analysing these changes while changing employer. The literature on job matching suggests that a significant number of workers who switch job also change employment characteristics, mainly industry and occupation.

Some studies show that job characteristics, particularly industry and occupation, collected in surveys are affected by measurement error. The effect of these errors is to exaggerate the occurrence of changes in such characteristics, at least when information is obtained at two points in time with independent interviews. Other studies demonstrate that, in general, industry is reported more accurately than occupation and that the agreement rate between employees and employers reports, classified according to a single-digit coding scheme, is higher than that resulting when reports are categorised according to a more detailed classification.

In this paper we deal with measurement error, and its potentially distorting role, in information on industry and professional status. As a case-study we consider two-wave panels one year apart collected by the Italian Quarterly Labour Force Survey (QLFS) in the period from April 1993 to April 2003.

The QLFS refers to a sample of the resident non-institutional population. Thus, it collects information on job characteristics of (almost) all the employed. The survey is cross-sectional with a 2-2-2 rotating design, which yields two-wave panels one quarter and one year apart. The focus of our analyses is on inconsistent information on employment characteristics industry and professional status resulting from yearly transition matrices for workers who reported that they were continuously employed over the year and did not change job.

First, we compute and comment upon some usual indicators of disagreement. We find clear evidence that there is sizable measurement error in both industry and professional status, and that industry is reported more accurately than professional status.

Then, we test whether the consistency of repeated information significantly increases when the number of categories is collapsed. As regards the detail of variable classification for cross-section estimates, Istat the Italian statistical agency provides the following indications. For professional status, a reliable classification reduces to a binary one: Employee and Self-employed. For industry, Istat asserts as dependable a classification in 12 categories. By means of the hierarchical Kappa coefficient, we obtain evidence that supports the first indication by Istat, but casts severe doubts on the second.

We further explore the patterns of inconsistencies among categories of variables by testing several specifications of Goodmans quasi-independence model, which is almost always rejected.

Lastly, we consider and compare alternative 4-category classifications obtained by collapsing professional status and industry into a single variable. The standard classification labels respondents as Self-employed, Employee in agriculture, Employee in industrial sector, and Employee in services. As an alternative, another 4-category classification was recently used in the study of worker turnover, with the four categories given by Self-employed, Employee in agriculture, Employee in industrial sector and private services, and Employee in Public Administration and social services. Interestingly enough, the latter classification turns out to be almost uniformly better than the former, standard one.

### 4.4.2 From texts to codes. Multivariate exploratory methods applied to open-ended questions

Anne-Sophie Cousteaux, Sciences Po (France)

**Abstract** Even if questionnaire designers introduce open-ended questions in their surveys, answers to open-ended questions remain too often underexploited. Indeed, statistical analysis require substantial efforts to code textual data using a limited number of pre-defined categories. The traditional manual technique, which consists in identifying the main themes in the textual corpus, appears often too arbitrary and subjective to quantitative researchers. At the same time, the shorter and necessarily more stereotyped responses collected from open-ended questions rather than from in-depth interviews cannot obviously reach the same degree of complexity. These reasons lead to neglecting open-ended questions. Yet combining qualitative and quantitative approaches could take advantage of the detailed responses and of the summary produced by statistical methods.

From the example of lay perceptions of health, this paper sets out to present how multivariate exploratory methods, namely correspondence analysis and cluster analysis, can successfully be applied to textual data to produce a standardized classification of the individual responses. The data used here come from the enquete permanente sur les conditions de vie des menages carried out by the French National Institute for Statistics and Economic studies (INSEE). In 2001, the rotating module focused on health and included an open-ended question: What does being healthy mean to you? (*Etre en bonne sante, quest-ce que cela signifie pour vous?*). Because of its size of 5000 individuals and its representative sample, this survey offers great possibilities to know the various perceptions of health in the whole population, to quantify their respective importance and to describe them by individual social and medical characteristics.

The aim of this paper is to discuss the advantages and the limits of the transition from texts to codes based on statistical methods. Turning the individual responses into a typology of health perceptions is not straightforward and fully carried out by the SPAD statistical package. At each stage, from grouping words together to producing the final typology by cluster analysis, the researcher is faced with subjective choices, which could be decisive in terms of conclusions but are also essential for making sense of textual data.

The statistical methods used (correspondence analysis then cluster analysis) aim at identifying several subgroups of individuals formed only on the basis of the words they used to define good health. From the 1409 distinct words grouped together under 120 lemmas, the cluster analysis reveals finally ten categories: healthy lifestyle, energy, absence of serious illness, well-being, never go to the doctors, absence of medicine, absence of pain, being happy and leading a normal life, capacity, dependency which seem sharply affected by biological and social age, i.e. the deterioration of health with age and the life cycle. In conclusion, these findings support the application of multivariate exploratory methods to textual data as a computer-aided coding which produces less subjective categorizations and reduces treatment costs of open-ended questions.

### 4.4.3 Analyzing open-ended questions by means of text analysis procedures

Roel Popping, University of Groningen (Netherlands)

**Abstract** Assume one has open-ended questions in a survey and seriously wants to analyse the answers to these questions. Now text analysis might be applied. This talk discusses a number of choices to be made when a thematic text analysis is to be applied. It starts with requirements to be posed to the open-ended questions themselves and sketches choices in the development of a category system. Here the coding comes immediately into view. Coding can be performed from an instrumental or a representational perspective. In the first the coding is performed from the point of view of the investigator, it can be performed in a run of a computer program. In the second the point of view of the respondent is acknowledged. Now the computer can be used as a management tool, but the coding itself must be performed by a human coder. The choice for one of these methods depends on what the investigator is looking for and has consequences for the way how to proceed. When the representational way of coding is applied also questions about intercoder reliability must be posed. The talk is illustrated by examples mainly based on data from Dutch election studies.

## 4.5 Methods of Evaluating the Quality of Data Collection in Surveys

### Coordinators:

Emanuela Sala, University of Essex (United Kingdom)

Noah Uhrig, University of Essex (United Kingdom)

### 4.5.1 A Framework for Understanding the Quality of Survey Data

Noah Uhrig, University of Essex (United Kingdom)

Emanuela Sala, University of Essex (United Kingdom)

**Abstract** The quality of survey data is of concern to social scientists and policy makers alike. It is determined by many aspects of the survey process including sampling, patterns of response and non-response as well as survey design and data collection procedures. To provide an overview of our session, we begin by comparing and contrasting cognitive interviewing, behaviour coding, and traditional pilots used to evaluate aspect of survey and question design in order to expose their underlying assumptions about the sources of survey data quality. Drawing on examples from the British Household Panel Study, we expose how each method can lead to inconsistent conclusions about data quality. We offer examples of ways the strengths of various evaluative methods can be combined to better inform survey design decisions. Our session is then comprised of four papers. The two initial papers provide new models for assessing total survey quality. The remaining two contain examinations of data quality post fieldwork.

### 4.5.2 Towards a Taxonomy for the classification of surveys

Tao Chen, Napier University Business School (United Kingdom)

Robert Raeside, Napier University Business School (United Kingdom)

**Abstract** Research has been under taken to improve the quality and effectiveness of surveys and from reviewing this work a model to assess the quality of surveys has been developed. The model has been applied to over 100 national and international surveys. This generates data on formulation, design and application quality. These data have been used along with attributes of the surveys examined to create a database. Various measures of the survey were taken such as average question length, number of questions, number of citations in papers were taken and input in to the database. This database has been analysed using tools of statistical classification. This has allowed the identification of structural features of surveys which are associated with the quality of the survey. It is these features which are reported on in this paper. The paper is concluded with a discussion as to how the approach of six sigma can be used to improve the quality of surveys. Recommendations are made as to how a systematic approach can be taken to improve survey quality and effectiveness.

### 4.5.3 Do Survey Respondents Give the Same Answers When Their Spouse is Present?

Arnaud Régnier-Loilier, Institut National d'Etudes Démographiques (France)

**Abstract** The quality of survey data depends on the way in which information is collected. To limit bias, the interviewers are instructed to ensure that the interviews take place in private without anyone else present. However, they cannot impose the interview conditions and in many cases one or more outsiders are present during all or part of the interview. To evaluate bias on respondents' answers, the interviewer indicates at the end of the interview if a third person was present, if they intervened and if, in the interviewer's opinion, this presence influenced the respondent's answers. Though it is essential to know if a third person was present, it is more problematic to expect the interviewer to assess the resulting bias. Is he or she really capable of doing so?

These reservations, combined with the constraints of survey design, led to the adoption of a different strategy for the first wave of the *\_Generations and Gender Survey\_* conducted in France by INED and INSEE in 2005. Rather than obtaining retrospective information (at the end of the survey), factual indicators of the presence of third persons were included at different points in the questionnaire. The interviewer was asked to answer the

following question: "Interviewer: Are there, at this point in the interview, any other persons present in addition to the respondent? Yes / No" and, if yes "Please indicate who is now present: Spouse / Children / Brothers and Sisters / Friends / Other". A total of five indicators are included in the course of the questionnaire.

This approach gives a more precise description of interview conditions than do other surveys, and provides a means to detect comings and goings (persons present at the beginning, but not at the end, or the arrival of someone during the interview). Although one quarter of interviews were affected by the presence of a third person at the beginning or the end, the proportion rose to one-third when these comings and goings were taken into account. The third person was usually the respondent's spouse. It is important to have a precise description of the interview conditions because of the reporting biases that can result. The example of the breakdown of parental tasks within the couple tends to indicate that the presence of the spouse exercises a certain control over the respondents answers, while his or her absence encourages the respondent to "idealize" by presenting a more even breakdown, more in keeping with a certain egalitarian standard. The presence of the spouse also seems to encourage recall when the respondent is trying to date certain events in their conjugal life. However, although these results seem to indicate that, contrary to the instructions given to interviewers, it is preferable for interviews to take place in company of the spouse, the presence of a third party particularly that of the spouse means that past conjugal histories are reported less fully, especially on the male side.

#### 4.5.4 The impact of forgiving wording and question context on social desirability bias in sensitive surveys

Anatol-Fiete Näher, University of Leipzig (Germany)

Ivar Krumpal, University of Leipzig (Germany)

**Abstract** Sensitive behavioral questions in surveys often result in self-reports which are distorted by social desirability bias. Interviewees underreport socially undesirable behavior and overreport socially desirable activities. Such systematic measurement error in turn generates erroneous prevalence estimates of the behavior in question. Standard literature on survey methodology often recommends (1) positive loading of sensitive questions, e.g. using forgiving wording, or (2) choosing a permissive question context to encourage interviewees to answer more honestly. However, only few attempts to systematically validate these recommendations can be found in the experimental literature. Based on theories of cognitive dissonance (Festinger 1957; Aronson 1999) and the inclusion-/exclusion-model (Schwarz & Bless 1992, 2007), we derive explanations how manipulations of question wording and context could elicit more socially undesirable answers in sensitive surveys. In an experimental online survey (N=1176, 4 splits), we evaluate the effects of (1) forgiving-wording and (2) question context (permissive versus restrictive) on social desirability bias in different sensitive behavioral questions. Consistent with former experimental findings (Catania et al. 1996; Holtgraves et al. 1997; Presser 1990; Tourangeau & Smith 1996) the empirical evidence of the predicted effects is mixed. Thus, the assumed social desirability bias-reducing effect of forgiving wording and permissive question context respectively should not be taken for granted.

## 4.6 Questionnaire design in panel surveys

### Coordinator:

Peter Lugtig, University of Utrecht (Netherlands)

### 4.6.1 Tracing life courses with prospective Panel surveys - the life history calendar of the pairfam German Family Panel Study

Laura Castiglioni, University of Mannheim (Germany)

Ulrich Krieger, University of Mannheim (Germany)

Volker Ludwig, University of Mannheim (Germany)

Klaus Pforr, University of Mannheim (Germany)

**Abstract** Collecting valid information on the occurrence, timing and spacing of events during the life course is a crucial issue for applied research. A prospective panel design can help to improve data quality enormously. Within this framework, it is possible to address the major problem respondents are typically confronted with when asked about what happened in their life: recalling events accurately over a longer time span. Using calendar techniques eases the process of information retrieval to some extent. However, gathering event history data on shorter periods at several points during the life course still produces the problem of spurious errors at the time of the interviews, also known as seam effects.

In this presentation we argue that measurement error can be reduced considerably by combining life history calendar and dependent interviewing techniques. Drawing on pretest data from the recently started German Family Panel Study, we provide evidence on the power of this approach in reducing the seam effect. We report results from a controlled experiment about implementation of dependent interviewing in a graphical life history calendar.

In wave two of the pretest study, the "Mini Panel", we administered a paper-and-pencil retrospective life history calendar from the age of 14 up until the date of the interview, covering a wide range of topics (relationships, children, employment, education and residential mobility). In wave three, we then updated the calendar with the events that occurred in between interviews. Using a split-ballot design we either enriched the calendar by the information of the last known state in the domains or started the interview with a blank calendar. Results will reveal how dependent interviewing helps to reduce the seam effect and thus improve data quality. Finally, an outlook to the computerized, graphical and interactive version of the life history calendar of the main Study, the pairfam German Family Panel Study, will be given.

### 4.6.2 Standardized recall aids for online life course surveys

Tina Glasner, University of Wageningen (Netherlands)

**Abstract** The main aim of our study was to develop an effective and cost-efficient calendar recall aid for a retrospective questionnaire to be used in a large online panel. Calendar instruments, such as the Event History Calendar, have been shown to improve data quality in face-to-face and telephone surveys, ensuring higher consistency, completeness, and accuracy of retrospective reports. They combine two important aided recall techniques, namely temporal bounding cues (or landmark events) and visual feedback of the respondents answers (i.e. domain timelines). Using an experimental approach we investigate which of these components contribute the most to the calendars positive effects on data quality. We optimize the calendar method by identifying its most effective elements and eliminating redundant or ineffective cues, while taking into account the requirements of online survey methods in terms of clarity and user-friendliness.

In our paper, we present the results of a split-ballot experiment in which we tested three versions of a calendar recall aid and compared them to a regular standardized questionnaire. Approximately 2000 participants of an existing online household panel were randomly assigned to four conditions. Respondents in condition 1 (control) filled in a regular online retrospective questionnaire. Respondents in the experimental conditions filled in the same questionnaire with either visual feedback but no public or personalized landmarks (condition 2: visual feedback), a timeline with public and personal landmarks (condition 3: landmarks only), or a standardized questionnaire with both, visual feedback and landmark events from the reference period (condition 4: full calendar). We compare the data from the four conditions with regard to the total numbers of reported transitions, completion time, number of self-corrections, refusal and break-off rates, and respondent evaluations of the method.

### 4.6.3 Improving retrospective life course data by combining modularized self-reports and event history calendars

Katrin Drasch, Institute for Employment Research (Germany)

Britta Matthes, Institute for Employment Research (Germany)

**Abstract** Event history calendars (EHC) have proven to be a useful tool for collecting retrospective autobiographic life course data. One problem is that they are only standardized to some extent. This limits their applicability in large-scale surveys. However, in such surveys a modularized retrospective CATI design can be combined with EHC as a data revision module by using insights from cognitive psychology. The data revision module stimulates the respondent's memory retrieval by detecting temporal inconsistencies such as gaps as well as overlapping and parallel events. The interviewers are guided through the editing process with the help of an automated questioning routine. However, to guarantee flexibility probing for the purpose of correction is possible.

This approach has been followed in the ALWA study (Work and Learning in a Changing World) a large-scale representative telephone survey with 10,000 respondents conducted in Germany in 2007/2008. In this study, a data revision module was implemented which merged complete general and vocational education histories as well as (un)employment histories and times of parental leave. Then, appearing inconsistencies were resolved during the interview. We investigate to what extent this data revision module improves time consistency and dating accuracy of individual reports by comparing the uncorrected data with the final data after revision. Especially unemployment episodes were likely to be omitted without the inclusion of the data revision module. Concerning dating accuracy all types of episodes are corrected to a similar extent. Additionally, the time-lag of the episode to the interview data has an influence on completeness and dating accuracy. Also, individual characteristics matter: women have a lower chance to add events; the same is true for younger respondents. In sum, we can conclude that this module is a powerful tool to improve time consistency and dating accuracy in retrospective surveys.

### 4.6.4 Evaluating Data Quality on the Measurement of Change using Dependent Interviewing

Peter Lugtig, University of Utrecht (Netherlands)

**Abstract** Dependent Interviewing is the process of reminding respondents of their earlier answer in panel surveys. In this paper we examine two Dependent Interviewing designs that are frequently used and show how the relative stability of the variable and the reliability of its measurement both contribute to the choice of the best Dependent Interviewing technique possible. Our data consist of income reports of respondents from a four-wave panel study conducted in the Netherlands. We manipulate the income questions using a 3x2 factorial design (three Dependent Interviewing conditions with two types of questions) The effects of our experiment on data quality are modelled with a quasi-simplex structure to enable the division of variances into measurement error and real change. Our main conclusion is that the use of Proactive Dependent Interviewing is problematic when the reliability of the variable of interest is low. We finally discuss our results and suggest guidelines for the design of Dependent Interviewing procedures.

### 4.6.5 The Impact of Dependent Interviewing on Measurement Error

Annette Jäckle, University of Essex (United Kingdom)

**Abstract** Repeated measures data are used, amongst other things, to estimate and predict the extent of change experienced by individuals, of time spent in particular states, or to control for unobserved individual effects in explanatory models. Repeated measures data from panel surveys, where sample units are interviewed at regular intervals, are affected by measurement error at each interview. The causes of measurement error are similar to those in cross-sectional surveys. The impact on estimates may however be worse. To improve the quality of panel data, survey organisations are increasingly using dependent interviewing (DI) techniques, whereby information about the sample member from prior interviews is incorporated into the question wording and routing.

This article assesses in which way, and to what extent, DI alters the distributional properties of measurement error in repeated measures data. The study examines reports on the timing and amount of various income sources, collected in the context of the British Household Panel Survey. The survey included a randomized experiment comparing different methods of independent and dependent interviewing and obtained validation data for respondents from government administrative records.



Previous validation studies have focused on the impact of DI on measurement error in a single wave. The results indicate that DI reduces under-reporting and hence reduces bias in cross-sectional estimates of prevalence. The impact on time series properties of errors across panel waves have however been ignored so far. Previous studies have also shown that DI increases the consistency of reports across waves, thereby reducing spurious change. Whether this is achieved by reducing errors, or merely by replacing random error at each wave with systematic (correlated) error, is however unknown.

This study firstly tests assumptions about the distributional properties of measurement error, typically made in the analysis of repeated measures data. Independent and dependent interviewing data are compared, to assess how and to what extent DI changes the properties of errors. Secondly, to illustrate the implications for actual estimates, the biases in exemplary analyses of repeated measures data from independent and dependent interviewing data are compared.

## 4.7 Questionnaire Developing and Testing

### Coordinators:

Stefania Macchia, Statistics Italy (Italy)  
Manuela Murgia, Statistics Italy (Italy)

### 4.7.1 The Force of the Forced Choice

Gabriella Fazzi, Sapienza Università di Roma (Italy)  
Fabrizio Martire, Sapienza Università di Roma (Italy)  
Rita Pavsic, Università degli Studi di Catania (Italy)  
Maria Concetta Pitrone, Sapienza Università di Roma (Italy)

**Abstract** Surveying attitudes, the subjects answer does not reflect only his actual status, but also: 1) the perception he has of it; 2) his awareness of having a status; 3) reactions to the instrument, affected also by the ability of using it. As a matter of fact, such ability is related to both the subjects specificities (especially education) and the structure of the instrument itself.

The most used techniques, which are usually also the most analysed through specific methodological researches, are the ones implying the use of point items sentences expressed as an assertion, a question, or an object description whose formulation clarify only one aspect of the issue investigated through the overall question. Beside the fact that such an instrument is considered affected by several kinds of distortion, a large amount of scientific literature has also shown that it often stimulates mechanical and/or acquiescent answers.

On the contrary, the technique known as the Forced Choice (FC) seems to be an interesting instrument, able to bypass successfully such a problems. In particular, we think that the accurately designed FC, even presenting some difficulties, allows to avoid the acquiescence risk as well as other common distortions.

To support this assumption, in the paper some research findings are presented. In particular, we will compare the FC with one of the most common technique as the Likert scale, showing the FCs efficacy in the face-to-face interviews, in the web surveys, and in the chat interviews. According to the empirical evidence, the research support the assumption that the indubitable design complexity of the FC, in this work, is largely paid off by its ability to provide more reliable answers and quasi-cardinal variables, which allow complex factorial analysis.

### 4.7.2 Using a Questionnaire Appraisal System to Improve Cross-National Questionnaire Design

Sally Widdop, City University London (United Kingdom)

**Abstract** Despite the range of pre-testing and development tools that can be used to design questionnaires, errors can often still occur leading to low measurement quality in the final items that are fielded. These errors are more problematic in cross-national surveys because of an increase in the range of things that can go wrong and reduce equivalence both within and between countries.

By adapting and building on the Question Appraisal Systems (QAS) developed by Willis and Lessler (1999) and Dean et al. (2007), a QAS is currently being developed for use during the question design process of the European Social Survey (ESS). The QAS is a structured questionnaire that acts as a check-list for assessing potential problems with draft versions of questions.

In every round of the ESS, two 50 item rotating modules are included in the questionnaire. These focus on a specific topic and are designed by multi-national teams following a European-wide competition. Members of the Central Coordinating Team of the ESS and individuals from the successful team are involved in the design and development process for the modules. This combines expert review, country consultation and a detailed set of translation procedures with statistical modelling techniques such as multi-trait, multi-method experiments (MTMM). Based on the results of this process, individual items are adopted, re-designed or discarded.

The QAS being developed for the ESS will be split into several parts, which will be used at different stages of the design process that is already in place. In addition, the cross-cultural elements from QAS-04 designed by Dean et al. will be adapted for cross-national implementation.

This paper will assess the contribution made by the QAS to the question design process used on the ESS and will seek to answer the following key questions:

- 1 Did using the QAS help to identify problems that were not identified by unaided expert review?

- 2 Did unaided expert review identify problems that were not identified by the QAS?
- 3 Was using the QAS helpful in producing solutions to the problems identified compared to the expert review process?
- 4 Was it helpful to divide the QAS into several parts and to apply these to different stages of the question design process?
- 5 Was the QAS useful at identifying problems specifically associated with cross-national research?

### 4.7.3 Meeting new challenges in the development of the 2009 Rehearsal Census Coverage Survey questionnaire for England and Wales

Sue Dalton, Office for National Statistics (United Kingdom)

Ruth Wallis, Office for National Statistics (United Kingdom)

**Abstract** Social change since 2001 and the need to capture more information about short-term migrants have led to new question content and changes to the enumeration base for the 2011 UK Census. These changes have impacted on the data collection requirements for the Census Coverage Survey (CCS) which will take place approximately six weeks after the Census. The 2011 CCS will need a questionnaire that can accommodate the question content changes, and gather basic data on visitors and short term migrants, as well as usual residents. In addition, the 2011 CCS will need to collect data about the previous address of households who have moved since Census day. This has presented the challenge of designing a questionnaire that is flexible enough to collect data about the current address of continuously resident respondents as well as data about the previous address from households who have moved. The CCS is an interviewer administered survey, involving a short doorstep interview. Thus any additional requirements need to be incorporated within this basic design. This paper reports on the iterative development of the UK 2009 Rehearsal CCS questionnaire to meet these challenges. In particular, it will detail how the challenges associated with the additional requirements were assessed via three waves of cognitive interviewing and ongoing consultation and collaboration with data analysis and field operation teams.

### 4.7.4 QDDS III: A tool for documenting survey questionnaires for researchers and data archives

Rainer Schnell, University of Duisburg-Essen (Germany)

Anja Zwingenberger, University of Duisburg-Essen (Germany)

Max Stempfhuber, GESIS - Leibniz Institute for the Social Sciences (Germany)

Oliver Hopt, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** The paper describes a new software system for the documentation of the development process of the questionnaire. Furthermore, it allows searching for formal and structural aspects of questions in social science data archives.

During the development of a questionnaire dozens of versions of the instruments are generated and subsequently discarded. Usually, only the final version of the questionnaire is documented. Due to this policy, details of previous versions and the rationale behind the final details in phrasing and layout is lost. The knowledge generated during the revisions can not be used for improving future research.

We developed QDDS III as a freely available Java-program. The system allows the electronic documentation of the development of Paper&Pencil, CATI and CAPI questionnaires. The program conforms to the DDI-standard of data archives.

Finally, QDDS III allows the search and retrieval of questionnaires and survey meta-data for large survey data archives. Currently about 50 widely used surveys are documented within QDDS III. The program is unique in its ability for searching according to formal characteristics of questions. For example, a data archive repository containing hundreds of surveys can be used to find all those surveys, in which a battery of dichotomous questions is followed by an open question. Such queries are useful for the methodological inspired re-analysis of survey data.

### 4.7.5 Design and Composition of Questionnaires: Techniques for Response Behavior by Extremely Long Instruments

Gediminas Merkys, Kaunas University of Technology (Lithuania)

Ruta Braziene, Kaunas University of Technology (Lithuania)

**Abstract** Research problem and methodological context of the paper. There are criteria of traditional empirical social research methodological quality: objectivity, representativeness, reliability, validity, efficiency and utility. They are also valid for survey research. In this paper we supposing that existing criteria could be reasonably complemented by the following criteria: variety and substantiality of inquiring indicators.

Psycho-social reality is complicated, multidimensional, contextual, stochastic, and constructive. Consequently research, that is operating by the small number of variables, extracted from the complex reality could not provide about such a reality anything eloquent. As an exception could be considered experimental or confirmational studies, founded theoretically and acting in previously explored area.

Long questionnaire as a trait of quality eventually could become anti-quality. Long questionnaires are difficult in filling because of tiredness of respondents, de-motivation, formal filling, low questionnaires reciprocity rate and etc. Considerable part of discarded questionnaires is an important factor and should be taken into account. On a whole as a result: unreliable and non valid research.

High number of primary items leads towards to long survey instrument. This paper presents methodological advices by filling extremely long questionnaires, based on experience of empirical studies. The most important of them are the following: research on quality of life and municipality provided services, work organizations personnel research and dissertational projects. Experience about construction of extremely long questionnaires is collected by working in very diverse populations (in common residents population and in such target populations as: school children, teachers, students, prisoners, and staff of very diverse work organizations).

There are a questionnaire construction, composition means and techniques, based on specially organized graphical design. By using extremely long questionnaires these methods and means help increase respondents motivation, attitudes and quality of the survey. The samples of conducted studies are conditionally high from 500 till 2000 and more respondents (conventional definition: long questionnaire starts from 200 primary items). Despite application of extremely long questionnaires, very high rate questionnaire reciprocity 80-98% is achieved. Based on our experience we created a framework of graphical composition. This framework involves the main points of application of graphical design: questionnaires cover list shaping and motivating instruction, application of graphical symbols, metaphors and analogies, replicated anonymous survey of the same respondents, application of principle of matrixes for construction of questions and answers, elimination of excessive information and informational noise, the harmony of syntax and graphical means, creation of compositional rhythmic and use of relaxing pictures.



## **Chapter 5**

# **Adding Data to Surveys**

## 5.1 Connecting data from independent surveys

### Coordinators:

Jacques Hagenaars, Tilburg University (Netherlands)

Ruud Luijkx, Tilburg University (Netherlands)

Ellen Verbakel, Tilburg University (Netherlands)

### 5.1.1 Enriching surveys by surveys

Jacques Hagenaars, Tilburg University (Netherlands)

Ruud Luijkx, Tilburg University (Netherlands)

Ellen Verbakel, Tilburg University (Netherlands)

**Abstract** The data from past and present day sample surveys are made more and more easily accessible and better and better documented by the great Social Science Archives. Together they provide an enormous amount of potentially useful information on social life that may be at least as important as the secondary data sources from formal (governmental) registers. The connection of independent surveys can potentially increase our knowledge compared to a single survey, thereby stimulating new research without having costs for a new data collection. The main uses of these survey data for social science research will be discussed, along with a number of methodological difficulties the secondary researcher may encounter. More specifically, this presentation will outline three possible ways of how linking independent surveys can be used to answer substantial research questions in the social sciences: (a) linking surveys for analyses at the aggregate (regional, cohort etc) level, (b) connecting surveys to incorporate contextual effects in individual-level analyses, and (c) combining surveys to estimate individual relationships that are not available in a single data source.

### 5.1.2 Statistical matching based on multiple imputation: Outline of the procedure and a substantive illustration

Bojan Todosijevic, University of Twente (Netherlands)

**Abstract** Given two methodologically similar surveys, a question not asked in one survey could be seen as a special case of the missing data problem. Hence, the transfer of data across data sets, or statistical matching, could be achieved applying the procedures for Bayesian multiple imputation of missing values.

This paper presents a model for transferring data between different datasets based on multiple imputation (MI) approach. The first part of the paper demonstrates the application of the MI procedure, and outlines its main characteristics by comparing the imputed and actual respondents responses. The second part of the paper presents a substantive research example focused on the moderating influence of group identification onto the relationship between authoritarianism and ethnocentric attitudes. The exercises are based on the integrated analysis of two Dutch studies: Social and Cultural Developments in The Netherlands (SOCON 2000) and the Dutch Election Study (NKO 2002). The results show that statistical matching based on MI principles can be a useful research tool. The entire enterprise is finally interpreted in the sense of taking the previous research into account seriously.

Key words: statistical matching, multiple imputation, Dutch Election Study

### 5.1.3 International analysis of student achievement and labour market outcomes based on data combined from different surveys

Maciej Jakubowski, University of Warsaw (Poland)

**Abstract** Student literacy surveys are measuring student achievement in scholastic domains. PISA international survey is conducted every 3 years since 2000. The survey collects 15-year-old student literacy test scores for all OECD and numerous non-OECD countries. PISA is supposed to measure skills needed on the labour market, however, no direct test of such presumptions is possible. The paper relates student test scores to income data trying to assess whether there are any visible similarities. We try to confirm that literacy as measured by PISA is able to predict labour market outcomes. Different sources of income data are employed including United Nations databases with income inequality statistics and EU-SILC individual datasets. The focus, however, is

not on comparing country means in literacy and income, but in relating to each other whole distributions and statistics describing their shape. The paper discusses results of such comparative analysis and methodological problems which have to be addressed. The most basic problems are related to differences in variable coding and scaling of background variables which could be potentially used to adjust samples making them directly comparable. More general problems are also discussed, which in this case are mainly associated with discrepancies in complex sampling designs and validity of linking samples of dissimilar populations collected using distinct survey methodologies. Robustness and reliability of different approaches to overcome these problems and doubts are discussed. Finally, the results of comparative analysis are presented. These are based on simple comparisons of basic statistics measuring score and income levels and dispersion, but also on regression and semi-parametric methods of adjusting and comparing two distributions (see Tarozzi, 2006). Main results are in line with intuitions and partially confirm validity of the approach. However, some findings are counter intuitive and their interpretation is uncertain. Examples given make clear that results based on linked datasets are less robust to methodological criticism. In some cases it cannot be assured that they are not artifacts caused by flaws in linking methodology, rather than real phenomena revealed by researchers.

### 5.1.4 Normative context as a determinant for labour market participation

Ellen Verbakel, Tilburg University (Netherlands)

Ruud Luijckx, Tilburg University (Netherlands)

**Abstract** Decisions on labour market participation and working hours are often supposed to be related to norms and values about gender roles and the division of (paid) labour. The shift from traditional values, which emphasize the household and caring role of women and the breadwinner role of men, to modern values, which emphasize an equal division of labour, has often been proposed as an explanation for the huge increase in female labour market participation, but also for the recent increase in part-time working fathers. Not only may values held in ones social context affect working hours of individual men and women, they may also play a role in the way spouses influence each others careers. Labour market careers of husbands and wives are supposed to be related since they need to arrange their careers in such a way that there is enough income and time available in the household. In an earlier study, I have hypothesized that the normative context affects the way spouses influence each others careers. In times that the traditional norms were widely embraced, wives were likely to stop working as soon as the couple could afford it. This results in a negative association between husbands resources and wifes labour market participation. The opposite is expected in times when norms have become less traditional. Then, husbands will try to support their wives careers, and it will be the husbands with most resources who will be most successful in this. As a result, in modern times, there is supposedly a positive association between resources of the husband and his wifes labour market participation.

Empirical tests for the hypotheses that the normative context affects working hours of men and women and the way spouses influence each others careers have either been completely absent (i.e. the observed trends are not tried to be explained) or have used indicators for normative context on the country level. Such a measure neglects the fact that much variety in values exists within a country, even at a single point in time, and it seems plausible that people adhere more strongly to norms in their direct environment than to the average norm held by the population in general. In this study, I will investigate the influence of the normative context that is more directly relevant for couples decisions: those of the people in the same birth cohort and with the educational level. These are supposedly the groups people can identify with without having the problem of endogeneity, which is the case if we would consider the norms of friends or of the region people live in. The research question runs as follows: To what extent does the normative context, defined as the people in the same birth cohort and with the same educational level, affect (a) working hours adjustments of men and women, and (b) the influence of the partners resources on working hours adjustments?

The Family Surveys Dutch Population has a retrospective design that provides us with complete labour market, relational and fertility careers of both partners, including the exact timing of events. This feature offers a unique opportunity to investigate transitions in the labour market careers within couples. Information on the normative context is lacking, however. To overcome this problem, I will aggregate information on values for each cohort x education group that will be derived from the European Values Study (waves in 1981, 1990, 1999/2000, and 2008) and SOCON (waves in 1979, 1985, 1990, 1995, and 2000). Average values on working mothers and gender roles will be calculated for five-year periods. I will study transitions that take place between 1980 and 2003. Event history analysis is used to estimate the influence of the normative context on (a) the probability of a change in labour market participation (entry, exit, more hours, fewer hours) of men and women, and (b) on the effect of partners resources on this probability. Implications that arise from the merging of several datasets, such as comparability of samples, timing of data collection, formulation of questions, will be discussed.



## 5.2 Media-reported events: the context of surveys

### Coordinator:

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

### 5.2.1 Event data and social surveys

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

**Abstract** One of the reasons why event data are collected alongside the European Social Survey is that it was expected that, for example, a question about the subjective interest in politics of a respondent may well be answered differently at the height of a national campaign for a general election compared to a time when no election is imminent. This could be important because general elections are national events that occur periodically and not at the same time in different countries. Therefore, they could influence survey outcomes in a way which is not uniform across the countries in the ESS and over time. As the ESS will in the long run become an important asset for historical micro analysis information on elections and other major events has to be available. As a consequence, from the beginning an event data base has been built in the ESS.

After three rounds of survey data collection, it can be analysed whether political interest is influenced by the presence of national parliamentary elections. This has been checked not by using data from the event database but by directly by relating political interest to the date of the elections. Political interest has been defined here as the share of media use devoted to news, politics and current affairs. The presentation will show that the national elections do not result in a clearly observable shift towards more attention for news and politics. A number of factors behind this will be discussed.

### 5.2.2 Identification of Societal Events

Cornelia Zuell, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Societal events are events like elections, significant changes in laws, demonstrations, but also extreme weather conditions. All such events can have an effect on a society and, consequently, influence the attitudes of its population. The assumption for surveys is that respondent behavior or answers to some questions are influenced by significant societal events in various areas. When conducting a survey, the impact of an event must be considered and, whenever possible, controlled especially in survey projects in which different countries participate. However, manual identification of significant events that occur during the data collection phase is a very error-prone and time-consuming task. Therefore, we have developed a procedure to identify events using a combination of two different approaches of the (quantitative) computer-assisted content analysis: the reference text technique and the statistical association approach. On the basis of distinctive features of word usage in a so called reference text corpus and in newspaper texts of a specific time period in which events should be located, words are selected and classified by means of an exploratory factor analysis. The identified factors used as indicators for societal events. Our procedure offers a new opportunity for research in survey-based practice. I will demonstrate this procedure by means of an example in which we will identify events automatically in a specific time period. As a final point, I will discuss the advantages and the restrictions of such an automatic technique to identify events. Considerations regarding the composition of the text base, language specific problems (for example, lemmatization, or the role of verbs), the selection of the final events (for example, number of events per week, weeks with no events, etc.) will be presented.

### 5.2.3 Event Data analysis: relating events with respondent's attitudes and opinions in three ESS' rounds (2002, 2004, 2006)

Analia Torres, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Paula Jerónimo, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

**Abstract** Event Data collection represents an added value on trying to explain differences among countries and changes in respondents opinions and attitudes over time. Political events like elections, the unpopularity of a specific government in a specific country, corruption processes in politics and economy and passionate debates over law changes in certain moral aspects (gay marriage, abortion, divorce) are frequently captured in the media event reports and may help to explain respondents answers.

In this paper, analysing results from the three rounds (2002, 2004 and 2006) we compare some major political events like elections reported in the media, and other debates related to moral values with changes in respondents answers comparing Portugal with all other European countries. Comparing data from the three rounds for all other countries we will question also why certain indicators in the majority of the countries remain stable - right, left positioning for instance - while in some others the change is so clear. Was there any particular event to explain it?

Accounting for some hot debates occurring in Portugal in 2008 over education, health and justice reforms we will try to advance some possible effects on respondents opinions for round 4.

#### 5.2.4 Analysis of media-reported event: transformation of the plot. Trace-study of the story about "seven new mortal sins declared by Vatican"

Victor Khroul, Moscow State University (Russia)

**Abstract** The research conducted in the field of Russian media clarifies functioning of mass information spreading mechanisms "media flows" and shows the relations between real life event and its coverage in mass media.

We used a complex method which we called "trace-study", focused on spreading one news, but very visible with easily recognized trace. We analysed 233 texts about "seven new mortal sins declared by Vatican", published in Russian media (news agencies, newspapers, weeklies, radio, TV and Internet media) during one week after the event (10-16 March 2008). All the texts was searched using several categories time, territorial region, type of media, level of experts involved into the estimation and commenting the problem, genre of the text, reliability and correctness of links and sources, grade of distortion of the original publication etc.

New methods lets the researchers find out the row of characteristic and mechanisms of delivering and spreading of information. It could be applied for easily homed in from the moment of their birth or creation "comets" of media reality which "trajectory" researchers could follow and study due to the modern "optics" of high quality. Computer data bases and searching systems (in Russia there are Yandex, Rambler, Integrum etc.) provide researchers with complete set of media texts.

Data of above described trace-study survey point that authentic sense and reliability seem to be secondary criteria for spreading the information on religious topics through mass media, primary ones are sensational character of the news and its correspondence with mass stereotypes. As well they clarify the lack of critical attitude to religious news reported from abroad. This combination of factors caused deep distortion and misrepresentation of real event and made media the tool of stereotypes spreading.

#### 5.2.5 Parliamentary elections and media behaviour

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

Jurjen Iedema, Social and Cultural Planning office (SCP) (Netherlands)

**Abstract** In preparing the Blueprint for the European Social Survey in the last years of the previous century, the need for providing context was foreseen as a background to survey data: It was .. well known from earlier comparative survey research that in some fields, such as electoral analysis, individual reactions to certain questions will be influenced by contextual factors and by significant events. For example, a question about the subjective interest in politics of a respondent may well be answered differently at the height of a national campaign for a general election compared to a time when no election is imminent. When fieldwork takes place during elections in some countries and not in others, possible effects of elections can hamper comparison over time and across countries. This can especially be a problem because ESS is foreseen to become an important asset for historical micro analysis. For this reason, during each round of the ESS and event database is constructed, comprising information on major events during fieldwork.

The presentation uses one type of events only, namely the presence of an election during fieldwork. It analyses whether interest in politics and news, as reflected by media use devoted to these topics, is higher in periods close to national elections. As will turn out, there is no clear relationship between the presence of elections and more

attention to news and politics in the media. This can be due to a lack of relevant information in the data, or to the fact that closeness to the election date is not the deciding factor, but crucial heights in the election campaign.

## **Chapter 6**

# **Mode of Data Collection and Data Enhancement**

## 6.1 Analysis and Evaluation of Survey Paradata

### Coordinators:

Gabriele Durrant, University of Southampton (United Kingdom)  
Frauke Kreuter, University of Maryland (United States)

### 6.1.1 Using paradata to inform best times of contact: A multilevel event history modelling approach

Gabriele Durrant, University of Southampton (United Kingdom)  
Julia D'Arrigo, University of Southampton (United Kingdom)  
Fiona Steele, University of Bristol (United Kingdom)

**Abstract** Establishing contact is an important part of the response process and effective interviewer calling behaviours are critical in achieving contact and subsequent possible cooperation. Recent developments in the survey data collection process have led to the collection of so-called process data or paradata, which greatly extend the basic information on interviewer calls. Research is needed to establish how best to use such data to inform nonresponse processes, as well as further methodological development in the specification of response propensity models.

This paper aims to build and improve response propensity models based on paradata to predict the likelihood of contact at each call, conditioning on household and interviewer characteristics. We explore the best times of contact for different types of households controlling for influences of the interviewer on establishing contact.

An advantage of this study is that we have access to rich paradata, including detailed call-record data, i.e. information recorded by the interviewer during the data collection stage at each call to the household even if contact was not made, interviewer observation data about the sample unit also recorded by the interviewer and rich information about the interviewers themselves. A key strength is that individual and household characteristics are linked to these data. The resulting data have a multilevel structure with individuals nested within households, which are nested within a cross-classification of interviewers and areas.

To model the propensity of contact at each call, we use multilevel discrete-time event history analysis. The model conditions on all information available to the interviewer, such as from administrative data and prior calls, and include call record data as time-varying covariates.

The paper aims to provide guidance to academic researchers and survey practitioners on how to model and use such paradata for the design of effective and efficient interviewer calling strategies. It is anticipated that this research will inform the improvements of responsive survey designs and the design of call-backs and follow-ups of nonrespondents, with implications for survey agencies for the allocation of time and staff resources.

### 6.1.2 Interviewer voice characteristics and productivity in telephone surveys

Henning Best, University of Mannheim (Germany)  
Gerrit Bauer, University of Mannheim (Germany)  
Leander Steinkopf, University of Mannheim (Germany)

**Abstract** Telephone interviewing has gained great importance in empirical social research, as it enables fast and cost effective data-gathering as well as the use of random samples (in contrast to online surveys). Compared to face-to-face interviews, however, the response rates are relatively low. To gain extended understanding of the mechanisms behind unit nonresponse, this study relates response rates to interviewers' voice characteristics. We use contact data generated in a survey with 78 interviewers and 606 respondents (gross sample  $N=2227$ ) and supplement these data with information interviewers' voices. The survey was conducted during 2007 and 2008 at the University of Mannheim, Germany. To obtain metadata, we recorded the interviewers' voices and analyzed the recordings using praat, a software for acoustic analysis. In a second step, each voice was evaluated on a number of subjective dimensions by 20 raters each ( $n=156$ ). Using this combined dataset, we study the determinants of interviewer effectiveness with a special focus on objective voice characteristics (e.g. pitch, loudness etc). Additionally, we account for gender, local and foreign accent as well as a variety of subjective voice attributes (friendliness, trustworthiness etc). First results show that (a) objective voice characteristics have greater explanatory power than subjective indicators and (b) a differentiation by gender is indispensable to obtain valid results. Since subjective indicators are related to the objective measures and interviewer effectiveness

only weakly, the further study of objective voice characteristics seems to be a worthwhile opportunity for future research.

### 6.1.3 Nonresponse and Measurement Error in Employment Research

Frauke Kreuter, University of Maryland (United States)

Mark Trappmann, Institute for Employment Research (Germany)

Gerrit Mueller, Institute for Employment Research (Germany)

**Abstract** Survey methodologists are increasingly concerned with the interaction of multiple error sources. Particularly prominent are discussions about nonresponse and measurement error. One hypothesis that is often found among practitioners is that sample cases that are brought into the survey only after repeated attempts and alternated recruitment strategies, are more likely to provide low quality data (e.g. Groves and Couper 1998). Data quality is often internally assessed through the proportion of missing items, proportion of don't knows and the like (e.g. Fricker 2007). Rarely, in these studies, are external data available to evaluate the quality of respondents' answers (e.g. Cannell & Fowler 1963, Olsen 2006).

The panel study PASS (Trappmann et al. 2009) is a novel dataset in the field of labor market, welfare state and poverty research in Germany. With almost 19,000 interviewed persons (in CATI or CAPI-mode) in more than 12,500 households, PASS is currently one of the most comprising panel surveys in Germany. In PASS, survey data on the employment and unemployment history, income and education of participants can be linked to corresponding data from respondents' administrative records.

Based on this study, we give an assessment of data quality as a function of contactability and response propensity as estimated from the survey's paradata. Only for some variables, the measurement error (variance or bias) assessed through the administrative records is increased with decreasing contactability and response propensity of the target persons. In particular, this is found in case of retrospective questions. Here, the differing length of time between date of interview and event explains a large part of the difference in measurement error between respondents with high vs. low response propensity.

### 6.1.4 Implementation of the Responsive Design Strategy

Milana Karaganis, Statistics Canada (Canada)

François Laflamme, Statistics Canada (Canada)

**Abstract** With the recent emphasis on analysis of collection process data (paradata), paradata research has been focusing on a better understanding of the data collection processes, phases in the collection process, and a link between the collection strategies and collection outcomes. Various tools have been developed for survey managers to better assess the data collection progress and put them in a position to adapt their collection strategies on the fly (active management). The need for a more flexible approach was acknowledged and Responsive Design Strategy was proposed (Groves and Heeringa (2006), Laflamme and Mohl (2007)).

In essence, Responsive Design Strategy aims at adjusting data collection strategies throughout the collection period in response to how well the collection progresses. Responsive Design needs to be simple, cost effective and relatively low-risk to avoid any negative effect on the quality of data and resulting estimates. Responsive Design has to take operational constraints into consideration and has to be endorsed and supported by the data collection managers for this Strategy to work. Finally, Responsive Design should ideally help better balance collection costs and data quality by refocusing the collection effort accordingly.

The first part of the paper introduces the key components of the Responsive Design Strategy proposed within Statistics Canada. The second part of the paper presents practical issues related to implementation of the Responsive Design Strategy and how the theoretical approach has to be transformed to adapt to the realities of data collection mechanisms on a specific example of a survey conducted by Statistics Canada.

### 6.1.5 Ethical Considerations in the Use of Paradata in Web Surveys

Mick Couper, University of Michigan (United States)

Eleanor Singer, University of Michigan (United States)

**Abstract** Paradata—the data captured by computerized systems during the process of data collection—are increasingly being used in Web surveys. Web paradata includes characteristics of a respondent's browser captured from the server logs (server-side paradata) and respondent behaviour on each Web page as they answer the survey (client-side paradata). The latter are typically collected by embedding JavaScript code in survey instruments, and include such measures as response latencies (e.g., time to first click), order of responding, changing of answers, keystrokes, mouse movements, and the like. Initially researchers collected paradata to improve the design or performance of the instruments; increasingly they are being used to understand or even control respondent behaviour. This raises questions about whether and how survey respondents should be informed about the collection of paradata.

We conducted a vignette-based experiment in the probability-based online LISS panel developed by the University of Tilburg. Respondents were exposed to a description of a hypothetical survey in which we varied 1) whether we mentioned the collection of paradata and how much information was provided, 2) the topic of the survey (sensitive or not) and the sponsor of the survey (government versus commercial). A total of 5,198 responded to the vignette and answered a set of background measures in a prior wave.

Respondents were significantly more willing to participate in the hypothetical survey in the condition with no mention of paradata than in any of the three conditions in which paradata was mentioned. Willingness to participate (WTP) also varied by survey topic and sponsor. Among those who expressed WTP in the survey, there was also variation in willingness to permit capture of the paradata. In this paper we examine the sociodemographic, behavioural, and attitudinal correlates of WTP under different paradata disclosure scenarios, and discuss the implications for informing respondents about the capture and use of paradata in Web surveys.

## 6.2 Detecting and Adjusting for Nonresponse Bias using Contextual and Paradata

### Coordinator:

Tom W. Smith, University of Chicago (United States)

### 6.2.1 The Multi-level Integrated Database Approach for Detecting and Adjusting for Nonresponse Bias

Tom W. Smith, University of Chicago (United States)

**Abstract** With support from the National Science Foundation, an initial assessment of the multi-level integrated database approach (MIDA) was carried out. A nationally representative sample of 400 addresses was drawn. These addresses were linked to dozens of case-level and are-level databases. These linkages confirm the feasibility of using such databases and document the rich amount of information that can be compiled about sample addresses prior to and independent of conducting a survey of the addresses. MIDA has many uses, but assessing and adjusting for nonresponse bias is probably the most important.

### 6.2.2 Contextual data and non-respondents: some insights from face to face surveys in Switzerland

Dominique Joye, University of Lausanne (Switzerland)

Sylvie Leuenberger-Zanetta, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Alexandre Pollien, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Marlene Sapin, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Nicole Schöbi, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

**Abstract** These last years, the response rates was rather low in Switzerland: even with a lot of effort, it was difficult to obtain more than 50% in face to face surveys. In order to identify the structure of the non-respondents and possibly to correct for biases, it is important to try to know more about non-respondents.

In this paper, we propose to follow two paths in this direction:

First, use the information concerning the context of the interviewee's address: urban-rural, center or periphery, linguistic region as well as description of the surroundings that we have in the contact forms of the surveys. Very often, these informations are difficult to use, because they could be influenced by the characteristics of the interviewers active in a given region. The idea is to pool the records of many surveys together, to begin with 3 waves of ESS, one of EVS and three of ISSP in Switzerland in order to have a more stable geographic distribution.

Second, use a non-response survey in order to establish not only the possible bias in the demographics but the one on the content of the answers themselves. We have realized such a study for the third wave of the ESS as well as for the EVS 2008 in Switzerland and have obtained answers on a reduced set of questions from more than 50% of the non-respondents.

By crossing information on context and non-respondent surveys, it will also be possible to measure the structure of the bias according the geographical context.

### 6.2.3 Predicting response propensity using paradata

Simone Bartsch, University of Bremen (Germany)

Uwe Engel, University of Bremen (Germany)

Christiane Schnabel, University of Bremen (Germany)

Helen Vehre, University of Bremen (Germany)



**Abstract** Based on a landline telephone probability sample of the adult population of Hamburg (Germany), a rich set of paradata was collected to predict response propensity. Available paradata include detailed information about the contact course, the number of contact attempts, an interviewers rating of respondents degree of reluctance, as well as a detailed coding scheme of interviewers convincing efforts. The survey design is flexible in terms of questionnaire length (full, core, just one exit question) and interviewer tailoring.

A logistic regression model is used to explain the probability of response to the full questionnaire by features of the contact process, convincing efforts and interviewers rating of respondents reluctance. Subsequently, the relationship between these quantities and respondent attitudes toward surveys is analyzed.

The survey is one out of a series of four surveys conducted to prepare one aspect of the main study of the project, namely the design of its recruiting interviews for building an actively recruited large multi-mode access panel.

The project is part of the Priority Programme on Survey Methodology of the German Research Foundation (DFG).

### 6.2.4 Survey participation ‘under pressure’: are working people too busy for survey cooperation?

Anina Verduyssen, Ghent University (Belgium)

Bart Van de Putte, Ghent University (Belgium)

**Abstract** Although balancing work and family demands is often more problematic for women than for men, work-family conflict (WFC) in Western societies is increasing for both women and men. International research also shows that the feeling of time pressure keeps increasing. This can make one wonder whether those busy working people still find the time to participate in surveys. The investigation of non-response bias due to busyness in surveys is often hampered by inconsistent collection of doorstep reactions. In this paper, the relation between WFC, contactability, doubt about participation and motivation during the interview is investigated using the drop-off data of International Social Survey Program (ISSP, 2002) on Family and Changing Gender Roles for Flanders. Distinction will be made for the direction of WFC: work-interfering-with-family (WIF) and family-interfering-with-work (FIW) are tested separately. Regressions show that WFC is related to doubt about participation: the more FIW respondents have, the more doubts they have about participating. But no significant relations are found between WFC and motivation of the respondents (as perceived by the interviewer). The continuum of resistance-model (Stoop, 2005) is applied to distinguish easy-to-contact respondents from hard-to-contact respondents and test the effect of WFC on contactability. When distinguishing the direction of the WFC, hard-to-contact respondents are more likely to experience WIF than early respondents. After controlling for objective indications of time pressure, namely the number of hours spent on work and on household tasks, and work-role and family-role overload, this effect of WIF is still significant. And even controlling for work stress and family stress, as more subjective measures of pressure, cannot undo the direct effect of WIF on contactability. It can be seen that the more respondents experience WIF, the harder they are to contact. Furthermore, gender differences in the effect of WFC on contactability are found: womens contactability is hindered more often than that of men. This again confirms that WFC is more often problematic for women than for men.

## 6.3 Mobile Phones and other ICT in Survey Research: Implications for Data Quality

### Coordinator:

Marek Fuchs, University of Kassel (Germany)

### 6.3.1 Mobile Phone Surveys in Germany – Response rates and response behaviour

Sabine Häder, GESIS - Leibniz Institute for the Social Sciences (Germany)

Götz Schneiderat, Dresden University of Technology (Germany)

**Abstract** The number of households solely equipped with fixed line phones has decreased in Germany since 2003. Simultaneously the number of people using mobile phones exclusively has increased; especially among young males living in single households. Latest estimations indicate that 8-12 percent of households have mobile telephones only. Experiences from other countries show even higher rates and, in addition, show that the speed at which fixed phone lines are replaced with mobile phones will increase in the future. This change in telephone coverage cannot be ignored in survey practice, since current frames for the selection of households include fixed line phones only.

To shed light on this issue the German Research Foundation was funding a joint research project of the University of Technology Dresden and GESIS, Mannheim. The project aimed to develop possibilities for combining fixed line and mobile phone surveys. For this, we conducted two surveys one on fixed line phones and one on mobile phones with around 1000 interviews each.

In the paper we will explain the outline of the project and present the results of the survey concerning response rates and response behaviour with the help of the disposition codes and further survey indicators. We will show that in contrast to the fixed line phone surveys the cooperation rate in the mobile phone survey is much higher. However, it is much more difficult to reach a respondent because many mobile phone numbers do not lead to working phones or the status working/not working is unclear. Hence further research has to be done.

### 6.3.2 Pushing Digital Pen Based Data Entry to its Limits – The Experience of the English Housing Survey

Simon Nicol, Housing Centre, Building Research Establishment (United Kingdom)

**Abstract** In 2008 the English Housing Survey physical inspection changed over from a standard paper survey form, with keyed data entry, to a system based on electronic data capture through digital pens. This paper describes how and why the change came about and the problems that had to be overcome, and evaluates the success and potential further applications following the first year in the field.

The English Housing Survey (EHS), formerly the English House Condition Survey (EHCS) is the longest running national survey of housing conditions in the world, and is considered to be the benchmark for other surveys that collect data on homes, their facilities, their state of repair, health and safety, energy efficiency and who lives in them. The first quinquennial survey was in 1967 and since 2002 it has been continuous, with annual results being published. The survey is used to develop and monitor government policies towards the housing stock, its improvement and replacement. The current sample size is 8,000 randomly selected homes of all tenures which are inspected each year.

Some 200 qualified inspectors work on the EHS on part time contracts, each surveying around 40 homes per annum. Prior to 2008 they sent their completed survey forms away for remote data entry. While this had worked well for years, there was a concern that the data took too long to key in, validate and compile into SPSS files for analysis. There was a concern over data quality, as those responsible for validating the survey forms had not visited the properties. And there were those who considered the survey to be old fashioned because of the common use of electronic data capture systems on other surveys.

Following an evaluation of a range of alternative methodologies, and a full scale pilot in 2007, it was decided that the most appropriate alternative method to the standard paper form was a web based system based on digital pen technology. In this, the inspectors complete a survey form printed on Anoto paper using a digital pen which collects the data directly. When the surveyor gets home he/she downloads the data from the pen to a PC and it is sent via the internet to the BRE EHS web site, where it is run through a series of checks and

returned for the surveyor to validate. Once validated, the data is lodged on the BRE site for compiling/analysis. While the system is complex to develop and has to be fully tested well before launch, the pens themselves are inexpensive, which was a critical factor in a survey with a large number of inspectors completing a small number of surveys each.

The system has proved to be a great success. Inspectors have adapted to the new technology well, and a key component is the built-in validation system which ensures that they take personal ownership of data quality. Now, the first the survey forms can be analysed within days of being undertaken, and the full data set can be compiled into SPSS files months earlier than previously.

### 6.3.3 Is a cell phone really a personal device? Results from the first wave of a mobile phone panel on sharing in Germany

Marek Fuchs, University of Kassel (Germany)

Britta Busse, University of Kassel (Germany)

**Abstract** In recent years cell phone coverage rates have increased rapidly. At the same time, more and more people have abandoned their landline telephone service. As a consequence, the proportion of the so called mobile-only population that has a cell phone only has increased. Thus, traditional RDD sampling procedures that rely on the landline telephone numbering frame only are questioned because of the potential coverage bias of samples drawn from this frame. In order to compensate for this coverage bias more and more telephone surveys make use of a dual frame sampling scheme consisting of randomly generated landline telephone numbers and mobile phone numbers.

Since a landline telephone number might reach a household with multiple members of the target population (e.g. adults) this number is subsequently weighted with the inverse of the number of persons in that particular household who belong to the target population multiplied with the number of landline telephone numbers that can be used to reach this household. By contrast, mobile phones are treated as personal devices, thus, the weight for these numbers is build on the number of mobile phone numbers of that particular person, only.

In this paper we will assess whether the underlying assumption is justified according to which a mobile phone is a personal device that is not shared among a group of potential respondents. Based on the first wave of a small scale mobile phone panel we will report estimates for the proportion of mobile phone numbers that can be used to contact more than one respondent. In addition, we will provide socio-demographic background information on the sharing population in Germany.

## 6.4 Use of the Internet and Mixed-Mode Surveys to Survey the General Public (I)

### Coordinator:

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 6.4.1 An experimental mixed mode design on a general population survey

Gillian Eva, City University London (United Kingdom)

**Abstract** The European Social Survey, now in its fourth round, insists on face-to-face interviewing in all countries to encourage equivalence of data. But with response rates under increasing pressure and fieldwork costs increasing, alternative data collection modes (including mixed modes) are now under active consideration.

Any change in mode, whether to another single mode or a mixed mode design, may have an impact not only on cross-national comparability but on the time series. As a result the ESS is undertaking a programme of research into mixed mode data collection in order to ensure that, if a change in data collection mode is made, the effects are known and minimised. The first phases of research examined the extent and cause of differential measurement error across modes. The focus then turned to the feasibility of conducting the ESS, an hour-long questionnaire, by telephone, and the effect of length of interview on response rates and data quality.

The latest piece of research investigates the feasibility of using mixed mode data collection on the ESS, a general population survey. The type of mixed mode design being considered in this experiment is one where different members of the same sample answer the same questions in different modes. The research focuses in particular on the response rates, costs and data quality of two different mixed mode designs; one sequential and one concurrent, compared with the main face-to-face survey. The concurrent design offers respondents the choice of mode (face-to-face or telephone interview or web self-completion) while in the sequential design respondents are offered the cheapest mode (web self-completion) first, followed by telephone and then face-to-face interviews. These designs offer the possibility of reducing (or at least better managing) costs, increasing response rates and reducing nonresponse bias.

This paper will describe the experimental design and discuss the practical implications of a possible future switch to a mixed mode design, such as questionnaire adaptation, available sample frames, and contacting sample members. The paper will then present some initial results, including response rate comparisons.

### 6.4.2 Comparison between Liss panel (web) and ESS data (face to face)

Melanie Revilla, Universitat Pompeu Fabra (Spain)

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

**Abstract** Web surveys are becoming more and more popular in survey research, mainly because their costs are usually lower than for other modes of data collection. With the increase of the internet coverage of the population in most of the European countries, the response rates obtained are also becoming high enough to collect huge amount of data and in a short period of time. However, the problems of comparability of the results with studies using more traditional modes of data collection (across time or countries) and of the potential differences in quality lead to several studies comparing data collected using web to data collected with mail or telephone or face-to-face surveys (Kaplowitz, Hadlock, Levine, 2004; Fricker et al., 2005; Heerwegh, Loosveldt, 2008; Heerwegh, 2009; Kreuter, Presser, Tourangeau, 2009). Our analyses are in this line and compare one study completed by the LISS panel (Web) with the same questions asked in the frame of the European Social Survey (face-to-face) in the Netherlands (Rounds 1, 2 and 3). First the general design and response rates for the surveys are compared, as well as the main characteristics of the samples with respect to some classic background variables. Then, we focus on an example of a composite score (using variables about immigration), and test for invariance across the different surveys before to compare the means and the qualities of this composite score. We conclude that so far, it seems possible to compare composite scores between the Liss panel and the ESS rounds. The expected mode effects are not really found. This could be linked to the specific design of these surveys: more than the mode itself, the way of implementing the survey might create differences.

### 6.4.3 A comparison of two mixed mode designs: cati-capi and web-cati-capi

Dirkjan Beukenhorst, Statistics Netherlands (CBS) (Netherlands)

Willem Wetzels, Statistics Netherlands (CBS) (Netherlands)

**Abstract** Statistics Netherlands conducts many surveys of the general population. Till a few years ago these were conducted in one data collection mode, some like (the first round of) the Labour Force Survey were CAPI surveys, others were CATI. For both types sample frames were available. The last decades CAPI surveys suffer from ever increasing non-response rates and the quality of CATI surveys is threatened because of growing undercoverage caused by unlisted numbers.

Besides these non-response and undercoverage problems rapidly increasing costs of especially CAPI surveys prompt to look for new ways of collecting data. Single mode web surveys do not qualify because of undercoverage, self selection bias and very low response rates. Theoretically mixing modes of data collection (here defined as using (almost) the same questionnaire in different modes) can solve the undercoverage problems, can raise response rates by approaching respondents in their preferred mode and reduce costs by collecting part of the data in cheap modes.

Several experiments have been conducted by Statistics Netherlands to investigate how to design large-scale surveys that use different data collection modes. A design that is extensively tested is approaching respondents with an introduction letter asking them to log in on a web site where they can fill in a questionnaire. Non-response is then approached by telephone if a telephone number can be found and approached by a field interviewer if no number can be traced. Another mixed mode design uses only two modes: CAPI and CATI. Respondents are approached by phone if they have a listed number. If not, they are approached by CAPI.

Both designs have advantages and disadvantages. The first uses three modes and will show greater mode effects, especially because in one mode no interviewer is present. The absence of an interviewer on the other hand reduces costs enormously. The CATI-CAPI design can be expected to have smaller mode effects, but is more expensive and will become still more expensive if the number of unlisted numbers will continue to increase in future. In our paper we discuss these advantages and disadvantages based on the experiments that were carried out.

### 6.4.4 Do online translated questionnaires result in higher response rates for patient surveys?

Jason Boyd, Picker Institute Europe (United Kingdom)

Arwenna Davis, Picker Institute Europe (United Kingdom)

**Abstract** Background: High response rates to patient surveys are important for ensuring that results are representative of the views of their local population. In the English NHS Patient Survey Programme, areas with high ethnic diversity and younger populations have lower response rates, with London-based organisations having the lowest response rates in England. One reason commonly cited for this reduced response rate is the higher proportion of patients in the areas for whom English is not their first language.

Objective: To investigate if offering the option to complete the inpatient questionnaire online in English or a number of other commonly spoken languages helps to increase response rates, particularly for hard-to-reach groups.

Methods: This online pilot survey was run alongside the national 2008 Inpatient Survey. Five London hospital organisations each generated an additional sample of 500 recent inpatients. Identical questionnaires were mailed to all patients but the additional sample were given the opportunity to complete the questionnaire online, in English or one of the ten most commonly spoken non-English languages in London. These participants were also able to call a free telephone number to request a paper copy of the translated questionnaire (in one of the ten non-English languages).

We compared the response rates and response patterns of the additional sample to the 2008 National Inpatient Survey sample to investigate the effect of providing an online completion option in the native language of the participants and uptake of telephone translation services.

Results: In four of the five participating organisations there was no difference between overall response rates for the online completion option and national survey samples. Generally, uptake of the online option was low in all languages (1% of sample), especially so for non-English completion (0.2%). Cost per response was unacceptably high for this research. Telephone completion of the questionnaire, in up to 127 languages, has traditionally been offered in the National Inpatients Survey. Uptake of this option in a non-English language halved from 2007 to 2008, with no difference between the online and national samples.

Conclusions: Offering an online completion option did not increase response rates to the survey overall, nor did providing translated questionnaires specifically improve the response rates for those from non-White ethnic groups. There are issues other than the language spoken which result in decreased response rates for those from non-White ethnic groups.

## 6.5 Use of the Internet and Mixed-Mode Surveys to Survey the General Public (II)

### Coordinator:

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 6.5.1 Using survey paradata to evaluate the benefits of mixed mode data collection

Caroline Roberts, University of Lausanne (Switzerland)

**Abstract** One argument in favour of mixing modes of data collection in surveys is that using a combination of modes may help to reduce the likelihood of non-response bias by encouraging a potentially greater number and wider range of respondents to participate. Indeed, alternatives to the principal survey mode are increasingly being considered as a means to compensate for the fact that certain subgroups may be underrepresented in samples, either as a result of coverage bias, or due to variations in contactability and response propensity associated with that method. This paper compares patterns of nonresponse on two implementations of the European Social Survey: 1) the standard face-to-face data collection; and 2) a version of the study conducted by telephone as part of a wider programme of research into the possibility of mixing modes. Using contact data from both surveys, we examine cooperation among different subgroups in each of the two modes and draw conclusions about the potential benefits of combining face-to-face and telephone interviewing as a way to reduce non-response error in survey data.

### 6.5.2 Mode Effect on Data Quality: Testing Measurement and Structural Invariance of Sensitive Latent Constructs among Minorities

Galit Gordoni, The Open University of Israel (Israel)

Peter Schmidt, University of Giessen (Germany)

Yiftach Gordoni, University of Haifa (Israel)

**Abstract** Latent constructs measured by multi-item scales have a central position in theory and empirical research in the social sciences. A valid analysis of models with latent constructs requires a two-step approach: 1. Assessing the relations between a construct and its observed indicators (measurement model testing). 2. Analyzing the relations between the construct and other constructs (structural model testing). Yet, most methodological studies testing mode effect on measurement error in social surveys, lack an appropriate attention to measurement model testing. This lack is especially evident in mode effect studies dealing with sensitive subjects and among special populations, in which data quality is especially vulnerable to measurement invariance problems between the modes. Since mean comparison is a central goal in mode effect studies, measurement invariance should be assessed in order to insure that potential means differences can be interpreted correctly.

Surveying minorities on sensitive topics has recently become the focus of interest in survey methodology. The present study illustrates the use of measurement invariance procedure, as an integral part of testing mode effect on response error, in sensitive latent constructs measured among minorities. The study tested the effect of mode of data collection on measurement error, both random and non-random, using multigroup confirmatory factor analysis (MG-CFA), which is a common method for measurement invariance testing. Representative face-to-face (N=701) and telephone (N=511) surveys, conducted in 2003 among the Arab minority in Israel, were used to test measurement and scalar invariance of three highly sensitive attitude scales concerning coexistence.

We tested configural invariance (same factor structure in the two modes) and metric invariance (same metric in the two modes), which is a necessary condition to ensure equivalence of the meaning of the construct. Scalar invariance (same systematic response bias in the two modes) was also tested since it is a precondition for latent mean comparison. Configural, metric and scalar invariances of two scales, social integration and threats, have been found and therefore latent mean comparison was conducted. However, for the prejudice scale only metric invariance holds. We conclude that for prejudice there is a mode effect on the measurement model and therefore also on the latent means. This lack in measurement invariance prevents a valid testing of structural invariance, due to confounding results.

The interpretation of results refers to possible causes of measurement error in face-to-face and telephone surveys in general and in the context of measuring sensitive topics among minorities. Conclusions of the specific study refer also to the advantages and limitations inherent in the techniques used, for mode effect studies.

## **6.6 WEBSURVNET: Network on challenges in survey methodology for web surveys**

### **Coordinators:**

Pablo de Pedraza, Universidad de Salamanca (Spain)

Stephanie Steinmetz, Erasmus University (Netherlands)

Kea Tijdens, Erasmus University (Netherlands)





## **Chapter 7**

# **Social Indicators**

## 7.1 Comparative social surveys in Europe: issues in methodology and implementation

### Coordinator:

Vijay Verma, University of Siena (Italy)

### 7.1.1 Comparative social surveys in Europe: issues in methodology and implementation

Vijay Verma, University of Siena (Italy)

**Abstract** Within an international framework, the keynote presentation will highlight salient issues in designing and conducting nationally representative but comparative social surveys in the European context. Some of the sampling and other methodological issues are common to different types of comparative social surveys, including to the major EU social survey programmes such as the LFS on economic activity, EU-SILC on income and related aspects of living conditions, AES on adult education, EHIS on health, ICT on internet and computer technologies, SES on structure of earnings, and also the ESS, the European social survey mostly concerned with the general population at the national level and with comparability across countries.

The presentation will consider each of the following aspects or concerns in turn.

- 1 Introduction: why do social surveys need to be comparable?
  - 2 What defines a set of surveys as comparable?
  - 3 What is the role of common frameworks, standards and procedures in comparative surveys?
  - 4 What are the implications of the basic distinction between collection and estimation aspects of the surveys, as they concern comparability?
  - 5 Standardisation versus flexibility: where are the appropriate compromises?
  - 6 The strong model: comparability in the EU Labour Force Surveys (LFS).
  - 7 The weak model: the Household Budget Surveys (HBS).
  - 8 The alternative model: the European Social Survey (ESS).
  - 9 Comparability in EU-SILC: detailed discussion of objectives and aspects of implementation in concrete terms.
- (10) Concluding remarks: are there alternative to nationally representative surveys, such as surveys designed and implemented directly at EU level?

### 7.1.2 Measuring work life balance in comparative social surveys

Anke Plagnol, University of Cambridge (United Kingdom)

Jacqueline Scott, University of Cambridge (United Kingdom)

**Abstract** The study of work-life balance is a very timely topic which has recently gained increased attention from policy makers, partly because of the conflicts that mothers often face in reconciling employment with family responsibilities. Measures of work-life balance have been included in several large scale cross-national surveys, including the European Social Survey (ESS) and International Social Survey Programme (ISSP). However, the measurement of work-life balance has been criticized because typical survey items mostly relate to paid work and neglect the life component. Not surprisingly, researchers often conclude that work-related aspects, such as work hours are most important for work-life balance, although these results could be driven by the wording of the question.

The aim of our study is to investigate what is captured by current indicators of work-life balance and how these measures could be refined. Particular attention will be paid to issues of cross-national equivalence, using data from the UK, Germany, Ireland and France. We investigate which domains of life, including paid and unpaid work, are captured by the measures of work-life balance that are included in the second round of the ESS and the 2002 ISSP. We examine problems with measuring unpaid work and differences in perceptions of time spent on household chores between men and women. We also explore an item from the third round of the ESS, which measures how satisfied respondents are with the balance between the time they spend on paid work and the time

they spend on other aspects of life, to investigate which domains of life are important for work life balance. We use the results of these analyses to propose alternative measures that would conceptualize more adequately the life in work life balance.

### 7.1.3 Data quality influence on the new housing indicators based on EU-SILC

Anna Rybkowska, European Communities Statistical Office (Eurostat) (Luxembourg)

Gara Rojas Gonzalez, European Communities Statistical Office (Eurostat) (Luxembourg)

**Abstract** Comparability is one of the most important characteristics of data in the context of cross national and decentralised data collection. Comparability is indeed particularly essential for EU wide statistics as it allows creating meaningful EU aggregate and indicators, enable countries to assess how they are performing in relation to others and it could even show how to create a sort of best practice model. For Eurostat, comparability is one of the six special features that define quality, namely relevance, accuracy, timeliness and punctuality, accessibility and clarity, coherence and comparability.

For EU-SILC (EU Statistics on Income and Living Conditions), comparability of data between MS is a priority as stated in the EC Regulation 1177/2003. In the first part of the presentation, an overview of the main issues related to comparability in EU-SILC will be given.

The second part of the presentation will focus on the newly built housing related indicators. Since the launch of the EU social inclusion strategy, poor housing conditions have been highlighted as a key factor of social exclusion. The 2001 Social Protection Committee report on indicators, to monitor the social inclusion process, recommended 'improving comparable information and reporting on decency of housing, housing costs and homelessness'. Eurostat reviewed the various SILC variables related to quality of housing and housing affordability, assessed their quality and comparability across countries and proposed on this basis two indicators for adoption by the ISG (Indicators Sub-Group of the Social Protection Committee): one based on material deprivation in the field of housing and the second based on housing costs. In this part of the presentation the focus is given to the quality of housing variables and the problems concerning comparability of these variables across countries.

### 7.1.4 Comparative procedures for net-to-gross conversion of reported incomes in EU-SILC using Siena Micro-simulation Model

Gabriella Donatiello, Statistics Italy (Italy)

Gianni Betti, University of Siena (Italy)

Paolo Consolini, Statistics Italy (Italy)

**Abstract** According to the EU Regulation on European Statistics on Income and Living Conditions (EU-SILC), gross income variables become mandatory starting from survey year 2007.

Member States could collect data on tax and social insurance contributions directly from the survey, they could also use register data or impute them by a tax-contribution model. The different procedures used by Member States may affect data quality and comparability.

For the net-gross conversion of EU-SILC variables, the Siena Micro-Simulation Model (SM2) has been developed as a practical tool. Italy, Spain, Portugal, Greece and Turkey have used SM2 for the grossing up of 2007 income variables.

This paper summarises the gross data production process and its main results in the five countries, focussing in particularly on the construction of gross income variables in Italy, where a mixed strategy has been chosen, using in conjunction the microsimulation model and an exact record linkage between survey and fiscal data at micro level.

The integration of microsimulation with register data improves the quality and the amount of information on gross income and could be considered as a refinement of the procedure for netgross conversion of income data.

### 7.1.5 Imputing fictive rents for owners using the Heckman correction: Application to CH-SILC.

Beat Hulliger, University of Applied Sciences Northwestern Switzerland (FHNW) (Switzerland)

Gordon Wiegand, University of Applied Sciences Northwestern Switzerland (FHNW) (Switzerland)

**Abstract** Comparing the income of tenants and home-owners is difficult since owners pay the service of their dwelling in a very different way than tenants. In order to allow a cross-European comparison of income, Eurostat suggests adding an imputed rent to the income of owners. Important indicators on poverty and social exclusion like the Laeken indicators are based on these adjusted incomes.

Besides owners there are tenants paying a rent which is below a market rent. For both groups rents have to be imputed. The preferred imputation method is based on a regression model for the rent which is developed for observed market rents. Rents for owners are then the fitted values based on this model. The model may have a selection bias since it is developed for tenants with market rent only. In principle, the Heckman correction allows an unbiased estimation. The Heckman method is a two-step procedure which involves a logistic regression to estimate the probability to rent at market price and a main model to estimate the amount of rent. The main model involves a correction term to account for the selection.

Application of the Heckman method to survey data needs adjustments: The models have to be estimated for the population and not only for the sample. In other words, the sampling weights should be taken into account. In addition missing values and outliers have to be taken into account.

The presentation discusses the application of the Heckman method to the data of the Swiss Statistics on Income and Living Conditions (CH-SILC). The modelling process including pitfalls will be shown.

The quality of the imputations depends strongly on the covariates at hand. Therefore the results for the imputation in SILC will be compared with the application of the method to other surveys.

## 7.2 Economic indicators and indices

### Coordinator:

Nick Longford, SNTL and University Pompeu Fabra (Spain)

### 7.2.1 A house price index based on the potential outcomes framework

Nick Longford, SNTL and University Pompeu Fabra (Spain)

**Abstract** The purchase of a residential property is the largest single expenditure of a typical middle-class household. The trends in the house prices are therefore of interest to professionals and agencies with remit in town planning, welfare, construction, taxation policy and the like, in both the private and public sphere. A house price index (HPI) is a summary of the trends in house prices. It compares the price of a typical property at one time point with the price of a similar property at another time point, such as a year ago. HPI has some features of a consumer price index (CPI). The principal difference is that a residential property is a complex product which includes not only the structure (a building or its part) and its immediate surroundings (garden, paving, fencing and the like), but also its environment, interpreted widely: access to the property, proximity of shops, schools, sources of entertainment and location for leisure activity, and other services, absence of dilapidation, criminal activity and (industrial) pollution, and the like. In contrast, the items in the basket of goods for a CPI are standardised, with their qualities and functions subject to no alteration.

A HPI is based on the census of all transactions of residential properties (houses, flats, etc.), with the attributes of the properties and circumstances of the transaction (market or private sale, mortgage or paid in cash, type of tenure, etc.). The ideal HPI compares like-with-like transactions in one period with the transactions in another. This is impossible to arrange, because we can exercise no control over the attributes of the properties involved in the transactions. This problem is addressed by the established methods by regression adjustment (hedonic house price indices) and repeated-sales analysis. The presentation will discuss the deficiencies of these, and will propose the potential outcomes framework (POF) as an approach to constructing HPI to a greater standard of integrity. In POF, the time of the transaction is regarded as the treatment, and the treatment effect is the difference in the prices paid for an identical property at one time point and another. However, a property at a fixed address is not the same consumer item as it was, say, five years ago, because of the wear and tear, maintenance and other factors (including fashion) that bring about changes, including the attributes of the environment (a new park, school, employment opportunities, reputation for petty crime, and the like). The strength of POF is that it can respond to this concern, at least in principle. A reference stock of properties is considered together with the effects of time for each property. The HPI is defined as the average treatment effect for the stock. This is practical to define on the multiplicative (log) scale, so that the calculus of percentages is simplified.

An example of a proposal of HPI for New Zealand will be discussed, together with the various options of how POF can be implemented, and their connection with the methodology for missing data, and multiple imputation in particular.

### 7.2.2 Comparing the economical growth of some countries in the European Union

Andreana Stoykova-Kanalieva, University of National and World Economy (Bulgaria)

**Abstract** In the current economical situation comparing the degree of economical development of different countries in EU is of great interest. Although there are some already established methods the search for newer approaches and research mechanisms continues.

The present report proposes one possible method, based on the statistical analysis of increases and differences. Specifically, it is based on a modification of the method of analysis suggested by V. Tzonev and T. Kanaliiev. The method is used in three ways to compare the increase of the Gross domestic product (GDP) and its sources for the countries of Southeastern Europe Bulgaria, Romania and Greece in the period 2003-2006.

The first way of the analysis assesses and compares the relative shares of two components of sources of increase in GDP component connected with the change in labor productivity in national economics and the corresponding component connected with the change of the employees (quantity of labor).

The second way of the analysis assesses and compares the relative shares of seven components of the sources of increase in GDP, which can be aggregated, to promote easier interpretation into three components. One of

these components reflects the contribution of change in the structure of employees in national economics by sectors.

The third way assesses and compares the magnitude of the parameters defining the size of the so called structural effect on the increase of GDP, that is to say, the intensity of the structural change, degree of variability of the sectored labor productivity and coefficient of correlation between the change in the structure of labor and labor productivity.

The proposed second scheme serves as a basis for other comparative analysis between the economics of different countries, as for example, to establish a conditional structural effect on GDP.

The advantages of the method in international comparative researches generally are:

First, the contribution of every source of increase is expressed through a relation between two or more parameters. Thus the algebraic form of the source (absolute or relative) has no influence on his interpretation.

Second, it is not difficult to supply the required information for the application of the method. The published statistical data is a basis for estimation both the common and of each of the partial increases.

Third, the degree of effectiveness on the accomplished structural changes in the different countries can be estimated.

Fourth, the discussed method for international comparative analysis provides useful and rich in content results with only moderate outlay of time and money.

### 7.2.3 Aspects of Ukrainian LFS Data Quality Improvement

Olha Lysa, Ukrainian National Academy of Sciences (Ukraine)

**Abstract** The data from the Ukrainian Labour Force Survey (LFS) are widely used for estimation of key socio-economic indicators for social policy making by different users at the national and regional levels. This motivates the intensive development of research on survey data quality improvement in the Ukrainian official statistical system. Pursuing modern international standards of statistical information quality, research is directed toward development and implementation of complex approaches to estimation of quality indicators and improvement of quality.

The presentation will discuss the system of key quality indicators for the assessment of quality of LFS data and the influence exercised on it by the data processing procedures. It is shown that using a system of complex statistical weights, which is built with an accurate account of the sample rotation scheme, non-response patterns and auxiliary information, enables us to improve the reliability of LFS data.

Special attention is paid to developing methods of indirect estimation of monthly labour force indicators and assessment of their influence on some quality indicators of reliability, timeliness, comparability and coherence. The basic results of these approaches realization are presented.

Issues related to the development of special approaches which allow improving of LFS data quality for the lowest data aggregation levels will be discussed.

## 7.3 Integration of objective and subjective indicators: methodological and technical issues

### Coordinator:

Filomena Maggino, Università degli Studi di Firenze (Italy)

### 7.3.1 Methodologies to integrate subjective and objective information to build well-being indicators

Filomena Maggino, Università degli Studi di Firenze (Italy)

**Abstract** The integration of objective and subjective indicators represents a crucial approach in order to come up with a solid scientific result and understanding of relevant social phenomena. The subjective and objective dimensions can provide new perspectives and allow the units of interest to be evaluated also in social policy.

This is confirmed also by many international initiatives and events revealing the increasing attention on individual perception of living conditions and quality of life and the necessity of a correct scientific approach to their measurement and analysis. In this perspective, we can include, for example, the theoretical debate around the relationship between economics and happiness.

The need to integrate subjective and objective information comes from different sources (statistical offices and survey) and is causing a growing demand in the study of well-being and happiness of societies (see OECD and its agenda of the Measuring and Fostering the Progress of Societies World Fora on Statistics, Knowledge and Policy).

In a policy perspective, the need for subjective indicators arises during (i) the assessment of policy results and (ii) in the selection of policy objectives (Veenhoven, 2002). The first item concerns the need to assess if a policy has been successfully implemented (e.g. is there perception of more security in the streets after increasing police staff?). The second item refers to what people desire (e.g. less air pollution, more cultural events in the city). The need arises also in consideration of the limits of objective indicators (Veenhoven, 2002), as reality cannot be fully reduced only to objective facts. Moreover, objective facts are measured referring to a design or a model that is subjective in its definition.

For this reason the definition of an integrating model is needed. This model requires firstly the definition of a conceptual framework from which it is possible to identify the proper analytical approach (causal analysis, multilevel analysis, life-course analysis, or explorative analyses). Secondly, it requires an organizational context in which the integration can be accomplished by relying on structured and systematic data, observed in long-term longitudinal perspective (e.g. systems of indicators) and in which particular technical issues (i.e. aggregation issues) can be managed.

The paper will discuss these aspects by referring in particular to the feasibility of the different statistical approaches taking into account their specific assumptions. The goal is to describe a procedure able to yield results, not only statistically valid and consistent with reference to the defined conceptual framework, but also easy to read and interpret at policy level.

### 7.3.2 The gap between objective relevance and subjective recognition of handicaps. How to measure physical, mental and social handicaps adequately

Rossalina Latcheva, Institute for Advanced Studies (Austria)

Angela Wroblewski, Institute for Advanced Studies (Austria)

**Abstract** In surveys it is a challenge to measure the existence of physical, mental and social handicaps in a way that corresponds to official statistics as well as to the perception of respondents. Experience with surveys among students showed that there is a significant gap between the official status of a handicapped person (e.g. a person that is registered to be handicapped and receives state assistance accordingly) and the individual perception of being handicapped (self assessment). This dilemma has to be solved in surveys that want to deal with two research questions: Firstly, the survey should give an overview of the share of handicapped students that is comparable to official statistics. Secondly, the survey should provide a basis to analyse existing barriers for handicapped students in everyday life at universities. Especially the latter requires a self-assessment of students of being physically, mentally or socially handicapped which leads to disadvantages or specific forms of discrimination.



In the paper we will discuss possible solutions to that dilemma by referring to a survey among students in Austria that has been conducted three times (2002, 2006, 2009). The questionnaire (online) is developed on the basis of a four stage pre-test including cognitive pre-tests as well as focus groups. Especially these qualitative parts of the pre-test give valuable insights in the gap between the official status and respondents perception of social reality.

### 7.3.3 Measurement of Teaching Productivity (Integration of Subjective and Objective Indicators)

Magdalena Jelonek, Jagiellonian University (Poland)

**Abstract** The aim of this paper is to present a measurement model for teaching productivity of university. Typically concept of teaching (educational) productivity is composed both of objective and subjective indicators (objectively defined teaching productivity versus subjective perception) which generate a lot of conceptual and methodological problems.

This paper examines the possibility of constructing measures of the performance of university composing both subjective and objective indicators. Model of educational productivity is developed by reconstructing goals and objectives of university (European, national, regional level) and exploring social perception of role of university (using qualitative methods).

The implication of the findings with regard to future research on teaching productivity of university are discussed.

Keywords: educational, teaching productivity, indicators, higher education

### 7.3.4 Subjective and objective indicators in the measurement of immigration flows

Mónica Méndez, Centro de Investigaciones Sociológicas (CIS) (Spain)

**Abstract** The goal of this paper is to discuss the combination of subjective and objective indicators to measure the attitudes towards immigration. This will be done by using a question of the European Social Survey (2002) that asked respondents to give an estimate of the number of foreign born out of every 100 nationals in each country. The answers to that question can be combined with an objective indicator, namely, the actual number of foreign born in each country. This offers a good occasion to study how subjective and objective indicators can be combined.

In addition to the ESS data, the Spanish case will be studied in greater depth. The same question that was asked in the 2002 round of ESS was included in a survey on attitudes towards migration in 2007 and 2008 (the surveys were carried out by CIS, Centro de Investigaciones Sociológicas), and, in 2008 the question also made reference to the region where the interviewee lived. These data will be analysed to check on how this subjective and objective indicators can be combined in order to further understand and interpret other attitudes to migration (also included in the survey).

### 7.3.5 The application of Extended Latent Class Models to the integration of objective and subjective indicators for scientific and policy analysis purposes

Karolina Keler, Jagiellonian University (Poland)

**Abstract** The problem of integrating subjective and objective indicators is essential in scientific and social policy analysis conduct. The scientific importance is an effect of the growing need to find methodological tools for this integration. Moreover, the combination of these two indicative components, and thus meaningful reduction of data, could be fruitful for social policy aims. The main aim of this paper is to present the approach which allows both the scientific and policy purposes to meet at the same time. Furthermore, it should show why trying to answer both types of questions helps to find the solution of the problem of combining the subjective and objective information.

The main thesis of the paper is that the Extended Latent Class Models, as an example of a more general approach to the latent variable modelling, serves as an appropriate tool to provide scientific understanding and recommendations for public policy by an integration of subjective and objective indicators. To make an attempt

to prove this thesis, survey data from 2007 is used. This methodological survey (carried out in one region of Southern Poland - Malopolska) captured different dimensions of the peoples evaluation of well-being and their objective situation. Furthermore, it was conducted to collect in one study the most important data gathered regularly by different public institutions (the statistical office and welfare institutions such as labour or public assistance). This approach makes it possible to choose proper indicators on the basis of rather easily available data. The data from this survey is used to analyse the (objective) Laeken Indicators in terms of subjective well-being, that should be considered as an exemplary analysis using the proposed mode of analysis. The results of this analysis provide not only the answer to the main question how to combine subjective and objective indicators, but additional conclusions important to the subject matter. These conclusions include the answers to the following questions: how to conduct methodological surveys for indicators development, how to choose indicators for portfolios used to monitor some phenomena as well as understand the mechanisms that lay behind them. However, in this paper the main emphasis is put on the role of flexible structural models in the process of substantive combining objective and subjective indicators, using the example of the indicators of social exclusion.

### 7.3.6 Integrating subjective and objective information: an application on European data

Elena Ruvigliani, Università degli Studi di Firenze (Italy)

**Abstract** The quality of life is a multidimensional concept composed of different component, traditionally classified in term of objective and subjective components. Many research measure the reality considering only objective information or subjective. A more realistic picture of the quality of life concept should be obtained if subjective and objective components are suitably combined. Some assumptions have to be made concerning the relationship between the two components. In broad outline, two perspectives can be alternatively defined: (a) objective quality of life at macro level can be considered an antecedent with respect to subjective quality of life (subjective well-being). In this case, objective indicators (input) can be interpreted in terms of contextual conditions that can explain the subjective indicators (output); (b) objective quality of life conditions at macro-level and subjective quality of life (perceptions) are independent; perceptions are influenced by individual characteristics and not by the objective living conditions. In this case, subjective indicators (input) can be considered as an important component driving the improvement of objective conditions.

The paper describes a procedure aimed at integrating objective and subjective data by showing an applicative study.

The possibility to develop integration between objective and subjective indicators is based upon the chance to rely on organized data, observed in long-term longitudinal perspective. This is particularly true with reference to subjective indicator, which are strongly dependent and linked to a great use of resources. In this perspective, a composite analytic approach should be defined. The study has been carried out on both subjective (from the European Social Survey project) and objective (provided by the Joint Research Centre JRC European Commission) data. The application shows how a simple integrative approach is able to reveal its efficacy in the perspective objective and subjective information.

The Econometrics and Applied Statistics Unit (EAS) of the JRC has actually and fruitfully contributed to the preparation of this study.

The approach is defined composite as carried out through subsequent steps (multi-stages) and by different analytical approaches (multi-techniques). In particular, four stages can be identified in order to complete the procedure.

## 7.4 Measuring Civil Society - What are the issues?

### Coordinators:

Helmut Anheier, University of Heidelberg (Germany)

Michael Hoelscher, University of Heidelberg (Germany)

### 7.4.1 Civil society items in international surveys – existing resources and possible developments

Sally Stares, London School of Economics (United Kingdom)

**Abstract** Civil society is in many ways a foundational concept in sociology and social policy. In many studies of civil society, therefore, researchers find it expedient to make use of survey data which have already been collected for a broader or different purpose than focusing specifically on civil society. Indeed, a great deal of information about civil society can be gleaned from existing surveys. However, it is often the case that their items are not quite as closely focused as we would wish them to be to capture civil society; or conversely that they are too closely focused on certain manifestations of it to enable comprehensive international comparisons to be made from them.

This paper will provide an overview of the types of questions which have been used to date in international surveys to capture civil society, and will suggest which theoretical models of civil society do and do not fit with them. Its aim is to provide the background for the subsequent papers and discussions, with a view to developing and improving survey items for international studies of civil society.

### 7.4.2 Identifying Civil Society Measurement Gaps – Evidence from Germany

Norman Spengler, University of Heidelberg (Germany)

**Abstract** Development in researching civil society requires more observation tools to allow testing theoretically derived hypotheses. Improving conceptual assumptions and operational descriptions must be based on thorough empirical analysis. Established survey systems deliver an important asset of raw material of useful unexploited data. By exploring these available potentials for civil society measurement, a better understanding of this complex research object could be contributed.

The paper explores and investigates the inventory of selected survey data for civil society measurement purposes by considering the German case. The following issues are intended to be addressed and discussed:

- 1 When considering civil society theory, which aspects and topics could be operationalized by survey systems?
- 2 Which gaps can be identified that may be from civil society research perspective of interest for further survey improvement?

This paper neither seeks to introduce a new maximalist operable measurement model of civil society, nor tries to offer a wish list of imaginable survey items. However, the paper aims to deliver an input for a stimulating discussion whether similar conclusions can be reported from other countries and in what way these lessons learned can be applied on European level.

### 7.4.3 Challenges for Comparative Research on Philanthropy in Europe

Rene Bekkers, Vrije Universiteit Amsterdam (Netherlands)

Pamala Wiepking, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** A framework is presented for comparative research on philanthropy in Europe. Previous studies of charitable giving have been conducted mostly in the US, the UK and the Netherlands. Questionnaires and survey designs of these studies are discussed with respect to their potential use in comparative research. The optimal research design is compared with actual research designs used in the European Social Survey 2002 and the Eurobarometer 62.2. The same regression equation is estimated using data from both surveys, and differences in parameter estimates are evaluated in the light of methodological differences between the two surveys.

#### 7.4.4 The improved 'Civil Society Index' as a pragmatic approach of combining theory and available data

Michael Hoelscher, University of Heidelberg (Germany)

**Abstract** The Civil Society Index by CIVICUS has been used in extensive research in more than 50 countries until now. Recently, the methodology has been improved, streamlining the applied techniques, fitting them better to their purpose and allowing for better international comparability. The paper will present the advanced toolkit for the CSI and discuss some problems in the light of prior experiences.

Especially the following questions will be addressed:

- The CSI has an important activity-component for civil society actors in the specific countries attached to it. How can this be combined in a useful way with scientific rigour?
- The situation of the civil society sector is very different in different countries. How can the CSI measure the CS in each country in a valid way and guarantee at the same time international comparability?
- How is the CSI connected to international available survey data?
- The CSI comprises different tools, reaching from a population survey, to an organizational survey, to expert interviews, and media analyzes. How can the results from the different instruments best be integrated?
- What are the next steps to further improve the CSI?

As these questions are connected to all sorts of methodological problems, this session wont offer final solutions. However, the idea is to discuss these topics by finding certain particular aspects that can be tackled in more detail and may serve as a basis for an interesting discussion afterwards.

## 7.5 Quality of measures for concepts of Social sciences (I)

### Coordinator:

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 7.5.1 Evaluation of concepts of Social Science

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 7.5.2 How to deal with scales from 0 to 10 in Geometric Data Analysis?

Flora Chanvrlil, Sciences Po (France)

**Abstract** The debates on response modalities measurement are virtually endless between those who prefer questions which force respondents to take side (the agree / disagree format) and those who advocate for both a median answer and a wide scale of response. Clearly the European Social Survey teams have taking side since the questionnaires contain a high number of response modalities ranging individuals on 0 to 10 scales. This preference in itself is justified but creates several substantial and statistical problems when using methods of Geometric data analysis, particularly specific Multiple Correspondence Analysis (Le Roux, Rouanet, 2004).

Most of the problems are caused by the median answer (5) chosen regularly by a high proportion of respondents. This kind of distribution creates quite systematically a Guttman effect, a well-known problem in GDA which ends up with a factor opposing extreme and median answers. Substantially, this effect challenge the aims of MCA: basically 1) to identify the main cleavages upon which the electorates oppose and 2) to sort out groups of respondents on value dimensions.

We propose a solution of this Guttman effect based on the Benzecris design (Benzecri, 1980) and apply it to the ESS round 2 and 3 French data. This method helps to solve the Guttman effects problem and give substantial results regarding the political and normative space of the French respondents on dimensions such as politics, Europe and globalization, ethnocentrism, institutions, economy.

### 7.5.3 Latent class analysis in identification of relation between optimists and pessimists in Poland based on The State of The Households Survey

Piotr Bialowolski, Warsaw School of Economics (Poland)

Dorota Weziak-Bialowolska, Warsaw School of Economics (Poland)

**Abstract** The paper aims at presenting an approach to optimism measurement among Polish households with application of data from a Consumer Tendency Survey (CTS). One of such surveys is conducted at Warsaw School of Economics (The State of The Households Survey) and it is based on a questionnaire harmonized for the EU countries methodology developed by the European Commission. CTS usually focus on trend analysis and are designed to describe and forecast changes in the consumer behaviour. However, as they measure forecasts of economic situation with application of different indicators (questions) it is also possible to grasp a latent category optimism.

In this paper authors approach the problem of measuring optimism with application of Latent Class Analysis. It is their objective to determine the number of latent classes that can be distinguished with respect to households attitude towards the economic situation. They try to explain the existing pattern referring to the situation in Poland. With application of Latent Transition Analysis they also try to evaluate changes in households optimism concerning the economic situation in Poland between 2006 and 2008.

### 7.5.4 A measure for political satisfaction: formative indicators or reflective indicators?

Maria Cristiana Martini, University of Modena and Reggio Emilia (Italy)

**Abstract** Political satisfaction is one of the many concepts in social sciences that can be measured (and are often measured) by means of multiple indicators. The operationalisation and measurement of political satisfaction is not straightforward: satisfaction with democracy or satisfaction with government are often used to measure political satisfaction as a whole, while it has also been shown that economic performances affect the popular political support, as well as system performances (in terms of quality of education and health services) can influence system support and political satisfaction. However, at least for some of these concepts, the question arises if they should be considered as reflective indicators or formative indicators of the more general construct. After a brief discussion of the operationalisation of political satisfaction, based on the European Social Survey (ESS) data, a measurement model for political satisfaction is fitted, that allows the three sub-constructs satisfaction with democracy, satisfaction with government and satisfaction with economy to act both as formative and/or reflective indicators. The model is then tested for measurement invariance across the different countries in the ESS.

### 7.5.5 An ethnographical look at survey questions

Analia Torres, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Rui Brites, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Bernardo Coelho, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

**Abstract** In this paper and based on the experience of comparative analysis with three rounds of the European Social Survey we propose the need for an ethnographic look at survey questions. We define this ethnographic look at three levels.

First, and a part from a special attention to the way questions are formulated we all know how hard it is to find the perfect question, the good indicator for measuring variables - we propose an insightful and descriptive look before starting multivariate analysis and procedures. We will give several examples of answers to questions with Likert scales from European Social Survey, but also from other surveys, where very often the modal answer concentrates on the position *neither agree, nor disagree*. How far can we go in interpretation when it is this special position that attracts the large majority of the respondents? And is it not true that when analysing data, producing images by charting the results, or with multivariate analysis, we often forget the first descriptive information?

Secondly, and insisting on the way questions are phrased, we propose focusing our attention on a mixed way of making questions. It is very common the use of questions representing clear stereotypes. Using also neutral questions may be also more insightful. We will give some examples related to gender stereotypes showing how conclusions can be different when questions are phrased in different ways.

Finally, and based on the great advantage that represents the replication of the same questions in ESS three rounds, we will give some examples showing how careful we must be when concluding only from one application when comparing countries. A difference showing up in 2002, can disappear in 2004 and reappear in 2006...

In our view a mixing methods stance can also be adopting an ethnographic look at questionnaires. Mixing methods is, at least partially, the adoption of a sociological prudent standpoint, avoiding misconceptions and misinterpretations of data, on one side, and avoiding enchantment with the graphical representation that can be misleading, on the other. The output of combining a quantitative approach with ethnographic sensitivity leads to interpretations more faithful to the data and fulfilling the heuristic possibilities of mixing methods.

## 7.6 Quality of measures for concepts of Social sciences (II)

### Coordinator:

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

### 7.6.1 Women's work and care orientations in Europe

Analia Torres, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Rui Brites, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Bernardo Coelho, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

Inês Cardoso, Instituto Superior do Ciências do Trabalho e da Empresa - ISCTE (Portugal)

**Abstract** Family and paid work are fundamental dimensions for individual projects constituting nowadays a mutual dependence relationship. Gender relations play also an essential role on the different ways of organizing daily experience around work and family practices, expectations and responsibilities.

Exploring the European Social Survey Round 2 (2004) and Round 3 (2006) data we analyse and discuss the connection between the different ways of organizing life between work, family and gender relations. We have tested the hypothesis that womens orientations towards work will vary less among countries than orientations to care. In some countries, orientations towards paid work can be modern but orientations to care can change more according to a set of institutional, cultural and economic constraints. This picture is drawn, nevertheless, in a scenario of a clear passage from more traditional gender values towards gender equality, opening room for gender relation negotiation in family context.

This private negotiation is a double folded process: there is a cleft between the ideological and the practices levels in the gender relations in the family context. If at the ideological level, both men and women tend to agree in more egalitarian family relationship it is also truth that a more traditional stance appears when gender roles referring to womens place in the family are directly at stake.

### 7.6.2 Support for European Level of Political decision making

Lluís Coromina Soler, University of Girona (Spain)

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

**Abstract** In this paper we look for patterns of the opinion about different policy levels governance - for eight different topics, which are protecting the environment, fighting against organized crime, defense, social welfare, aid to developing countries, immigration and refugees, interest rates and agriculture that is already decided at European level.

Our aim is to use the opinion patterns in order to create a score or index for the respondents. Mokken cumulative scale will be used because the policy levels (national or supranational) can be considered as ordinal from a higher to a lower level of governance. The consistency of the cumulative scale is measured by Loevinger's H, which can be considered as a measure of quality. Data are concerning 22 European countries form the first round of the European Social Survey. Differences between European countries can be studied in order to check if exist different patterns of support in the different European countries.

We will also relate the personal levels of support of European decision making with other personal characteristics of the respondents.

Keywords: subsidiarity, governance, legitimacy, cumulative scale, Mokken scale

### 7.6.3 The measurement of public attitudes towards surveys. An illustration with a Spanish "survey on surveys"

Marta Fraile, Consejo Superior de Investigaciones Científicas (CSIC) (Spain)

Joan Font, Consejo Superior de Investigaciones Científicas (CSIC) (Spain)

Mónica Méndez, Centro de Investigaciones Sociológicas (CIS) (Spain)

**Abstract** The goal of this paper is to discuss the development of indicators towards surveys. We argue that attitudes to surveys can be measured through several substantive dimensions that are all potentially relevant for their impact on survey quality: the degree of knowledge that people declare to have about surveys, the frequency of attendance to surveys' results; the degree of confidence and the perceived utility of surveys. Second, we explore the social and attitudinal basis of these attitudes. Third, we analyse how these five different theoretical dimensions work empirically. We make use of factor analysis techniques and find that, in contrast to the five theoretical dimensions identified previously (based on the existing literature on this topic), there seems to be only two different empirical dimensions that characterise public opinion on surveys in the Spanish case. The first one refers to the supply side: the technical characteristics of surveys and their credibility as instruments to measure public opinion, whereas the second relates to the public demand of surveys, i.e. the consumption and enjoyment of surveys from the public. The paper finishes with a discussion of the implications of this findings for relevant issues such as response rates.

To address these questions we will analyze a survey on the views citizens have of surveys that was carried out by CIS (Centro de Investigaciones Sociologicas) in March 2007 to a representative sample of the Spanish population.

#### 7.6.4 On the measurement of attitudes in surveys: from index construction to learning from data via nonparametric modelling

A. Jan Kutylowski, University of Oslo (Norway)

**Abstract** Aim of the presentation is to review a number of fundamental issues in the modelling of attitudes on the basis of (ordinal) categorical indicators, with emphasis on the role of attitude measurement in survey practice. As an example, data concerning self-reported locus of control from two national Polish surveys, conducted in 1984 and 1998 will be (re)analyzed in reference to the original interpretations.

Conventional procedure of summated ratings, as well as agregation based on the assumption of pseudo-metrical category scores will be shown inadequate with respect to a number of issues arising in attitude measurement. These include choice of scale for representing the latent variable, verification of the ordering of response categories, including interpretation of the difficult to say category, numerical representation of the latent factor, inter-temporal comparisons, assessment of over-time reliability and differential item functioning.

It will be shown that these issues are adqueately solved by a paradigmatic switch to a functional nonparametric latent factor model (Kutylowski, 1997). Such a model, based on the idea of kernel smoothing of the category probabilities as functions of the latent factor, originally developed by Ramsay (1990, 2001), will be applied to the substantive task of quantifying locus of control via learning from data, separately and jointly for the two periods.

It will be shown that quantification of the locus of control in this flexible modelling framework leads to analytically satisfying treatment of the technical issues mentioned above, as well as to meaningful findings concerning comparisons across socioeconomic positions, associations with other attitudes, and role of the attitude in question as an intermittent variable contributing to political choices.

Wider questions concerning the use of attitude measurement will also be considered, in particular robust measurement of attitude scores in terms of ordinal rather than metrical scores, and inclusion of the attitude scores in multistage models involving several directly or indirectly observed variables.

The presentation will aim at being self-contained, for the purpose of quiding the reader as to similar applications in other substantive settings.

References:

- Kutylowski, A. J. (1997). Nonparametric latent factor analysis of occupational inventory data. Chapter 24 in: Rost, J. and Langeheine, R., eds.; Applications of latent trait and latent class models in the social sciences. New York/Munster: Vaxmann.
- Ramsay, J. (1991). Kernel smoothing approaches to nonparametric item chatacteristic curve estimation. *Psychometrika*, 56, 611-30.
- Ramsay, J. (2001). Testgraf 98. Computer program and manual. Available from [www.psych.mcgill.ca](http://www.psych.mcgill.ca).



## 7.7 Social Indicators of Trust in Criminal Justice

### Coordinators:

Stephen Farrall, University of Sheffield (United Kingdom)  
Mike Hough, King's College London (United Kingdom)  
Jonathan P. Jackson, London School of Economics (United Kingdom)

### 7.7.1 Trust, Legitimacy, Compliance: A Review of Concepts with Relation to Criminal Justice'

Jonathan P. Jackson, London School of Economics (United Kingdom)  
Ben Bradford, London School of Economics (United Kingdom)  
Mike Hough, King's College London (United Kingdom)  
Stephen Farrall, University of Sheffield (United Kingdom)

**Abstract** This paper offers a conceptual discussion of public trust, legitimacy and compliance in relation to legal authorities across Europe. We aim to shed some light on the relationship (and balance) between trust and legitimacy and how they relate to compliance. The discussion draws on a range of social-theoretical, psychological and sociological perspectives to show that trust and legitimacy are interrelated but distinct, and that both are based on and expressed through a range of underlying ideas, opinions and orientations about police fairness, effectiveness and engagement with the community, for example. We distinguish between: institutional trust and encounter-based interpersonal trust; legitimacy as justification and as duty to obey; and opinions and feelings which attach to institutions and those which attach to organisations. These concepts will be found in different combinations, and be of variable importance, in different jurisdictions and/or social settings. The paper concludes with a discussion of some of the implications of the conceptual work.

### 7.7.2 Situating police legitimacy: The example of England and Wales

Ben Bradford, London School of Economics (United Kingdom)

**Abstract** Although concepts such as legitimacy and trust are universal, local contexts and exigencies will affect how they play out from country to country. This paper examines how the legitimacy of the police in England and Wales has been both conceptualized theoretically and experienced by individuals on a day-to-day basis. It attempts to develop an understanding of the legitimacy of the police among the policed which is both true to local context but which also contains elements which may be transferable to other, possibly very different, jurisdictions. After sketching out a working understanding of legitimacy, the discussion moves on to outline the unusual position of the police in English and Welsh culture, emphasizing in particular the strong link that still exists between representations of the police and ideas of national identity, community, and belonging. Data from a large-scale sample survey of Londoners are then used to outline how police legitimacy appears to be built up in the light of this link (as well as other important ideas about the police).

### 7.7.3 Tracking trajectories in public confidence in the police: a time series analysis.

Katy Sindall, University of Surrey (United Kingdom)

**Abstract** Improving public confidence in policing is high on the public agenda and the focus of much policy making. Since its conception, patterns of public support and trust in the police have changed substantially, but while an abundance of research has examined the causes and consequences of citizens attitudes toward the police, it has primarily done so using individual-level, cross-sectional analysis, and in so doing has paid comparatively little attention to aggregate changes in confidence over time. Using the British Crime Survey over its first 25 years, and employing both descriptive and parametric time series analyses, this paper explores how trajectories in confidence in the police have evolved over time, how these evolutionary paths may differ for varying subgroups in the population and whether the relationships found between confidence and some of its key drivers at the individual, cross-sectional level, hold over the longer term.

### 7.7.4 Fear of crime and motivated social cognition: Evidence from round 3 of the European Social Survey

Carolyn Cote-Lussier, London School of Economics (United Kingdom)

Monica Gerber, London School of Economics (United Kingdom)

Jonathan P. Jackson, London School of Economics (United Kingdom)

**Abstract** In this paper we draw links between fear of crime and psychological needs and preferences regarding social structure and stability. We build upon existing evidence that fear of crime expresses public concerns about norms and values in society, about conflicts and divisions, forms of social and moral ordering, and the manner in which society is regulated and governed. We analyse data from the European Social Survey to explore the relationship between fear of crime and the cognitive-motivational preferences that drive desired social conditions. And we discuss the findings with regard to John Josts system justification theory and the work of Jonathan Haidt on moral instincts.



## **Chapter 8**

# **Data Analysis**

## 8.1 Analysis Strategies for Cross-Cultural Research

### Coordinators:

Michael Braun, GESIS - Leibniz Institute for the Social Sciences (Germany)  
 Timothy Johnson, University of Illinois at Chicago (United States)

### 8.1.1 An Illustrative Review of Techniques for Detecting Inequivalences

Michael Braun, GESIS - Leibniz Institute for the Social Sciences (Germany)  
 Timothy Johnson, University of Illinois at Chicago (United States)

**Abstract** There are currently several analytic approaches available that have been proposed as useful tools for assessing measurement comparability in cross-cultural research. This introduction gives a brief overview of the main issues involved in statistically dealing with comparability problems, describes the main approaches used in the literature and applies them all to the same data. In order to demonstrate all the techniques with one example, we must make several simplifications. We forego a comprehensive presentation of the results which can be obtained when applying each technique and concentrate instead on an illustration of the main insights the different techniques offer.

Though the more advanced techniques allow for a higher flexibility and thoroughness in analysis, the basic techniques are also not without merit. They are not only easier to implement for the researcher who is more interested in substantive problems than in methodology. The basic techniques are also useful to familiarize oneself with the data which might be crucial to avoid accepting some surprising, yet entirely artifactual results that an improper application of the more complicated techniques might yield. Moreover, as we will show, a superficial application of the more advanced techniques, e.g., exclusively taking into account fit indices without also examining other relevant information (such as modification indices in confirmatory factor analysis), might lead to a failure to detect problems. Fortunately, some of the statistically more advanced methods of analysis share the ease of applicability with the basic methods. Finally, as software becomes more advanced and user-friendly, all of the advanced techniques may become easier to use and therefore more popular.

### 8.1.2 Negative media portrayals of immigrants give rise to majority members' perceived group threat: A longitudinal analysis

Elmar Schlüter, Wissenschaftszentrum Berlin für Sozialforschung (WZB) (Germany)  
 Eldad Davidov, University of Köln (Germany)

**Abstract** Perceptions of group threat, whether related to majority members economic interests or cultural identity, have proven to be a key predictor of majority members anti-immigrant prejudice. However, research has only partially identified the contextual-level characteristics giving rise to such perceived threat. Most studies in this domain focus their attention on structural characteristics as explanatory factors, e.g. the demographic size of the immigrant population. Surprisingly, up to date the role negative media portrayals of immigrants might play for explaining majority members perceived group threat has received only little attention. To remedy part of this neglect, we examined the role of negative media portrayals of immigrants for explaining majority members perceived group threat during the period 1996 to 2007 in Spain. Controlling for various regional- and individual-level characteristics, our data show that an increase in negative media portrayals of immigrants is associated with an increase in majority members threat perceptions. Our findings highlight the importance of including the mass media as an explanatory factor in a group threat framework.

### 8.1.3 A Comparison of Measurement Equivalence Methods based on the Confirmatory Factor Analysis, Item Response Theory, and Latent Class Factor Analysis

Milos Kankarash, Tilburg University (Netherlands)  
 Jeroen Vermunt, Tilburg University (Netherlands)  
 Guy Moors, Tilburg University (Netherlands)

**Abstract** Three distinctive methods of assessing measurement equivalence, i.e. confirmatory factor analysis, differential item functioning using item response theory and latent class factor analysis, make different modeling assumptions and adopt different procedures. Simulation data are used to compare the performance of these three approaches in detecting the sources of measurement inequivalence. For this purpose, we simulated Likert-type data using two non-linear models, one with categorical and one with continuous latent variables. Inequivalence was set up in the slope parameters (loadings) as well as in the item intercept parameters in a form resembling agreement and extreme response styles. Results indicate that the item response theory and latent class factor models can relatively accurately detect and locate inequivalence in the intercept and slope parameters both at the scale and the item level. Confirmatory factor analysis performs well when inequivalence is located in the slope parameters, but wrongfully indicates inequivalence in the slope parameters when inequivalence is located in the intercept parameters. Influences of sample size, number of inequivalent items in a scale, and model fit criteria on the performance of the three methods are also analysed. The number of respondents per group (sample size) proved to be the most important factor influencing results in the three approaches.

### 8.1.4 Measurement equivalence vs. Representativeness: The influence of response enhancing measures on the comparability of answers.

Joost Kappelhof, Social and Cultural Planning office (SCP) (Netherlands)

**Abstract** In recent years several large scale surveys among difficult to survey groups have been conducted in the Netherlands. Tailor made approaches have been developed to increase the response among these difficult to survey groups. The reason to try and achieve a higher response is to reduce bias in population estimates which will occur when non-respondents systematically differ from respondents with regard to the variables under investigation. The probability of getting biased estimates will increase when response is unequal across different groups. In the Netherlands ethnic groups are seen as difficult to survey. The response among ethnic groups can be increased by the use of interviewers with the same ethnic background, translated questionnaires, longer fieldwork periods and an increased number of contact attempts.

The drawback of these response enhancing measures is that one cannot assume that a (latent) factor will be measurement invariant across these ethnic groups because of the ethnicity of the interviewer and language -and cultural differences. On the one hand because respondents with different cultural and ethnic backgrounds might differ in their opinion as to what is important about the factor being measured and on the other hand because of the perceived social undesirability of certain answers or opinions. With regard to social desirability the ethnicity of the interviewer and language might be of influence. The ethnicity of the interviewer would particularly play a role when the question is about specific ethnic issues. Also the gender match between interviewer and respondent will have an effect as will the sensitivity of the issue.

One of the main objectives of cross cultural survey research is to compare concepts across groups. Specifically in the case of difficult to survey groups this usually leads to a trade-off between representativeness (high and equal distributed response among groups) and measurement invariance. There is the need to adequately measure concepts among all groups involved in the survey and to be able to compare those concepts across group for which they need to be measurement invariant.

This research focuses on the effect of response enhancing measures, such as the use of ethnic interviewers and gender matching on the measurement invariance of concepts. The second aim of this research is to establish if the effect of these response enhancing measures on the measurement invariance of concepts increases if the concepts are more socially sensitive.

### 8.1.5 Symbolic data analysis for cross-country comparisons

Seppo Laaksonen, University of Helsinki (Finland)

**Abstract** Edwin Diday and Monique Noirhomme-Fraiture edited the book *Symbolic Data Analysis and the SODAS Software* that was published in the Wiley series in 2008. The book opens a new approach to analyse data with complex units, called symbolic objects. Such objects can be of various types but in this paper, I use countries on one hand, and countries by gender on the other. The variables in symbolic analysis can be more complex than in classic data analysis. In this paper, I use two types of variables, that is, intervals and modals. The former ones are such intervals that well describe the phenomenon concerned, e.g. certain quantiles of a micro variable, or confidence intervals, respectively.

The modal variables are in survey context factually the frequency distributions of micro variables. So, symbolic analysis is a certain type of aggregate approach that is better than the corresponding classic analysis

since less information will be lost. The paper presents this approach in more details but its major part concentrates on the multivariate analysis of 32 European Social Survey countries.

Two types of variables will be used in empirical examples: peoples values and political/trust variables, respectively. Several illustrations are given, using zoom graphics, classification trees and dissimilarity measures. A compact result is given in the form of bi-dimensional mapping. This shows nicely how close or far the symbolic objects (32 countries or 64 countries by gender) are from each other.

## 8.2 Causal Analysis Using Multi-Wave Panel Data: Problems and Solutions

### Coordinators:

Joshua K. Dubrow, Ohio State University (United States)

Kazimierz M. Slomczynski, Ohio State University (United States)

### 8.2.1 The Relevance of Group-Specific Markov Processes to Decomposition of Long-Term Changes in Political Opinions

Irina Tomescu-Dubrow, Polish Academy of Sciences (Poland)

**Abstract** This paper is an extension of my paper *Decomposition of Long-Term Changes in Political Opinions According to Group-Specific Markov Processes* (Tomescu-Dubrow, ASK 2008). In it I used two sources of longitudinal data for Poland, POLPAN and POLTEST, to test the assumption that political opinion change through time is not entirely due to some universal and time-constant processes; rather, it depends on the initial conditions in a persons state. Information on Poles evaluations of the past socialist regime available for repeated intervals, and over a sufficiently long time periodten yearsallowed me to decompose long-term changes in assessment of socialism into short-term change, and the reliability of responses according to group-specific Markov processes. Results demonstrated that Markov-type processes do not have significant explanatory power for long-term change in opinions about socialism. Substantively, this meant that the subjective legacy of the past, namely peoples views of the former regime, matters. The current project extends the analyses by adding information from the 2008 wave of the POLPAN survey, and by dealing with the causal interpretation of this type of results. Focus will be on different processes within different groups.

### 8.2.2 Democratic Values and Views of the Welfare State: A Cross-Lagged Panel Analysis

Sandra T. Marquart-Pyatt, Utah State University (United States)

**Abstract** Previous research suggests that political attitudes are taking shape in unique trajectories in former state socialist countries compared with previous waves of democratic transitions. However, many of these investigations focus on cross-sectional analyses and are thus unable to disentangle contemporaneous and causal effects among a set of political attitudes. This research utilizes POLPAN, a panel dataset collected every five years since 1988, to address the question of how political attitudes develop in synergistic fashion in Poland. Using a cross-lagged panel design, this research examines the relationship between democratic values and support for the welfare state during a period of rapid social change. In comparing factors driving the formation of these attitudes, cleavage, emerging political context, and generational hypotheses are examined. Results point to some support for each of the aforementioned hypotheses. Implications of this research are discussed.

### 8.2.3 Consistency in underemployment: a dynamic analysis

Anna Kiersztyn, University of Warsaw (Poland)

**Abstract** Today, analysing peoples' positions and opportunities on the labour market is becoming an increasingly complex issue. This is due to the raise in atypical employment relations, which provokes new questions concerning changes in the 'quality' of available jobs. Labour economists and sociologists express concern about the raise in underemployment and contingent work, understood mainly in terms of low-quality employment. In this context, some even argue that access to employment, by itself, can no longer be regarded as the major dimension of social inequality in contemporary societies. More detailed characteristics of peoples' jobs have to be taken into account. The aim of this study is to identify the conditions and characteristics which affect the likelihood of the persistence of underemployment and contingent employment over time, using data from five waves of the Polish Panel Survey (POLPAN 1988, 1993, 1998, 2003, and 2008). It is an exploratory analysis, based on dynamic, random effects models estimated using STATA. For the purpose of the study, I define underemployment as low (i.e.: poverty-level) wages and skill mismatch, which occurs when there is a substantial gap between workers educational attainment and their occupational status. Contingent work denotes short-term,



unstable employment relations, with high levels of worker insecurity and lack of attachment between employer and worker. The study is an attempt to overcome various methodological problems, particularly those associated with the operationalization and measurement of employment inadequacies for the purpose of panel analyses, given the limitations of multi-wave panel data.

### 8.2.4 Comparing Youth Migration in the United States: Evidence from the NLSY79 and NLSY97\*\*

Yan Guo, Utah State University (United States)

E. Helen Berry, Utah State University (United States)

Sandra T. Marquart-Pyatt, Utah State University (United States)

**Abstract** Americans are known for their highly mobility, as during the 1990s, the average person in the United States moved about eleven times throughout the course of his life. Youths especially have high mobility rates. Previous research demonstrates that period and cohort effects are important to address with regard to explaining migration across individuals and across time. This research compares youth migration experience (ages 16 to 26 years) in the U.S. using two national longitudinal surveys, NLSY79 and NLSY97, including ten waves of data for each cohort. First, migration rates of youths across these two groups are compared. Next, the factors influencing migration are examined for both age groups across three types of migration: primary, onward and return. We emphasize comparison of what motivates migration across the groups, focusing on a core set of explanatory factors. We also examine key changes occurring in youths life course: changes in marital status, educational attainment and employment as key dimensions of similarity or difference across the two time periods. Features of the longitudinal analysis are highlighted, especially with regard to using person-year periods and the question of repeated observations across individuals. We discuss key findings of the research as pertain to future studies of youth migration.

### 8.2.5 Postmaterialism as a lifetime learning process. A longitudinal analysis of intra-cohort value change in Western Europe

Raul Tormos, Universitat Autònoma de Barcelona (Spain)

**Abstract** Research on value change and stability tends to underline the importance of generational factors. One good example is the theory of postmaterialism developed by Ronald Inglehart. Although Inglehart admits the presence of period shocks as a short-term force, he basically points to the relevance of generational effects. Formative experiences shape different age cohorts through their lifetime, and the values of society shift progressively due to the force of generational replacement. In his many publications he shows empirical evidence of wide and constant generational differences. However if we analyze contemporary time series which cover a wider period (Eurobarometer), we also find dramatic changes inside each generation over time. Then it is possible to talk about an important adult learning process in the field of postmaterialism. It is not just a matter of minor short-term period influences affecting the normal change due to generational replacement, but a systematic intra-cohort tendency linked to the European economic prosperity of the last decades. Once we compare the contribution of generational replacement to the overall change in postmaterialism with the one of the period, we may find out that the latter could be almost bigger than the former. The fact of postmaterislm showing less intra-cohort stability than it should according to Ingleharts theory, can question the socialization hypothesis based on the formative experiences idea. The concept of generation itself could be adjusted to reflect the lifetime learning processes. I perform a longitudinal analysis of the Eurobarometer data, adding different economic system level information from Eurostat. I analyze a group of Western European countries applying time series and age-period-cohort multilevel techniques.

## 8.3 Comparing and Evaluating Autoregressive, Latent Trajectory, Autoregressive Latent Trajectory, and Continuous Time ALT Models (I)

### Coordinator:

Han Oud, Radboud Universiteit Nijmegen (Netherlands)

### 8.3.1 Second-order Stochastic Differential Equation Model as an Alternative for the ALT and CALT Models

Han Oud, Radboud Universiteit Nijmegen (Netherlands)

**Abstract** In the structural equation modelling (SEM) literature, two models for the analysis of longitudinal data become very popular in the past: autoregressive (AR) cross-lagged models and latent trajectory (LT) models. Curran and Bollen (2001) and Bollen and Curran (2004), however, argued that, theoretically, there are many instances when both the processes described by the AR model and the processes described by the LT model are plausible. They proposed the autoregressive latent trajectory (ALT) model, which captures features of both. The discrete-time approach in the ALT model has been criticized by Delsing and Oud (2008), who proposed a continuous time version of the ALT model, using stochastic differential equations, called continuous time autoregressive latent trajectory (CALT) model. In the paper, the linear component appearing in both the ALT and the CALT models will be criticized on several counts. It is shown that most of the problems associated with the linear component are solved by a second-order stochastic differential equation model.

### References:

- Bollen, K.A., and Curran, P.J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research*, 32, 336-383.
- Curran, P.J. and Bollen, K.A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In A. Sayer and L. Collins (Eds.), *New methods for the analysis of change* (pp. 107-135). Washington, DC: American Psychological Association.
- Delsing, M.J.M.H., & Oud, J.H.L. (2008). Analyzing reciprocal relationships by means of the continuous-time autoregressive latent trajectory model. *Statistica Neerlandica*, 62, 58-82.

### 8.3.2 Autoregressive Models: A Comparison of Discrete and Continuous Time Methods

Pascal R. Deboeck, University of Kansas (United States)

**Abstract** Time series involving an autoregressive component and (latent) trend are proving to be challenging to model, as evidenced by discussions about models such as the autoregressive latent trajectory model (ALT) and the continuous autoregressive latent trajectory model (CALT). More fundamentally, the use of models with autoregressive components are still prevalent in psychology as well as many other fields despite the ability to modeling time series using continuous time methods. This paper will examine the modeling of emotion time series using both discrete and continuous time methods. The discrete time models include the application of autoregressive models to intraindividual data which has been altered to mimic differing sampling rates. As is to be expected, differing parameter estimates occur corresponding to the changes in sampling rate. Continuous time methods are then considered, including the method described by Oud & Jansen (2000) for application of the Continuous Time State Equation, and Differential Equation Models as described in articles such as Boker & Nesselrode (2002). These analyses serve the purpose of demonstrating the conflicts that may occur in the literature due to the combination of differing sampling rates and discrete time methods. Consequently, the need for continuous time methods will be highlighted.

### 8.3.3 Three Approaches to Analyze Longitudinal Panel Data: Autoregressive Model, Latent Trajectory Model, and Autoregressive Latent Trajectory Model

Marc Delsing, Praktikon, Radboud University, Nijmegen (Netherlands)  
Johannes A. Landsheer, University of Utrecht (Netherlands)

**Abstract** This presentation addresses the analysis of change in longitudinal panel data, using three different structural equation models: the autoregressive (AR) model, the latent trajectory (LT) model, and the autoregressive latent trajectory (ALT) model. These models are applied to the data of 498 pupils on aggression, measured in 5 annual waves and starting at age 12. Special attention is paid to the interpretation of the various model parameters and the meaning they have for the description and understanding of developmental processes. All three models offer different possibilities to answer questions concerning change or development. Advantages and limitations of the models will be discussed.

### 8.3.4 Extending the CALT Model by Means of SDDE

René Hempel, TU Dresden (Germany)  
Lutz-Michael Alisch, TU Dresden (Germany)

**Abstract** Delsing and Out (Statistica Neerlandica, 2008, 62, 1, 58-82) have analysed models for the investigation of reciprocal relationships, e.g. AR cross-lagged models, LT models and the ALT model, a synthesis of features of both the AR and LT models. They argued that each model has its strengths, of course, but also weaknesses, some of which could be overcome with the CALT continuous-time model. CALT incorporates a continuous-time version of the AR model in terms of an SDE which generates the cross-lagged effect functions. SDEs have Markovian properties and therefore the cross-lagged effect functions have limited memories. In order to explore whether some attention has to be paid on this limitation in panel analysis, we extend CALT using an SDDE approach. To estimate a stochastic delay differential equation, prediction-based estimators (PBE) seem to be optimal (U. Kuchler & M. Sorensen, 2008, Statistical inference for discrete-time samples from affine stochastic delay differential equations; preprint). However, they are computationally much involved. An alternative with mild efficiency loss is the maximum pseudo-likelihood estimator, a particular case of the PBE. We demonstrate our approach using an empirical example and discuss the CALT and CALT/SDDE models.

## 8.4 Comparing and Evaluating Autoregressive, Latent Trajectory, Autoregressive Latent Trajectory, and Continuous Time ALT Models (II)

Coordinator:

Han Oud, Radboud Universiteit Nijmegen (Netherlands)

### 8.4.1 Three approaches to analyzing panel data: Latent growth curve models, autoregressive models and stochastic differential equations

Peter Schmidt, University of Giessen (Germany)

Manuel C. Voelke, University of Mannheim (Germany)

Eldad Davidov, University of Köln (Germany)

Han Oud, Radboud Universiteit Nijmegen (Netherlands)

**Abstract** Autoregressive (AR) models and latent growth curve (LGC) models are two well-established methods for the analysis of panel data. In the past, much research has focused on the comparison of the two approaches (e.g., Raykov, 1998), whereas the idea of combining them so called autoregressive latent trajectory (ALT) models is comparatively new (Bollen & Curran, 2004). For all three models, however, data are assessed at discrete time points and little is known about the robustness of the findings if time points would have been different. The present paper takes up this issue and shows how differential equation modeling provides a solution to this problem by linking the discrete time model parameters to the underlying continuous time model. The different approaches are compared and illustrated using data from the Group-Focused Enmity Survey, which was conducted at three time points between 2002 and 2004.

### 8.4.2 Comparison of a Continuous-Time Autoregressive Model with a Linear Mixed Model Framework

Zita Oravec, Katholieke Universiteit Leuven (Belgium)

Francis Tuerlinckx, Katholieke Universiteit Leuven (Belgium)

**Abstract** In this talk we will compare a hierarchical continuous-time autoregressive model with more traditional modeling approaches such as the linear mixed model (LMM) framework for analyzing intensive longitudinal data. Although both frameworks have been applied for such designs and they both model individual differences, their relation is unclear and has not been thoroughly investigated yet. Also, while the estimation in the LMM framework is quite straightforward and implemented in many general purpose and specific software packages, the estimation of continuous-time autoregressive models can prove relatively cumbersome. We will focus on one particular continuous-time autoregressive model, namely the hierarchical Ornstein-Uhlenbeck (OU) process with measurement error. Our aim is to introduce this dynamical model by interpreting its parameters as related to the mixed models framework. First we will show that under certain conditions, the OU model is equivalent to a linear mixed model. Next, we will demonstrate under which conditions the equivalence relation breaks down. We will conclude that the OU model based dynamical approach is especially relevant when we can assume several sources of inter-individual differences.

### 8.4.3 Growth Mixtures Models with Panel Data: Recent Applications with Adolescents' Delinquent Behaviour

Jost Reinecke, University of Bielefeld (Germany)

Luca Mariotti, University of Münster (Germany)

**Abstract** Growth curve models as structural equation models have been extensively discussed in various fields (e.g., Duncan, Duncan, Strycker, Li & Alpert, 2006; Curran & Muthen, 1999). Recent methodological and statistical extensions have taken into consideration unobserved heterogeneity in empirical data. Muthen extended the classical equation approach to mixture components, i.e. categorical latent classes (Muthen, 2001, 2002, 2004).

The paper will discuss applications of growth mixture models to data from one of the first panel studies in Germany which explore deviant and delinquent behavior of adolescents (Reinecke, 2006a, 2006b). Observed as well as unobserved heterogeneity will be considered with growth mixture models using the program Mplus (Muthen & Muthen, 2006). Special attention will be given to the distribution of the substantive dependent variables as count measures (Poisson distribution, zero-inflated Poisson distribution, cf. Nagin, 2005). Different model specifications with respect to substantive questions will also be emphasized.

#### 8.4.4 The Communication between GP and Patient Represented in a Continuous Non Linear Simulation Model

Cor van Dijkum, University of Utrecht (Netherlands)

Niek Lam, University of Utrecht (Netherlands)

William Verheul, Netherlands Institute for Health Services Research (Netherlands)

Joziën Bensing, Netherlands Institute for Health Services Research (Netherlands)

**Abstract** The communication between a professional and a patient is more complicated than was thought until now in the medical profession. The social aspect of the communication, most of the time not well included in the medical education and sometimes viewed as an obstacle, is nowadays accepted as a component that can facilitate the medical diagnosis. Research done in the GP practice (Bensing 2003) shows how it works. From video observations it became clear that the process to come to an adequate diagnosis depends on the communication skills, both on the GP as well on the patient. Moreover when looking at an ongoing clinical communication process it is evident that different components of the communication do not emerge as simple functions of the participants. Instead the emerging pattern is constantly influenced by the preceding communication. It is a matter of feedback between the two actors of the process, and the interaction of psychic components in the actors themselves that also can be described as (inner) feedback. Those processes can be simulated in continuous simulation models that are described with differential equations. However simple linear differential equations fail to capture that complexity of patient-provider communication. Non linear differential equations, has to serve as a foundation computer model on patient-provider interaction that takes the ongoing causal relationships between communication elements into account. The possibility that such models show that the communication can becoming chaotic and lead to surprises the GP only can anticipate as a surfer on waves, is nowadays discussed in the medical sciences (Suchman 2006). In earlier simulation studies of the authors the non chaotic dynamics of a model of the communication between the GP and patient, including 2 external feedback loops and 3 internal feedback loops, is reproduced with the aid of simulation software. The results of the simulation are compared with quantitative and qualitative representations of video observations of such communications and showed to be correct at face-value (Dijkum et al 2008).

In this paper the non linear chaotic dynamics of such models is further investigated by using Matlab to simulate the chaotic phases of the model, thereby extending studies of mathematical models of 2 coupled processes (Savi 2007) to 2x3 coupled processes.

With this simulation study the exact character of the communication between the GP and patient can be better understood and used as a pattern recognition technique to facilitate the communication.

References:

- Bensing et al (2003). Communication patterns of primary care physicians in the United States and the Netherlands. *Journal of General Internal Medicine*, 2003;18:335-342.
- Savi M., (2007). Effects of randomness on chaos and order of coupled logistic maps. *Physics Letters A*, 364 (2007) 389395.
- Suchman A., (2006). A new Theoretical foundation for relationship -centered care. *Journal of General International Medicine* 2006: 21: S40-44.
- Dijkum, C. van, Verheul W., Lam N., Bensing J. (2008). Non Linear Models for the Feedback between GP and Patients. In Trappl R. (Ed). *Cybernetics and Systems*. Vienna: Austrian Society for Cybernetic Studies, pp. 629-634.

## 8.5 IRT: Item Response Theory in Survey Methodology (I)

### Coordinator:

Wijbrandt van Schuur, University of Groningen (Netherlands)

### 8.5.1 IRT in Survey Methodology. A state of the art overview of the research in the field.

Wijbrandt van Schuur, University of Groningen (Netherlands)

**Abstract** Latent characteristics of respondents, such as degree of trust, religiosity, or job satisfaction, are often measured in survey research by interpreting item responses on an interval scale, based on Likerts (1932) defense of the five-point rating scale as measurement at interval level. This allows the calculation of means, standard deviations and correlations, which in turn are used in Reliability analysis, Factor analysis and Structural Equation Models. For parallel test measurement, items that indicate the same latent trait ideally have the same distribution.

Item Response Theory, starting with ideas of Thurstone (1925) and Guttman (1944), uses the differences in distribution of the items to allow for a measurement of the (dichotomous) items, along with the measurement of respondents. Later developments allowed for the measurement of the response categories of polytomous items. In a Guttman scale the measurement of respondents and items is based on interpreting the response of a respondent as a deterministic dominance relation between respondent and item.

Two types of development will be described: First, probabilistic dominance models have been developed for latent variables at the ratio level (e.g., the one parameter logistic (e.g., the Rasch model, 1960), the ordinal level (e.g., the Mokken scale, 1971), and the nominal level (Latent Class Analysis; e.g., Lazarsfeld & Henry, 1968).

Second, models have been developed in which the response of a respondent is interpreted as a proximity relation between respondent and item. The most important ones are the linear unfolding model (e.g., Coombs, 1950, Andrich and Luo, 1993, and Van Schuur, 1993), and the circular unfolding model (e.g., Leary, 1956, Brown, 1992, and Mokken, Van Schuur and Leeferink, 2001).

Examples of these developments will be given in the other papers in this session.

### 8.5.2 Design and Evaluation of 12 New Items to Measure Spanish National Identity

Eduard Bonet, Universitat Pompeu Fabra (Spain)

Jordi Muñoz, Universitat Pompeu Fabra (Spain)

**Abstract** National identity is an elusive and very difficult to attain and to measure concept. Several attempts have been made to develop a global set of items able to capture the different (and normally presented as opposed) tokens of national identities: civic and ethnic, primordial and constructed, etc. This paper presents an attempt to develop a set of good survey items to measure Spanish national identity, and the results of the evaluation of it. First we introduce the process by which we used Q-Methodology to create a set of 12 items that were included in a survey questionnaire that was used by the Centro de Investigaciones Sociológicas (CIS) to collect data from a representative sample of Spaniards (CIS-2667). Secondly we use that data to explore the dimensionality and the quality of the items proposed using Mokken and Rasch Models.

### 8.5.3 The measure of trust in institutions and its dimensionality: a non-parametric IRT approach

Cristiano Vezzoni, University of Milan (Italy)

Paolo Segatti, University of Milan (Italy)

**Abstract** The analyses of items addressing the measure of individual trust in institutions often result in factor solutions that underline the multidimensionality of the concept. However, little attention is paid to the nature of the items and to their distributions (e.g. dichotomous items, or four point-scales with skewed distributions), characteristics that hardly fulfill the requirements of factor analysis. The paper addresses this problem explicitly and shows how the cumulative nature of the data on institutional trust can be dealt with in a more efficient and convincing way through a non parametric IRT technique, namely Mokken scaling. The dominance model developed by Mokken (1971), considering explicitly the order of items in terms of absolute frequencies of positive answers, helps also to take into account an aspect often removed from the analysis of trust, that is the trustworthiness of specific institutions. The application of Mokken analysis also have conceptual consequences, as it is shown that once the order of the items is taken into account, the measure of trust in institutions can be reduced to one single dimension. The problem of cross-country comparisons is addressed and a systematic procedure to test invariance of item order across country is proposed. The data used to carry out the comparison come from international datasets, namely Eurobarometer and European Value Study.

### 8.5.4 Different methods to test a circumplex scale according to the Holland model of vocational interests

Alexandra Langmeyer, Universität der Bundeswehr München (Germany)

Christian Tarnai, Universität der Bundeswehr München (Germany)

Angelika Guglhör-Rudan, Universität der Bundeswehr München (Germany)

Nadja Pfuhl, University of Dortmund (Germany)

**Abstract** Using the theory of Holland (1997) persons and environments can be characterised by six predominant orientations, which are arranged in a hexagon: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. According to the theory of Holland (1997) persons aspire to environments which correspond with their orientations. Therefore, we can deduce that students consider fields of study from the point of view of their dominant orientations.

In two online surveys (N= 1465 and N=499) we investigated which fields of study students had taken into consideration when selecting their subjects.

We analysed whether the fields of study constitute a circumplex according to the Holland model by means of two different methods: non-linear structure analysis (Nagy, 2006) and non-parametric circular unfolding (Mokken, Van Schuur & Leeferink, 2001). Both models allow parameter estimation of subjects, hence for every person the scale score can be estimated. We found that the data can be depicted by circumplex structures with both models: most appropriate by non-parametric circular unfolding.

References:

Holland, J.L. (1997). *Making Vocational Choices. A Theory of Vocational Personalities and Work Environments* (3rd ed.). Lutz, FL: Psychological Assessment Resources.

Mokken, R.J. Van Schuur, W.H. & Leeferink, A.J. (2001). The circles of our minds: a nonparametric IRT model for the circumplex. In: Boomsma, A, Van Duijn, M.A.J. & Snijders, T.A.B. (eds.), *Essays in Item Response Theory. Lecture notes in Statistics*, Vol. 157, (pp. 339-356), New York: Springer

Nagy, G. (2006). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewahrung im Studium [Vocational interests, cognitive and scholastic abilities: Their role in choice of major and success at university]*. Doctoral thesis, Free University, Berlin, Germany.

### 8.5.5 A flexible modelling framework for the measurement of smartness using Raven's matrices in mass surveys

A. Jan Kutylowski, University of Oslo (Norway)

**Abstract** Nonspecific cognitive capabilities of respondents who participate in mass surveys are commonly considered to influence their opinions and attitudes, separately from other variables such as socioeconomic position, or class or educational attainment, yet are difficult to measure.

Following a brief review of previous approaches to such measurement, employed in survey studies of psychosocial functioning in relation to respondent's location in social structure by Melvin Kohn and his associates, a new method of measurement will be outlined, based on selected Raven's items related to progressive matrices, with emphasis on analysis of the resulting data from a national-level high quality Polish survey (PolPan) conducted in 2003.

The emphasis will be on predicting smartness in a flexible manner as a latent factor that explains joint distribution of indicator responses via nonlinear dependencies fitted via kernel smoothing.

This modelling framework reveals the structure of dependencies of the responses to Raven's items, enabling adjustments that arise in connection with possible nondistinguishability of categories, their nonmetrical character and differential functioning of the items. Robust prediction of smartness will also be addressed, leading to categorical measurement on ordinal scale.

Efficaciousness of the proposed analysis of smartness based on Raven's items will be evaluated by presenting selected results concerning distribution of smartness across educational and income strata, or association of smartness with political opinions.

Keywords: Raven's progressive matrices, smartness, psychosocial functioning, nonparametric latent factor model, mass surveys



## 8.6 IRT: Item Response Theory in Survey Methodology (II)

### Coordinator:

Wijbrandt van Schuur, University of Groningen (Netherlands)

### 8.6.1 Item and scale analysis using CTT and IRT approaches: Effects on scale evaluation and theory testing

Fanney Thorsdottir, University of Iceland (Iceland)

**Abstract** The primary objective of study 1 was to explore whether the conclusions drawn concerning the psychometric properties of Likert-items and scales depended on whether the classical test theory or the item response theory approaches were used to analyse the data. These techniques differ in terms of the requirements they make in relation to the data analysed. Methods based on CTT presume that the data obtained have metric properties whereas the IRT approach does not make this requirement. Those two models were used to analyse a scale measuring attitude towards homeownership and investigated whether conclusions regarding which items to include in the scale - on the basis of their discrimination power - and reliability depended on which of those techniques were used. In regards to the measurement precision of the scale, the findings showed that the marginal reliability of the IRT analysis was considerably higher than the alpha reliability. Moreover, in regards to the discrimination power of the items, the findings showed that the final scale did depend on which statistical technique was used to analyse the discrimination power of the items. Thus the use of the two methods resulted in scales which defined different constructs. In study 2 two different forms of the home-ownership scale were constructed on the basis of findings from the CTT and IRT analysis. The main purpose of study 2 was to explore whether theory testing was affected by which form of the scale was used to measure the relevant construct. The main findings from this study suggest that the choice of a statistical method to analyse Likert-items and scales may affect the substantial conclusions researchers may draw when testing theories which involve the construct these scales measure.

### 8.6.2 Scaling the Checklist of the Dutch Portage Program using IRT Modeling

Margo G.H. Jansen, University of Groningen (Netherlands)

Aafke T. Hoekstra, University of Groningen (Netherlands)

Bieuwe F. van der Meulen, University of Groningen (Netherlands)

**Abstract** Having a young child with learning problems or delayed development is likely to put considerable strain on the parents and many parents will require some kind of outside support. The Portage Program is a home-intervention program for children who have special educational needs. A developmental checklist is used to assess the childrens existing repertoire of skills in six developmental domains. For each domain the items consist of descriptions of often complex behaviors. A basic assumption underlying the checklist is that items are ordered on a developmental continuum which corresponds to the chronological age in case of normally developing children.

The present paper focuses on the psychometric characteristics of the revised version of the checklist in the Dutch setting. Parents of children from 0 to 5 years old who are developing normally, were recruited in several ways and sent a checklist, of which 736 were returned. The children were assigned to groups based on chronological age. For each domain, each age group received a different subset of items, targeted at their supposed developmental level. These item sets are called booklets. Within a booklet the items were arranged in supposed order of difficulty. Each booklet overlapped with the previous and subsequent booklets. Within a booklet item means can be compared directly, but between booklets we have to account for the age differences. Using the booklet overlap we can put the item difficulties on the same scale, but this can be done more easily within an IRT modeling framework. The scaling models we will use are the Rasch model and the one-parameter logistic model (OPLM), both implemented in the OPLM program (Verhelst, Glas & Verstralen, 1995). With OPLM a set of test items can be calibrated on a common scale, and several item oriented statistical tests become available. Moreover the OPLM program is suited to incomplete data designs.

At present the analyses are still in a preliminary stage. We conclude tentatively, that it is possible to find a reasonably fitting scale for each of the content domains, but at the cost of losing a large number of many items. It might still prove to be better to develop separate scales for the age-groups instead of trying to scale all items on the same dimension. Also the item difficulties were in general increasing with the original ordering, but there were many inversions.

### 8.6.3 IRT models in the assessment of the rater effect – the case of Polish external examinations

Henryk Szaleniec, Centralna Komisja Egzaminacyjna (Poland)

Dorota Weziak-Bialowolska, Warsaw School of Economics (Poland)

**Abstract** The aim of the paper is to present Polish experience in the assessment of the rater effect in the external examinations. The rater effect was measured by: (1) examining the dispersion of rater severity/leniency and (2) checking if there was a correlation between the amount of rater effect and chosen traits of raters.

The IRT model namely the Many-Facet Rasch Model was used to explore and measure the rater effect. At first, the difficulty level of tasks was assessed using the real data 10% sample of whole population of students. Then, the abilities of students as well as the levels of severity of raters were quantified with difficulty of tasks anchored. Applied data was obtained from a tailor-made research that was launched by Polish Central Examination Office in 2008. The examiners rated the real exam-sheets in two separate modes: at home paper version of the exam and once in the examinations center using tailored made rating system e-marking. Each task/exam was rated at least five times.

The aim of that research was to verify the hypothesis that the mode of rating (paper exam versus e-marking) influenced the rater severity/leniency. It was also verified whether the mode of rating affects the results.

Additionally it was verified if the rater effect is connected with the aptitude towards change (novelty). This aptitude was measured by several 5-point Likert scale statements and quantified by confirmatory factor analysis.

### 8.6.4 Impact of test score scaling model on regression and multilevel estimates

Maciej Jakubowski, University of Warsaw (Poland)

**Abstract** This paper analyzes how student achievement test score scaling affects final estimates obtained from the regression analysis. While general differences between test score scaling models are well established, little is known about their impact on the analysis of student achievement using linear regression and multilevel models (see Brown et al., 2006, for seminal paper discussing impact of IRT models on basic statistics). The paper explores data from TIMSS 1995 and PISA 2000 international surveys of student achievement. TIMSS and PISA data are analyzed separately not to confound the effect of study design with the impact of scaling methodology. For both surveys scores from two distinct scaling methodologies are available. In TIMSS 1995 student test scores were scaled using the one-parameter and the three-parameter IRT models. In PISA datasets the weighted likelihood estimates and plausible values were made available. While survey organizers advice data users on which model should be preferred and what are the drawbacks of another option, scores from biased methods are widely used in other studies. Thus, our research could be helpful to assess potential risks of using such scores in a secondary analysis. We tested three commonly employed regression models. The linear regression model, the multilevel model, and the quantile regression model. The models were estimated with a set of variables typically used in educational studies which attempts to explain student achievement. Individual and school level regressors were included. The results were presented for all countries and compared for differently scaled scores. Point estimates with their standard errors as well as variance decomposition were discussed. The results suggest that linear and multilevel regression point estimates obtained in a typical analysis are not importantly affected by the choice of scaling method. However, standard errors as well as variance estimates could be heavily biased notably affecting final conclusions. Results of analysis with quantile regression or based on smaller number of observations (e.g. for subgroups) could be also highly affected by a scaling model. The paper points out that researchers should fully understand what are the drawbacks and benefits from a scaling method used to produce scores they wish to analyze. From the other side, survey organizers should clearly describe scaling process and methodology discussing potential risks of using scaled test scores for more demanding examination.

### 8.6.5 Estimating and explaining order (sequence-) effects for items by means of generalized IRT models: I) Benefits and Limitations

Rainer Alexandrowicz, University of Klagenfurt (Austria)

Herbert Matschinger, University of Leipzig (Germany)

**Abstract** Item response models play an important role when studying position (aka order, sequence) effects, i.e. the fact that an item's difficulty is altered depending on its position within an instrument. The LLTM plays an outstanding role for that purpose as it allows for the differentiation between true item difficulty and an additional position parameter. This model usually incorporates CML as the standard estimation method. Rijmen et al. (2003) have shown that the LLTM can be reformulated in terms of a mixed logistic regression model. The current contribution elaborates on the benefits and limitations of this methodological approach: Along with a gentle introduction of the essential idea, several feasible parametrizations and possible extensions (eg. introducing a discrimination parameter sensu 2PL, cross-level-interactions, and quantitative as well as categorical predictors) compared to the classical LLTM are discussed. By means of a simulation study parameter recovery are explicated, different kinds of ordering effects are considered and the question of necessary sample sizes is being discussed.

References:

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A Nonlinear Mixed Model Framework for Item Response Theory. *Psychological Methods*, 8, 185-205.

### 8.6.6 Estimating and explaining order (sequence-) effects for items by means of generalized IRT models: II) A feasibility study

Rainer Alexandrowicz, University of Klagenfurt (Austria)

Herbert Matschinger, University of Leipzig (Germany)

**Abstract** One important, but almost neglected source of bias and/or DIF (differential item functioning) is the order of the items employed to measure the latent construct. Effects on the response for a particular item may result from learning or tiring processes and/or other sources of DIF. This is particularly true for item-sets frequently administered in population surveys, since respondents may be not familiar with the domain under investigation, but rather learn about the very topic in the course of responding to the questions. Order/position effects on the probability of an item may be both item-specific and depend on characteristics of the respondents.

Like other forms of item bias, the effect of the order of items can only be investigated by means of a model, which parsimoniously represents the data at hand. Only one-dimensional models will be considered here. To estimate the order effects a generalized exploratory IRT model is employed (De Boeck & Wilson, 2004) which is formulated as random intercept model with a logit link and a binomial error (Skrondal & Rabe-Hesketh, 2004). Items are considered to be nested within respondents. The effect of the item position (general, or item-specific) on both the item-difficulty and the item-discrimination is investigated employing 1- and 2-parameter IRT models.

A survey in the republic of Germany was conducted to estimate the effects of interest for 5 dichotomous items designed to measure health related quality of life. Each of the possible 120 (5!) orders were administered in a fully balanced design on 1016 respondents. Order effects with respect to the item-difficulties were estimated and evaluated with respect to the latent dimension under consideration. Interaction between item-specific order effects and respondent-specific characteristics like gender are estimated as cross-level interactions. Computations were carried out by the module `gllamm` within STATA 10.1 (StataCorp, 2007).

References:

De Boeck, P. & M. Wilson. (2004). Explanatory Item Response Models: A General Linear and Nonlinear Approach. Edited by Paul De Boeck and Mark Wilson. *Statistics for Social Science and Public Policy*. Ed.: Fienberg, Stephen and Van der Linden, Wim. New York, Berlin: Springer.

Skrondal, A. & S. Rabe-Hesketh. (2004). Generalized Latent Variable Modeling; Multilevel, Longitudinal, and Structural Equation Models. *Interdisciplinary Statistics Series*. Ed.: Keiding, N., Morgan, B., Speed, T., and Van der Heijden, Peter G. M. London, New York: Chapman & Hall/CRC.

StataCorp. (2007). *Stata Statistical Software: Release 10*. Edited by StataCorp LP. College Station, TX.

## 8.7 Marginal models for dependent data

### Coordinators:

Marcel Croon, Tilburg University (Netherlands)

Jacques Hagenaars, Tilburg University (Netherlands)

### 8.7.1 Marginal Models for Dependent, Clustered, and Longitudinal Categorical Data

Wicher Bergsma, London School of Economics (United Kingdom)

Marcel Croon, Tilburg University (Netherlands)

Jacques Hagenaars, Tilburg University (Netherlands)

**Abstract** In the social, behavioral, educational, economic, and biomedical sciences, data are often collected in ways that introduce dependencies in the observations to be compared. For example, the same respondents are interviewed at several occasions, several members of networks or groups are interviewed within the same survey, or, within families, both children and parents are investigated. Statistical methods that take the dependencies in the data into account must then be used, e.g., when observations at time one and time two are compared in longitudinal studies. At present, researchers almost automatically turn to multi-level models or to GEE estimation to deal with these dependencies. Despite the enormous potential and applicability of these recent developments, they require restrictive assumptions on the nature of the dependencies in the data. The marginal models of this talk provide another way of dealing with these dependencies, without the need for such assumptions, and can be used to answer research questions directly at the intended marginal level. The maximum likelihood method, with its attractive statistical properties, is used for fitting the models. This talk is based on a forthcoming book by the authors in the Springer series *Statistics for the Social Sciences*, see [www.cmm.st](http://www.cmm.st)

### 8.7.2 Marginal causal models

Jacques Hagenaars, Tilburg University (Netherlands)

**Abstract** Marginal modeling for dependent observations is often described as a population averaged approach and as such opposed to causal modeling in the form of structural equation modeling (SEM), where the latter can be characterized as an individual-level or conditional approach. It will be shown that principles of marginal modeling enhance the usefulness of SEMs, for dependent but also for independent observations. This will be exemplified by applying SEMs (with latent variables) to categorical data from quasi-experimental designs (the pretest-posttest design and the nonequivalent group design) from family research and from multiple group comparisons where the observations in the groups are dependent.

### 8.7.3 Marginal models and measurement issues

Marcel Croon, Tilburg University (Netherlands)

**Abstract** Marginal models for dependent observations offer opportunities to solve some questions pertaining to the quality of the measurement procedures in the social sciences.

It will first be made clear how marginal models can be applied to test hypotheses about various scalability and reliability coefficients for measurement scales. Up till now only a limited number of such tests based on normality assumptions and approximations were available. Marginal models enlarge the spectrum of statistical tests for hypothesis about scalability and reliability coefficients.

Marginal models also provide means to analyse data from multiple response items in which respondents may choose an arbitrary number of alternatives from a given set of response options. By means of marginal modelling, the relationship between a multiple response item and other variables (including eventually other multiple responses items) can be studied.

### 8.7.4 On the smoothness of intersections of marginal conditional independencies

Tamás Rudas, Eötvös Loránd University (Hungary)

Wicher Bergsma, London School of Economics (United Kingdom)

Renáta Németh, Eötvös Loránd University (Hungary)

**Abstract** Statistical models are often defined by superimposing several simpler restrictions. The resulting models are intersections and one often hopes to derive the properties of the more complex model from that of the simpler ones. In a categorical data setting, smoothness of the resulting model is of fundamental importance. If holds, it implies standard asymptotic behavior when testing using chi-squared related statistics, but is also needed in the interpretation of parameters. The talk will present a general combinatorial condition for intersections of marginal independencies to be smooth. The usefulness of the condition will be illustrated by showing that it holds in problems related to repeated measurements, Markov chains and in certain block recursive models associated with chain graphs.

### 8.7.5 Chain graph models of multivariate regression type for categorical variables

Giovanni M. Marchetti, Università degli Studi di Firenze (Italy)

Monia Lupparelli, Università di Bologna (Italy)

**Abstract** In this work we discuss a class of models for discrete distributions representable by chain graphs. These models, proposed first by Cox and Wermuth (1993) and recently discussed by Drton (2009), have an interpretation distinct both from the classical (Lauritzen, Wermuth, Frydenberg) and the alternative (Andersson, Madigan, Perlman) interpretation. Under a special block-recursive Markov property, these chain graph models have a multivariate regression interpretation.

We propose a parameterization for discrete distributions which is a recursive version of the multivariate logistic regression models of Glonek and McChullagh (1995). This parameterization satisfies the set of independencies under the block-recursive Markov property by imposing linear constraints. We also show that these models are equivalent to marginal log-linear models of Bergsma and Rudas (2002) with linear constraints. Moreover we propose an algorithm for fitting multivariate regression chain graph models based on the procedure developed by Lang (2005). The model is illustrated through an example based on a dataset containing joint responses, intermediate and purely explanatory variables taken from the General Social Survey.

## 8.8 Methodological Issues in Multilevel Analysis for Cross-national Research

### Coordinator:

Bart Meuleman, Tilburg University (Netherlands)

### 8.8.1 Methodological Issues in Multilevel Analysis for Cross-national Research

Bart Meuleman, Tilburg University (Netherlands)

**Abstract** Thanks to the increasing availability of international survey data, research involving cross-national comparisons is burgeoning in various domains of social science. A rapidly increasing number of studies deals with the simultaneous impact of individual and country characteristics on individual-level attributes, such as opinions, attitudes or behaviour. Very often, multi-level techniques are considered as the state-of-the-art analytical tool for this type of research.

At first sight, the choice for multi-level models in cross-national research is far from illogical. Because of its statistical properties, multi-level modelling is appropriate to take the hierarchical structure of cross-national data into account. Citizens are nested within countries rather than independent observations. Furthermore, multi-level models render it possible to study the interplay between variables that are situated at different levels (here: country and individual level).

Yet, despite these desirable characteristics, the application of multi-level models in the domain of cross-national research might not be as straightforward as it seems. Several methodological issues challenge the choice for multi-level techniques. First, the number of participating countries in cross-national surveys does usually not exceed 25 or 30. Simulation studies have shown that similar numbers of higher-order units are too small to guarantee unbiased multi-level estimates (the so-called small-N problem). Second, whether or not a country participates in an international survey is generally not a question of random sampling. Instead, participation depends on factors such as funding possibilities and research interests of the local scientific community. Third, in times of globalisation it can hardly be sustained that countries are units that develop independently, free from external influences. Especially in the case of Europe, the independence that multi-level models assume at the country-level is highly questionable (the Galton problem). For these reasons, it is uncertain to what extent the statistical inference at the country level implied by multi-level models is meaningful.

This session explores the usefulness of multi-level techniques for the analysis of cross-national survey data. The session welcomes papers that deal with the above-mentioned or other problems of multi-level modelling in cross-national research, papers that propose possible solutions to these issues, as well as applications of multi-level in cross-national research.

### 8.8.2 Introducing spatially weighted contextual variables in multi-level analyses of comparative survey data

Guy Elcherth, University of Lausanne (Switzerland)

Dominique Joye, University of Lausanne (Switzerland)

Dario Spini, University of Lausanne (Switzerland)

**Abstract** Survey research or electoral studies have repeatedly shown that political attitudes and behaviour are influenced by the collective economic and social situation, rather than the personal situation of individual citizens. This has led to an increased interest in time-series or comparative analyses of the relationship between collective experiences and public opinion. In this field, multilevel analyses are now widely used as a privileged tool for studying the effect of contextual variables on individual survey responses. The most frequent procedure is to use contextual indicators at the level of entire nations, based on either official statistics, or ad hoc aggregations from micro-level survey. However, these mainstream procedures convey a series of limitations, on both practical and conceptual grounds. The most obvious practical limitation is that the necessarily limited number of higher-level cases (i.e. nations) potentially undermines efforts to develop sensitive, unbiased, and precise statistical models. More problematic even, these traditional approaches imply that procedures for computing new survey-based contextual-level indicators are of particularly little cost-effectiveness, because they require a representative sample for each higher-level case. As a consequence, potential micro-level data sources for creating new contextual-level indicators are limited to a very small number of existing international surveys. This creates a sizeable and, in our

view, partially unnecessary obstacle for creative approaches to contextual data generation. The main conceptual flaws of these procedures derive from the fact that they introduce some rather arbitrary assumptions regarding the relationship between the communitarian backgrounds of collective experiences, and social representations. Why would the national community be the only relevant frame of influence for public opinion? And why would social and economic circumstances have a homogeneous impact on collective experiences within the national community, independently of the particular location of concrete events?

In order to overcome part of these shortcomings, we have started to develop an innovative procedure for generating contextual variables for multi-level analyses, which are spatially weighted. The starting point of our approach is close to geographically weighted regression analysis, which assumes that the effect of an explanatory variable on an outcome variable is a function of the geographical location of the cases (which can be individuals or areas). Extending this logic from physical distance/proximity between survey cases to multifaceted social interdependencies and influences, we suggest a range of continuous proximity functions for defining open-ended social contexts, focusing either on the territorial, symbolic, or practical foundations of social communities. This way, theoretical hypotheses regarding the social dimensions underlying collective experiences can be empirically tested. The conceptual assumptions and analytical procedures of this approach will be introduced. Then, a substantive research application will be presented. On the basis of a large-scale comparative survey across all countries of the former Yugoslavia, we are modelling the effects of collective experiences to violent conflict and massive social exclusion on public opinion within transitional societies. Research outcomes open new avenues for conceptualising social dynamics that shape public opinion at the meso-level of social reality, which are generally inconspicuous in multi-level models relying exclusively on individual-level and country-level indicators.

### 8.8.3 Beyond the “average” respondent: How quantile regression opens up possibilities in cross-national research

Katrin Hohl, London School of Economics (United Kingdom)

**Abstract** The paper explores the use of quantile regression method to cross-national social research. Drawing on data from the European Social Survey, the paper shows how - compared to conventional linear (multilevel) models - quantile regression can extract more information from the same data.

The method complements standard (multilevel) regression analysis in situations where the latter fails to pick up and adequately describe the effect of the explanatory variables on the distribution of the dependent variable.

Conventional regression estimates the effect of an explanatory variable on the mean of the distribution of the dependent variable. Summarizing the relationship between explanatory and response variable in this way is useful and valid if the significance, direction and strength of relationship is not systematically different at the lower, medium or upper end of the distribution or if such differences do exist, they can be deemed unimportant for the substantive research question.

However, this is often not the case. A range of research questions could be better addressed and important cross-national differences captured more adequately by considering differences in the effect of explanatory variables on the variance or the skew of the dependent variable (e.g. effects on inequality and heterogeneity in Y), regression effects on locations on the distribution of the dependent variable Y other than the mean (e.g. the 10% most deprived respondents rather than the mean respondent), and changes in the size and direction of the regression effect as we move along the entire distribution from lowest through to the highest percentile (e.g. the size of the gender gap in income might become smaller as we move from the least, to average and top earning men and women within one country, but not within another country).

Whilst standard regression limits the researcher to the estimation of regression effects on the mean of the response variable, quantile regression allows estimating the effects of explanatory variables on all these other features of the distribution of the response variable.

The paper gives a non-technical introduction to the method. It suggests four diagnostic questions that help the applied researcher identify when quantile regression is a useful alternative to conventional linear regression. Key differences between quantile and linear regression in terms of data requirements, model assumptions, the practicalities of model estimation in STATA, and presentation and interpretation of the results are illustrated using data from the subjective well-being module fielded in Round 3 of the ESS.

The findings from the quantile regression analysis challenge our current understanding of the relationship between life satisfaction and socio-economic circumstances, affect experience and social functioning, and offer further explanation of European national differences in subjective well-being.

### 8.8.4 Measurement inequivalence as a source of useful information. Using multi-level SEM to explain why measurements are inequivalent

Elmar Schlüter, Wissenschaftszentrum Berlin für Sozialforschung (WZB) (Germany)  
Bart Meuleman, Tilburg University (Netherlands)

**Abstract** In recent years, researchers using international survey data have become more concerned with the cross-cultural comparability of their measurements. Before attributes can be validly compared across cultural groups, it has to be shown that they are measured in a sufficiently equivalent way. By now, there is considerable agreement on the concrete operationalization of (various levels of) measurement equivalence and on analytical tools to test for equivalence (i.e. the multi-group CFA framework).

There is less clarity, however, on what one should do when confronted with measurement scales that are not cross-culturally equivalent something that occurs not infrequently in practice. In the literature, a number of strategies for dealing with measurement inequivalence have been put forward. A very promising proposal was made by Poortinga (1989), who suggests to give a substantive interpretation to measurement inequivalence. Inequivalence is then treated as a useful piece of information on cross-cultural differences. Yet, to the best of our knowledge, there have never been elaborated concrete guidelines on how researchers could proceed in a systematic way to interpret inequivalence. Probably because of this lacuna, this potentially very useful strategy is applied only very rarely.

This paper makes an attempt to fill this gap. We show how a relatively new statistical tool, namely multilevel structural equation modelling, can be used to test whether deviations from measurement equivalence can be explained by relevant context variables. Finding such context effects can provide insight on why particular items function differently across cultures. Applications on ISSP data are presented.

References:

Poortinga, Y.H. (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology*, 24, 737-756.

Key words: measurement equivalence, multilevel SEM, cross-cultural research

### 8.8.5 Database developing World: Unprecedented opportunities for comparative multilevel research on developing countries

Jeroen Smits, Radboud Universiteit Nijmegen (Netherlands)

**Abstract** In the last decades, many large-scale representative household surveys have been conducted in developing countries. In these surveys, for millions of individuals, demographic, socio-economic, health and other information has been gathered. At this moment, the data from these surveys cover over 100 countries in all regions of the developing world (Latin America, Caribbean, Sub-Saharan Africa, North Africa, South-East Europe, Middle East, South, Central and East Asia).

In the Database Developing World (DDW) these data sets are brought together, harmonized for a basic set of variables, and enriched with contextual data at the sub-national and national level. In this way, a unique window to the developing world is created that makes it possible to study important processes taking place there on a scale and with a degree of detail that has never been possible before.

The DDW is an open data infrastructure to which continuously new datasets are added. Currently it contains individual/household level data for over ten million individuals living in 1,200 sub-national regions within more than 100 developing countries. Because the information at the household level is supplemented with information about the sub-national and national context, the DDW offers unprecedented opportunities to study how processes at the household level depend on the context in which the households lives.

The major sources of the datasets included in the DDW are Demographic and Health Surveys, UNICEF Multiple Indicator Cluster Surveys, Pan Arab Family Project (PAPFAM) surveys, ILO-IPEC SIMPOC Surveys, and census datasets from the IPUMS project. Most of the context variables at the sub-national level are derived by aggregating from the household surveys.

In my presentation I will discuss the content of the database, the way it is structured, the research, opportunities offered by it and demonstrate the possibilities by presenting the findings of a multilevel analysis of household and context determinants of educational participation in 90 developing countries.



## 8.9 Testing Structural Equations Models

### Coordinator:

William van der Veld, Radboud Universiteit Nijmegen (Netherlands)

### 8.9.1 Finding Nemo, do we really want to? And, do we like Jaws to haunt the beach?

William van der Veld, Radboud Universiteit Nijmegen (Netherlands)

**Abstract** Recently, an alternative procedure has been suggested to evaluate structural equation models. This procedure was necessary because the popular procedures to evaluate model fit have a tendency to find very small and irrelevant misspecifications when a model under consideration is actually a very good model in terms of measurement error and explained variance. On the other hand, for not so good models, even large misspecifications go undetected. This is an awkward situation, good models are rejected (too often) while bad model are accepted (too of-ten). In order to change this situation we should control for type II errors. Recently Saris, Satorra, and Van der Veld (forthcoming) have suggested an alternative procedure, the detection of misspecifications, that will do just that. This procedure is quite laborious, even for fairly simple models. Therefore, we have developed a software program called JRrule, which for each restricted model parameter judges whether it is [1] not misspecified (high power, nonsignificant MI), [2] misspecified (low power, significant MI), [3] maybe misspecified (high power, significant MI), or whether [4] we cannot say anything at all (low power, non-significant MI). JRrule makes these judgements - 1, 2, 3, or 4 - based upon the output of standard SEM software. The advantage of JRrule, is that we immediately look for misspecifications in the model and also control for type I and type II errors. Moreover, in line with Steiger (1990), Browne & Cudeck (1993), and MacCallum, Browne & Sugawara (1996) we also believe that models are always misspecified to a certain level, that is why in JRrule it is also possible to define the size of the misspecifications one believes is irrelevant for all practical purposes. This prevents models being rejected, due to very small (irrelevant) misspecifications. In this presentation we will shortly discuss the shortcomings of the popular global model evaluation procedures and introduce JRrule as an alternative way to test structural equation models.

### 8.9.2 Social inequality in schools: A methodological summary of model testing

Lihong Huang, Norwegian Social Research (NOVA) (Norway)

**Abstract** This paper presents structural equation modeling tests and results of four studies carried out under an analytical framework of social inequality in education, by analysing data of Young in Norway 2002, i.e. a national questionnaire survey conducted in secondary schools in Norway in 2002. First, the paper offers a broad conceptual framework for approaching the analysis of student inequality in school based on an interdisciplinary perspective of student learning and student attainment. Then, it presents the data and analytical approach of using LISREL techniques in the analysis. Finally, it presents the results of the four studies and provides reflections in the methodological approach of these studies.

### 8.9.3 Some important considerations while testing for misspecifications in SEM

Willem E. Saris, ESADE, Universitat Ramon Llull (Spain)

**Abstract** Recently we have made a proposal to directly test for misspecifications in SEM (Saris et al 2009). Although we think that this approach has considerable advantages above the global tests and fit measures, normally suggested, there are also problems connected with this direct method. To mention some: firstly, one will get a large number of suggestions for misspecifications of which some can be significant by chance; secondly the suggestions are conditional on the starting model ; thirdly, the estimates of the modification indices can be highly correlated. These issues play also a role if one wants to use this information to correct the models. It is this type of problems and the solutions of these problems we will discuss in the presentation.

### 8.9.4 Incremental improvement of a 40 group mean and covariance structure model using JRule: keeping track and when to stop

Daniel Oberski, Tilburg University (Netherlands)

**Abstract** In this application, the goal was to estimate the means of three latent traits to do with political trust, each measured by three repetitions in an MTMM experiment. About 20 different countries were to be compared and ranked on their amount of trust in three objects. A comparison needed also be made across two waves of the ESS. This yielded a structural equation model with equality restrictions on about 40 groups.

Many restrictions in the model such as configural, metric, and scalar invariance might be wrong. They may also make a difference for the rank ordering of trust which was the primary focus of this study. In order to detect such mistakes I applied the procedure suggested by Saris, Satorra and van der Veld (forthcoming).

This presentation will focus on the advantages and problems of this approach when so many parameters are involved. I will discuss the use of a simple computer program which allows one to keep track of the changes made in the model and their effect on the estimates of interest; in this case the rank ordering of countries. My suggestion for a stopping criterion will be that the changes made do not affect the conclusions of the analysis.

### 8.9.5 Cross-national comparison of anti-immigrant attitudes: Test of construct equivalence

Marcel Coenders, University of Utrecht (Netherlands)

**Abstract** Due to globalisation and immigration, the ethnic diversity of European countries increases. Issues of multiculturalism and migration policies are high on the political agenda, fuelled by negative immigrant stances among the general public. In this presentation we focus on the public opinion regarding immigrants and ethnic minorities in European countries. That is, we are out to compare countries on their mean level of anti-immigrant attitudes, without explaining why there are differences.

The research field of ethnic relations and anti-immigrant attitudes offers a relevant and interesting application for methods of cross-cultural research. Even more so than in many other research fields, the question of cross-cultural equivalence of anti-immigrant measures is pressing here. Inter-group relations and individuals attitudes towards immigrants are shaped by individual characteristics as well as the specific national context. Countries differ for instance in the ethnic composition, the history of immigration waves, and the national discourse on citizenship and multiculturalism. Hence, can identical survey questions regarding attitudes towards immigrants really be compared across countries?

The main research question is whether or not survey items that are supposed to measure anti-immigrant attitudes really measure anti-immigrant attitudes in the same way across different countries? This is tested using the approach suggested by Van der Veld and Saris (work in progress). That means that we test for measurement equivalence in a bottom-up approach, i.e. we first test for configural invariance, then for metric invariance, and finally for scalar invariance. If the latter type of invariance holds, then it is meaningful to compare means and relationships across countries. The estimation of the factor models is done with LISREL (Joreskog & Sorbom) on data from the European Social Survey. The testing of the factor models is however done with JRule (Van der Veld, Saris, and Satorra, 2008). The results indicate that the set of items in the scales that we studied, i.e. ethnic threat, ethnic distance, and civil rights, can and should be reduced in order to obtain valid, reliable, and comparable measures.



## **Chapter 9**

# **Qualitative Methods**

## 9.1 Advancements in Cognitive Testing Methodology

### Coordinator:

Gordon Willis, National Institutes of Health (NIH) (United States)

### 9.1.1 Quality Criteria for Data Collection in Cognitive Interviewing: Adding Experiences of Qualitative Research

Hennie R. Boeije, University of Utrecht (Netherlands)

Ger Snijkers, Statistics Netherlands (CBS) (Netherlands)

Gordon Willis, National Institutes of Health (NIH) (United States)

**Abstract** Cognitive interviewing has gained importance as a pretesting practice for surveys but it is still unclear how to evaluate this practice. How do we know that pretesting makes survey questions better? What are the criteria to be used to determine the effectiveness of specific qualitative pretesting approaches? In this paper we would like to discuss qualitative questionnaire pretesting from the viewpoint of how qualitative research has been done outside this field. Here we focus on the data collection part of pretesting.

It is acknowledged that cognitive interviewing is a kind of qualitative research, but it brings some specifics to it in at least three ways. 1) Within the interpretive tradition, qualitative researchers strive to find out how people give meaning to everyday life, while in cognitive interviewing relevance is strictly related to the questionnaire. 2) In qualitative research emerging, flexible methods are used, in order to get to know the social worlds of the researched. In cognitive interviewing data collection is in any case structured by the survey questions. 3) The findings of qualitative research consist of detailed descriptions and should lead up to a better theoretical understanding of the phenomenon under study. The aim of cognitive interviewing is clearly to improve the questionnaire.

In pretesting practice quality is addressed from two different sides. One side pleads for further standardization of the interview protocol and the other side seeks the solution in training of interviewers so that they are competent for the job and in the transparency of the interviewing process. In current qualitative research practice, quality criteria crystallize even in the form of checklists to assess all the steps taken in the research process. What can we learn from these efforts to preserve quality in qualitative data collection more generally? We will have a critical look at the available criteria and suggest possible adjustments to fit the specific character of cognitive interviewing as a form of qualitative research.

### 9.1.2 Attempting to Validate Cognitive Testing Methodology

Theresa J. DeMaio, US Census Bureau (United States)

**Abstract** Cognitive interviewing has been an accepted part of the standard practice of pretesting questionnaires in government, non-profit, and commercial settings for almost two decades. Examples of its use (Lessler, Tourangeau, and Salter, 1989; DeMaio and Rothgeb, 1996; Pan et al, 2007) and guidelines for its use (Willis, 2005) appear in the survey literature. Yet the lack of formal evaluation of the method has subjected it to healthy doses of skepticism by members of the academic community (Tourangeau, Rips, and Rasinski, 2000; Conrad and Blair, 2004; Blair, Conrad, Ackermann, and Claxton, 2006). The time and resource constraints of the applied setting do not lend themselves to conducting the type of systematic evaluations that provide advances in the methodological literature. Perhaps, however, other methods of evaluation, such as using benchmarking of cognitively-improved questionnaires against known distributions, can be used to validate the results of cognitive interviews. In this paper, two examples of pretesting conducted by the U.S. Census Bureau are presented along with tentative validation measures. The examples include testing the Identity Theft Supplement for the National Crime Victimization Survey and testing materials that will be mailed to respondents in the 2010 Census.

### 9.1.3 The Use of Cognitive Testing to Understand Social Phenomena

Cynthia S. Robins, Westat (United States)

**Abstract** During the early 19th century, European social observers (Marx, et al.) recognized the critical role played by the construct of individualism in the emergence and instantiation of capitalist production. In the United States, where capitalism has been so integral to our cultural and historical development, the ideology of the individual frequently obscures our ability to recognize the social whole. Because social science research is merely one facet of the larger American political economy, psychological theories (mind, self, cognition) and research methods often have trumped those drawn from anthropology and sociology (culture, context, situated meaning). As cultural distinctions play an ever-increasing role in American society and politics, however, social science research must adapt accordingly. The goal of this presentation is to demonstrate how the methods and findings of cognitive interviewing - theoretically rooted in the science of cognition (memory, retrieval, information processing) - can be made even more robust through the careful application of anthropological and sociological approaches.

The Medicare Consumer Assessment of Health Plans Survey (MCAHPS) is a standardized, self-administered consumer survey used annually by the Centers for Medicare and Medicaid Services (CMS) to rank Medicare plans around the country. In 2002, statistical analyses conducted by CMS evinced differences in health plan ratings across U.S. geographic regions. While the regional differences were of interest, even more compelling was that the region (California) in which beneficiaries gave their plans the lowest satisfaction scores was the same region that received the highest Federal ratings in terms of quality of care. Because plan ratings emerged from individuals completing a standard questionnaire, one approach to solving the puzzle could have been to cognitively test the survey items among Medicare beneficiaries. Yet the statistical data indicated that we were not witnessing a psychological event, but rather a social fact. And Emile Durkheim pointed out more than a century ago:

The determining course of a social fact must be sought among antecedent social facts, and not among the states of the individual consciousness.

The MCAHPS data thus suggested the need for a research approach that could account for the social whole.

This presentation thus has three main objectives: First, the presenter will describe the data collection method, specifically, conducting cognitive testing within the context of a small-group discussion. Secondly, the presentation will describe how socio-cultural theories were used in data analysis to understand the historical and cultural context in which California seniors were assessing their Medicare health plans. A final objective of the presentation is to encourage researchers to recognize the inherent limitations of any field of social science and to borrow techniques and ideas across field boundaries where appropriate.

### 9.1.4 Cognitive interviewing and validation processes of psychological questionnaires

Jose Luis Padilla, University of Granada (Spain)

Miguel Castillo, University of Granada (Spain)

Isabel Benítez, University of Granada (Spain)

**Abstract** Cognitive pre-test methods began to be used in the USA and in Europe in the 80s by survey methodologist to reduce measurement error sources in survey questionnaires. On other hand, few studies have examined the usefulness of the cognitive pre-test methods, more specifically cognitive interviewing, for gathering validity evidence which support the intended use of psychological questionnaires and educational tests.

Since 90s Validity Theory has acquired great importance within the field of psychological and educational testing. Not only has validity a central position in the processes of evaluating tests and questionnaires, but also in the designing those (Standards for Educational and Psychological Testing, AERA, APA y NCME, 1999). Evidence based on the response processes" is one of the sources of validity evidence. The Standards suggests resorting to the empirical and theoretical analysis of the response processes of the respondents in order to obtain evidence about the fit between the construct which test or questionnaire intend to measure and the response process actually put in to practice by examinees.

The proposal on which the present study is based is that cognitive interviewing can provide evidence "based on the response processes" of the respondents, when responding to the items of the psychological questionnaires and tests. The aim of this paper was to illustrate how cognitive interviewing can be used to get validity evidence in psychological testing. Preliminary results obtained by using cognitive interviewing to get response process evidence for a family functioning questionnaire will be presented as an empirical example. The pre-tested questionnaire intends to measure five components of a family function construct: adaptability, partnership, growth, affection and resolving capabilities. The questionnaire consists of 5 items in a rating scale format with 3 response options. The function family questionnaire is included in a health survey questionnaire. 21 cognitive interviews were performed during the cognitive pre-test. The function family questionnaire is included in a national health survey. Validity evidence from cognitive interviewing will be compared with empirical

psychometric evidence of dimensionality of family function questionnaire. Finally, the conditions under which quantitative psychometric evidence and cognitive interviewing results could be complementary will be discussed.

## 9.2 Enhancing Survey Methodology with Qualitative Methods

### Coordinator:

Bojana Lobe, University of Ljubljana (Slovenia)

### 9.2.1 The explorative and qualitative character of the Multiple Correspondence Analysis and the Ascendant Hierarchical Classification

Laurie Hanquinet, Université Libre de Bruxelles (ULB) (Belgium)

**Abstract** In sociology of culture and consumption a major shift from snobs (Bourdieu) to omnivores (Peterson) has been highlighted. If a trend towards eclecticism has indeed been attested, this paper intends to show that the impression of a unidirectional change is the result of the choice of analysis methods. Using more descriptive and qualitative methods such as multiple correspondence analyses (MCA) and ascendant hierarchical classification (AHC) allows us to introduce more nuances in findings. Savage, Gayo-Cal, Warde and Tampubolon notably consider that research about cultural lifestyles must be in-depth explorative in order to circumscribe the various relations between the different aspects of lifestyles. Explanatory methods, although powerful, fail to give a detailed image of the cultural patterns and tend to reduce the description of complex lifestyles to some cultural variables (dependent variables) and to limit their explanations to sociodemographic determinants (independent variables). Moreover this contribution shows how interesting the MCA and a linked ascendant hierarchical classification can be to combine quantitative and qualitative methods. Indeed, the AHC forms different classes from the MCA. To each class are attributed characteristics and individuals. If a way to contact these individuals has been foreseen, it is possible to get a more complete idea of the meaning of this classification through interviews.

From the analyses of the data about visitors of six museums of modern and contemporary art in Belgium (1900 individuals) using the SPAD software and from interviews, this contribution puts forward the suitability of the multiple correspondence analysis to explore the data in a more contrasted way. As a matter of fact, a MCA and a related AHC allow us to investigate whether several cultural patterns coexist contrary to more reductionist methods and to give consistency to them through a classification and an associated qualitative approach.

### 9.2.2 Enhancing surveys with the theory - based approach

Seweryn Krupnik, Jagiellonian University (Poland)

Hanna Uhl, Jagiellonian University (Poland)

**Abstract** The purpose of the presentation is to depict the use of theory-based approach in survey research. Although the approach is applied in public policy analysis, the possible usage is much broader.

It is claimed that theory based approach helps to explain the program's success or failure namely how and why the program achieve or fail to achieve the intended outcomes. It becomes possible because the evaluation process focuses on the examination of the conceptual foundations of the program. Moreover, the method takes into account both the context of the analyzed process and inner causal relationships. The approach is based on realist paradigm which provides stable epistemological grounds for the applied procedures. The paradigm takes advantage of both scientific rigor of positivism and constructivist concern for taking into account diverse perspectives of agents involved in the analyzed process.

Application of the method enhances survey research in a twofold manner. Firstly, it specifies the research questions which are to be answered. Secondly, it identifies the relevant respondents' variety which shall be covered within research design.

The method was applied by authors in many evaluation studies. During the presentation the reference will be mostly made to evaluation of sub-measure 2.2.1 of Polish Sectoral Operational Programme Increase of Enterprises' Competitiveness. The main objective of the SOP ICE is to improve the competitiveness of enterprises established in Poland, operating on the Single European Market.

The approach will be presented as an good practice example of enhancing survey methodology with qualitative methods.

### 9.2.3 The use of vignettes in a quantitative-qualitative approach

Veronika Schmid, University of Marburg (Germany)



**Abstract** In the paper to be presented a vignette based online-study and a hitherto unfinished qualitative follow-up study to analyze the phenomenon of group-focused enmity is discussed. The core idea of group-focused enmity is that different types of prejudice (e.g. xenophobia, islamophobia, anti-semitism, sexism, homophobia) are interrelated and constitute a syndrome.

In both studies the very same vignettes are employed, yet with different research aims: The online-study focuses on testing the assumed relationship of a general value orientation towards adjustment with the phenomenon of group-focused enmity. In the qualitative vignette study the same vignettes are, then, employed in problem-centered interviews with semi-structured questions to interpret the previous data by eliciting the argumentative patterns and subjective theories of the respondents.

In the vignettes employed in both studies a protagonist is involved in a conflict between his/her own need for self-realization and the demands to adjust him-/herself. The story is interrupted and the respondents are asked to rate the sympathy of the protagonist and to give an advice how to behave. Then, the story goes on and the respondent is informed about the protagonist's decision and its consequences. The following factors are systematically varied: the protagonist's affiliation to a majority or minority group, his or her decision for or against adjustment and the consequences of the decision (positive/negative).

The advantages of such a quantitative-qualitative approach will be shown. Furthermore, the benefits of the vignette approach to analyze sensitive concepts in a less personal and, therefore, less threatening way is discussed.

### 9.2.4 Advantages and Challenges of Using Qualitative Methods in Conjunction with Surveys: Case Examples from Studies funded by US Government Agencies

Susan G. Berkowitz, Westat (United States)

Izabella Zandberg, Westat (United States)

**Abstract** Qualitative methods (focus groups, in-depth interviews, open-ended responses) can be profitably employed in conjunction with survey research in different ways and at different points in the design and conduct of a study. By casting light on areas or aspects of an issue of interest to potential respondents, focus groups can help to inform development of the survey questionnaire. Alternatively, carrying out focus groups or in-depth interviews with subsets of survey respondents after the administration of a sample survey may serve to contextualize, deepen and refine the understanding of survey responses. Similarly, affording survey respondents the opportunity to provide some answers in open-ended fashion on a largely close-ended survey instrument may reveal nascent issues or areas of concern that would otherwise go undetected and at the same time, provide the basis for additional response categories or new survey questions in subsequent iterations of the instrument. Ideally, the two types of methods should be mutually informing, operating in a continuous feedback loop involving clarification of basic concepts, revision and refinement of survey instruments, and analysis and interpretation of findings. Unfortunately, this loop is often left open or incomplete in such a way that the full benefits of combined use of both types of methods remain unrealized.

This presentation will present case examples of several different ways in which qualitative methods have been profitably combined with sample surveys in research studies carried out for US federal government agencies, with special emphasis on how in each case the qualitative component helped to enrich, refine, or critique the methodology and/or results of the survey component. We will conclude by discussing particular areas where the feedback loop between the qualitative and quantitative components could, and should, be further strengthened.

### 9.2.5 Using of system of open-ended questions for studying the students' attitudes to innovations of the Bologna process

Elena Stepanova, Russian Academy of Sciences (Russia)

**Abstract** By working at the thesis The success of modernisation of Russian education within the frames of the Bologna process the situation of participation of Russia in the Bologna process at micro-level was cleared up. Namely the perception of the transformational process by its "final" players the students was studied. This part of research is of qualitative type and is constructed on reflective interrogations when the students formulate themselves the statements concerning the variety of modern collisions in the education sphere. The methodology of open-ended questions is used in the research; the analysis is carried out with the support of computer technology VEGA which ensures diverse possibilities of storage, classification and analysis of the set of text files.

The method of open-ended questions represents as the independent type of research uniting the advantages of both qualitative and quantitative approaches. Using the method of open-ended questions the primary presentation of a problem is expressed by respondents themselves in textual actualisations.

In the given research the mastering of modern formats in the system of Russian higher education is studied. The research is conducted in 2007-2008 among the students of five high schools of Russia. The research sample is 150 persons. Over 6000 text judgments on 35 open-ended questions have been received and analysed.

The research is built up on the structured questionnaire with open-ended questions students are offered to formulate the judgments on a set of essential aspects of modern high school training, on the one hand. On the other hand, the researcher in this situation receives the staticized representations of different categories of respondents. As a whole, the file of the information received due to such type of questionnaire constructed on open-ended questions gives enough quantity of primary illustrations. Due to it is possible to make well-founded representations about vital values of respondents, about the value of education and its different aspects for them, about an estimation of quality of that education which they receive, about their knowledge and acceptance of the Bologna process.

The interrogation included the following parts: revealing of the system of vital values of the students and the place of education in this system, estimation of character and quality of high school education, realization of the Bologna innovations in the high schools, perception of the formats of the Bologna process by the respondents. In the questionnaire the wide scale of open-ended questions is used: the object-estimated, general, bipolar, mixed, "essay", and also some of the closed-ended questions. The factor of influence of specificity of high school/speciality is analyzed as an independent research problem.

The empirical research found out the following features of the situation with the Russian education: the value of education does not take leading places in the system of vital values of students, the guidelines on higher education execute mainly tool functions at modern youth, the high schools are essentially differentiated on degree and character of motivation of their students on education and the chosen speciality, the basic spectrum of the academic forms of study is insufficiently used in high schools, the students are frequently not satisfied by the quality of various educational means, the students have no accurate notions on the character of requests and educational formats of the Bologna process. It is revealed that ideas and formats of the Bologna process are familiar a little to students, many of them are evaluated as of little effectiveness. The educative forms introduced through the system of the Bologna process are not considered by students as insistent and attractive. The exception is the attractiveness to students of idea of education abroad.

The technique allowing to estimate the status of education in the system of vital values of an individual, to reveal the perception by users of the educational services and the innovations connected with the Bologna process is developed. The possibilities of wide using of the open-ended questions, the conclusive classification of text judgments and the Russian computer system VEGA for carrying out of textual researches are shown.

## 9.2.6 Combining qualitative and quantitative methodologies in surveys – from the respondents' perspective

Marika Wenemark, Linköping University (Sweden)

**Abstract** Self-administered public health questionnaires have, in Sweden, a long tradition as important means of monitoring populations health. During the last decade the design has been developed in collaboration with experts in qualitative methodology. Rather extensive open-ended questions and follow-up telephone interviews have been added to the traditional questionnaires. The purpose was, initially, to assure that all respondents could give correct and fair descriptions of their health problems. The open-ended questions were inspired by the way questions are asked in qualitative interviews and the way the researcher shows interest in what the informant has to say. In the first trials it was unclear how respondents would react to these questions. Survey literature often recommends avoidance of open-ended questions with the arguments that people don't want to answer open-ended questions or that it is a too burdensome task for many respondents.

This presentation is based on experiences from five different surveys during the years 2002-2008, where different designs have been used. It might shed some new light on some of these truths, i.e. do the respondents want to and have the ability to answer open-ended questions? What types of open-ended questions have been most successful, in terms of data quality and response rates? What are the pros and cons of using open questions from the respondents perspective?

The presentation will focus on what we have learned so far about the respondents reactions to open-ended questions and invitations to follow-up interviews. The presentation will also show examples on the usefulness of combining qualitative and quantitative methods for the interpretation of survey results.

### 9.3 Quantitative and qualitative approaches to validate psychological and educational questionnaires in cross-cultural research

#### Coordinators:

Juana Gómez Benito, University of Barcelona (Spain)

María-Dolores Hidalgo, University of Murcia (Spain)

Jose Luis Padilla, University of Granada (Spain)

#### 9.3.1 Response process of “proxy” versus “self-respondents” to a disability survey questionnaire: Gathering evidence via cognitive interviewing

Miguel Castillo, University of Granada (Spain)

Jose Luis Padilla, University of Granada (Spain)

**Abstract** Proxy respondents often are used in epidemiologic and health survey research when the subjects of interest (i.e., index or direct respondents) can not give information by themselves. Including proxy respondents increases sample size and study power. These benefits must be weighed against possible sources of measurement errors which can undermine the validity of survey questionnaire. The accuracy of proxy-derived versus self-derived data has received attention in the literature; less studied has been possible differences between proxy and self respondent response process.

The objective of this study was to assess the proxy respondent information for a wide range of variables common in disability research and to explore the response strategies of both types of informants. We compared data gathered from proxy respondents with self-respondents by means of cognitive interviewing. In this sense, we analysed sources of error in the response process for questions of a national disability survey. The questionnaire includes questions related to different limitations about daily life activities. 13 direct informants and 13 proxies answered to 11 target questions about the same content for both types of informants. Data were analysed by means AQUAD (Huber, 2008). Results will be discussed in order to the level of agreement between both type of informant, interviewer errors identifying potential disabled people and deviations to expected interpretation of key concepts for each target questions.

#### 9.3.2 Detecting polytomous differential item functioning in a health survey questionnaire

María-Dolores Hidalgo, University of Murcia (Spain)

Juana Gómez Benito, University of Barcelona (Spain)

Jose Luis Padilla, University of Granada (Spain)

Isabel Benítez, University of Granada (Spain)

**Abstract** Including psychological scales and questionnaires in survey questionnaires has become a common practice in the last years. Psychometric concepts and techniques can help survey researchers to improve quality of survey results while particular characteristics of survey research are taken into account. The focus of this paper is the Differential Item Functioning (DIF) analysis. Few studies have been conducted on Differential Item Functioning (DIF) analysis in survey data as compared with DIF research in educational test and psychological questionnaire. When survey purposes involve different cultural and linguistic groups, survey researchers could take advantage of DIF study results in order to make much more solid interpretations of measures rejecting alternative interpretations about lack of functional equivalence. As an empirical example, partial results of polytomous DIF analysis performed on the Spanish version of the General Health Questionnaire will be presented. The Spanish version of the GHQ is included in a national health survey. Polytomous DIF analysis comparing groups defined by a set of demographic variables were conducted by Logistic Regression analysis. Finally, the theoretical convergence between DIF results and evidence provided by qualitative methods will be also discussed.

### 9.3.3 Key issues for a proper adaptation of Psychological Tests from Spanish into Basque

Nekane Balluerka, Universidad del País Vasco (Spain)

Arantxa Gorostiaga, Universidad del País Vasco (Spain)

Xabier Isasi, Universidad del País Vasco (Spain)

P. Huizi, Universidad del País Vasco (Spain)

**Abstract** This work describes the process of adaptation of The Childrens Depression Scale (CDS, Lang & Tisher, 1983) from Spanish into Basque emphasizing some important issues that should be taken into account when adapting psychological tests from Spanish into Basque. The CDS was translated into Basque using the forward-backward translation method. Structural differences between Spanish and Basque languages were tackled in order to achieve functional equivalence between both versions of the test. After translating the test, a pilot study was carried out with 102 subjects (52 males and 50 females) with the aim of analysing if there were comprehension problems with the items that had been difficult to translate due to structural differences between Spanish and Basque languages and cultures. The psychometric properties of the questionnaire were analysed in a sample of 886 participants (432 males and 454 females) aged between 8 and 16. The factor analysis carried out to assess its dimensionality yielded a two-factor solution that accounted for 77% of the variance. These factors corresponded to the Depression and Pleasure dimensions from the original version. Criterion-related validity was tested by means of correlations with scores on emotionality and sincerity scales of the EPQ-J (the Junior Eysenck Personality Questionnaire), which shown to be suitable for both variables. Reliability with regard to internal consistency level and long-term stability was also appropriate.



## **Chapter 10**

# **Comparative Research**

## 10.1 Comparing Social Survey Data in a Global Perspective. The cases of the CSES and the ISSP.

### Coordinators:

Jessica Fortin, GESIS - Leibniz Institute for the Social Sciences (Germany)

Knut Kalgraff Skjåk, Norwegian Social Science Data Services (NSD) (Norway)

### 10.1.1 Comparing Social Survey Data in a Global Perspective. The cases of the CSES and the ISSP.

Knut Kalgraff Skjåk, Norwegian Social Science Data Services (NSD) (Norway)

Jessica Fortin, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** The Comparative Study of Electoral Systems (CSES) ([www.cses.org](http://www.cses.org)) and International Social Survey Programme (ISSP) ([www.issp.org](http://www.issp.org)) are well-established international research projects covering more than 40 countries around the world. While the studies are independent efforts both thematically and in governance, they face shared methodological challenges that are also common to other comparative research projects. In particular, the credibility of scholarly output from such research programmes relies on the ability of the data to facilitate comparison across many countries and cultures. One key requirement in ensuring quality of comparison is providing special attention to equivalence of concepts and measures. This introductory presentation will provide an overview of both the CSES and the ISSP, introduce some of the challenges that the project face in ensuring quality of comparison, and share their current approaches to the issues.

### 10.1.2 Assessing the Quality of Data in Cross-national Research

Jörg Blasius, University of Bonn (Germany)

Victor Thiessen, Dalhousie University (Canada)

**Abstract** Responses to a set of items are usually associated with socio-demographic characteristics such as age, gender, and education, but they are also associated with response styles, such as a tendency to strongly agree or strongly disagree, and with misunderstandings of words and phrases. Further, in different countries the meanings of words, phrases, or sentences may differ even when the translation is of high quality as it yields for the ISSP. Finally, differences in the quality of data often occur, since each country has its own research group and institute responsible for the field work. When analyzing a battery of items, responses are related to the substantive concept and to methodological effects (for example, response sets). Applying multiple correspondence analysis and categorical principal component analysis to an item battery of the ISSP data allows us to assess what part of the responses is due to substantive relationships and what part is attributable to methodological artifacts. Using as an example the items of political efficacy and trust, our findings call into question the substantive relationships between political efficacy and trust and socio-demographic indicators as reported in the literature.

### 10.1.3 An Analysis of Aberrant Response Patterns on Redistribution Attitudes in ISSP Data

Insa Bechert, GESIS - Leibniz Institute for the Social Sciences (Germany)

Jonas Edlund, Department of Sociology, University of Umea (Sweden)

**Abstract** Although there may be disagreement about the qualities of the standards and instruments developed within survey research, few would disagree that the uncertainty in measurement precision is especially pertinent when performing cross-national research.

The variables that are designed to measure popular attitudes towards progressive taxation in the ISSP Role of Government module of 2006 provide an example of this uncertainty for at least some of the countries surveyed. In the vast majority of countries there is a clear pattern indicating support for the concept of progressive taxation in society. However, the samples of a few developing countries display a very deviant pattern with many respondents asking for the poor to carry a higher share of the tax burden. Our research question is whether these outcomes reflect a substantive difference in the support for progressive taxation or if the deviant pattern is

a methodological artefact rather reflecting measurement error. The paper attempts to deal with this issue from a number of angles. The analysis will focus on whether, firstly, the deviant patterns could be caused by sampling procedures or methodological indicator non-equivalence, such as shifts in the meaning of items or categories due to translation problems. Secondly, we will try to investigate whether the reasons behind the outcomes depend on a more substantive level of indicator non-equivalence in terms of different interpretations based on different cultural backgrounds, and country characteristics. Not until these sources of errors have been excluded can we believe that country characteristics on the system level or fundamentally different public attitudes towards progressive taxation have led to the aberrant patterns we observe. The presentation will introduce some first results of various checking analyses.

#### 10.1.4 Voter Turnout Stability in Comparative Perspective

Mikolaj Czesnik, Polish Academy of Sciences (Poland)

Michal Kotnarowski, Polish Academy of Sciences (Poland)

**Abstract** Voter turnout is a core theme for political science. Its dynamics has been puzzling political scientists for a long time, and thus it has been investigated since the very first studies on voting. However, most of these studies have referred to macro-level stability (or instability) of voter turnout; our knowledge about micro-level voter turnout stability is poor and needs further in-depth elaboration. Thus the central aim of this paper is to analyze micro-level voter turnout stability in detail.

Our main research question is what the covariates and determinants of voter turnout stability (on micro-level) are. Most precisely, we want to find out why some citizens are very stable as far as electoral participation is concerned, while others repeatedly transit from voting to abstention or vice versa. We also want to answer the question why in some countries voter turnout stability is a widespread practice, while elsewhere it is a nearly non-existing phenomenon. Hence we offer two types of hypotheses. Firstly, we propose hypotheses linking voter turnout stability with micro-level variables. Secondly, we propose hypotheses linking voter turnout stability with macro-level, institutional characteristics.

Our second principal aim in this paper is to evaluate the quality of the measurements in the CSES, especially cross-nationally and cross-culturally. In analysis of voter turnout stability we must first of all rely on questions about electoral participation of respondents. Meanwhile, over-reporting of voter turnout is a widespread and well-known phenomenon (CSES data is not an exception in this regard), which of course must have an impact on analysis of voter turnout stability. Moreover, there are significant cross-national differences as far as this phenomenon is concerned. Thus our aim is to study cross-national and cross-cultural differences in over-reporting in CSES data thoroughly.

As our theory pertains both to micro-level variables as well as macro-level characteristics, we employ a multi-level research design. We study effects of both micro-level variables and macro-level variables. We also investigate interactions between them and their impact on voter turnout stability. Our hypotheses are tested using CSES data (module 2). Our results indicate that voter turnout stability varies importantly across countries, also in reference to its determinants and covariates. Our central finding is that both micro-level variables and macro-level characteristics have at best mixed effects. Thus we find no evidence for a single pattern of voter turnout stability in modern democracies.

#### 10.1.5 Individual Attitudes Toward Others, Misanthropy Analysis in a Cross-Country Perspective

Natalia Melgar, Universidad de la República (Uruguay)

Máximo Rossi, Universidad de la República (Uruguay)

Tom W. Smith, University of Chicago (United States)

**Abstract** In general, misanthropy has been analyzed taking into account residents of one country and just comparing results. Instead of doing that, we employ 2004 International Social Survey Program dataset and we analyze its determinants in a cross-country model taking into account both individual characteristics and country effects.

Our models show that misanthropy could be explained by some socio-demographic and economic individuals characteristics. For instance, being a woman, education and being married are characteristics that lessen misanthropy while being young; having a low income, being divorced or being self-employed have the opposite effect.



Moreover, in order to capture fix effects, we include dummies per country of residence and most of them result significant in determining misanthropy. This last result indicates that not only individual characteristics matter but also some factors regarding context also play a significant role. Hence, we estimate two additional models including Gross Domestic Product per capita and GINI index. We find that better economic performance reduces misanthropy while inequality hikes it. Finally, we show that there is a strong relationship among our misanthropy ranking (trust in other people) and two corruption perception rankings (trust in public institutions).

### 10.1.6 An Example of Rasch Analysis in a Comparative Setting – Scaling National Pride with ISSP Data

Markus Quandt, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** When the comparability of attitude scales in international surveys is to be assessed, sociologists habitually think of structural equation modelling (SEM) as the method of choice. Alternative methodological approaches such as the Rasch model, however, offer some conceptual and practical advantages of their own. The present paper will demonstrate how Rasch models can be used to investigate measurement equivalence in comparative research settings.

Among survey researchers, the basic Rasch model is (in)famous for rarely fitting real world data. Still, several different routes can be taken to apply Rasch scaling fruitfully in the social sciences even when a straightforward application fails, without giving up too much of their theoretical promise. One of these routes is to simply adopt the Rasch perspective on measurement, and to accept that not every respondent may be measurable with sufficient reliability by the instrument at hand. A second, somewhat different route is to allow for some heterogeneity in each of the population(s) surveyed and to scale different (sub-) populations with similar, but not identically parametrized Rasch models, using manifest criteria to distinguish the populations. Obviously, samples from different nations provide a manifest grouping criterion on which an analysis of Differential Item Functioning can be based. As a third route, with the advent of discrete mixture models, latent subgroups with different item parameter sets can be distinguished in a more exploratory fashion as well.

These strategies will be briefly demonstrated for a well established scale, namely that of general national pride, as measured in ISSP 2003. For the sake of demonstration, the analyses will be limited to a few national samples.

## 10.2 Designing and testing survey instruments for comparative research

### Coordinators:

Brad Edwards, Westat (United States)  
Sue Ellen Hansen, University of Michigan (United States)  
Janet Harkness, University of Nebraska, Lincoln (United States)

### 10.2.1 The perception of distance between Likert's labels: an international comparison.

Ester Macri, Università degli Studi di Firenze (Italy)

**Abstract** The well-known survey instrument created by the psychologist Rensis Likert ( Likert Scale), is still now widely used in social research (in ESS questionnaire we can find various sets of Likert scale). However, this useful instrument has stimulated a great debate centred on the following question: are the category labels perceived as equidistant by respondents? The distance between agree and neither agree or disagree is not always the same between 1 and 2, but usually researchers assign cardinal values to the verbal response options .

One of the answers is that differences in perceiving labels' meaning (and consequently distance between labels) could be related to respondents' social, cultural and linguistic context. The aim of this work is to estimate differences in perceiving distances between the various Likert's labels according to countries and native languages by using ESS data. In this perspective, ESS Likert items have been analysed through a particular analytical approach, Multi Dimensional Scaling.

### 10.2.2 Job Vacancy Surveys in Europe: How different methods cause different data

Anja Kettner, Institute for Employment Research (Germany)  
Michael Stops, Institute for Employment Research (Germany)

**Abstract** Recently most European countries started to collect data on job vacancies by business surveys. These data are highly relevant for economic research and policy consulting. Until now analysis is limited by the lack of data on the total number of vacancies, usually only data on the registered vacancies are available from administrative sources of the national public employment services. However, national surveys need to use comparable definitions and methods to allow country comparisons and to permit the calculation of European data for comparisons with the United States or Asia regarding the Lisbon strategy.

On basis of the German Job Vacancy Survey we discuss, how the number and structure of job vacancies as well as the job vacancy rates differ using different approaches applied in Europe at present. One example is the treatment of outliers, that are either defined as zero (no vacancy), as missing value, or the outlier values are restricted to a specific maximum value or they are not at all treated in a specific way. All methods yield highly different result in the total number of vacancies grossed up to the whole economy.

Our examples and results underline the need to harmonize input (methods) instead of output (data). The last is usually favoured by the European Commission in European statistics. But job vacancy surveys are very special: at the reference date only a small share of firms has one or more job vacancies, about 15 to 20 percent in Germany for instance, in weak economic situations even less. Most firms have no vacancy at the reference day. The applied methodologies therefore highly influence the data output and determine, if national survey data are comparable to each other or not.

### 10.2.3 Computer-Assisted Audio Recording (CARI): Repurposing a Tool for Evaluating Comparative Instrument Design

Brad Edwards, Westat (United States)  
Wendy Hicks, Westat (United States)  
Karen Tourangeau, Westat (United States)  
Lauren Harris-Kojetin, National Center for Health Statistics (United States)  
Abbie Moss, National Center for Health Statistics (United States)

**Abstract** The most common CARI objective (to date) has been to identify potential errors (interviewer falsification, other interviewer errors, response errors), but identification really results from CARI coupled with other tools. This paper discusses a specific application of CARI within a larger system designed to quickly identify both interviewer errors, and potential measurement error resulting from question design and implementation. Using CARI within this larger system provided the necessary tools to identify problems early in the data collection period, diminishing the potential for a negative impact on final data quality. The system has three components: audio-recording selected questions but random subsets recorded during any one interview; behavior coding of recordings; and report generation – analysis tables generated and reviewed weekly. The system provides visibility into both interviewer and question performance; early and frequent review of information on both interviewers/questions, with flexibility in terms of amount sampled/coded and approach towards coding; and quantitative summary of results that can be verified, diminishing some of the caveats with a pure qualitative approach

Our discussion of the specific application is based on experience with a comparative establishment survey. Comparative surveys collect data from multiple groups (people living in different countries, speaking different languages, aging in different cohorts) in ways that support group comparisons. Establishment surveys are often concerned with different types of organizations. In 2007 we used CARI on the establishment-based National Home and Hospice Care Survey (NHHCS), sponsored by the National Center for Health Statistics. The NHHCS was designed to produce nationally representative data on home health care agencies and hospice care agencies. Although both agency types provide health care in the home and their services overlap, they differ in some important ways. Hospices provide palliative end-of-life care; home health agencies provide care for people with a wide range of conditions and functional limitations, for treatment and rehabilitation purposes.

#### 10.2.4 Questionnaire Design for Multipopulation Studies

Janet Harkness, University of Nebraska, Lincoln (United States)

**Abstract** The paper presents an overview of both conventional and less conventional options for the design and realization of survey questionnaires deliberately designed for implementation in multiple cultures and languages. While we aim to provide concrete information on general considerations of comparative instrument design, we also wish to point to and explain preferred strategies and discipline-determined requirements in a variety of fields (educational testing, health research and social and behavioural sciences).

## 10.3 Investigating social change with surveys: problems of comparability, harmonization and cumulation

### Coordinator:

Christof Wolf, GESIS - Leibniz Institute for the Social Sciences (Germany)

### 10.3.1 Investigating social change with surveys: problems of comparability, harmonization and cumulation

Christof Wolf, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** As more and more surveys are conducted either across countries and / or across time the need for strategies and procedures that yield comparable data, i.e. harmonized data, is increasing. Current practices in cross-national surveys vary widely: While, for example, the European Social Survey aims at standardizing the whole data production process as much as possible (input harmonization) other studies, such as the International Social Survey Programme or even more so the Comparative Study of Electoral Systems and the surveys conducted under the auspices of Eurostat rely mostly on defining target measures without prescribing the procedures by which these measures are reached, i.e. ex-ante output harmonization.

Similar problems are encountered when social change is studied on the basis of cumulated cross-sectional surveys. Measures must have to be (ex-post) harmonized; files have to be merged; documentation and metadata have to be created.

The session invites presentations dealing with comparability of survey data and any aspect involved in the harmonization of survey data: harmonization of measures; harmonization of instruments; harmonization of data collection, data editing, data documentation and the creation of appropriate metadata. In addition presentations focusing on evaluating the level of harmonization achieved in a given dataset and measuring the quality of a harmonized variable would be greatly appreciated.

### 10.3.2 Measuring legal marital status

Anne Katrine Mortensen, Norwegian Social Science Data Services (NSD) (Norway)

Hilde Orten, Norwegian Social Science Data Services (NSD) (Norway)

**Abstract** Information regarding a person's legal marital status and cohabitation makes up an important part of the background characteristics covered in most surveys. However, measuring these concepts is not a straightforward task.

Respondents within and across different countries do not always share the same understanding of the concept legal marital status. While the term itself refers to the de jure status of the relationship, some respondents tend to answer about their de facto relationship to a partner. Furthermore, the fact that different partnership forms are recognised as legal in different countries adds to the complexity of the matter.

The European Social Survey (ESS) is an academically driven multi-country general social survey, fielded biannually since 2002. The ESS has made an effort in trying to measure legal marital status and cohabitation separately, as distinguishing between de jure and de facto marital status is required for the exploration of certain research questions.

The ESS is focusing on the measurement of marital status and cohabitation. From its third round, the question on legal marital status was changed to better reflect the diverse legal frameworks in the participating countries. As a result, the approach was changed from input harmonisation to ex ante output harmonisation. However, preliminary analyses indicate that even with this new instrument, misclassifications occur.

We will give a short overview of problems related to the measurement of legal marital status. Our main focus will be on how questions on marital status work in a Norwegian context. This will be investigated through cognitive interviews. In addition to focusing on the ESS questions, we will also explore alternative measurements of this concept.

### 10.3.3 Nationality, Citizenship, Race, Ethnicity and Migration as Socio-demographic Background Variables in Comparative Social Surveys

Jürgen H.P. Hoffmeyer-Zlotnik, GESIS - Leibniz Institute for the Social Sciences (Germany)

Uwe Warner, Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques (CEPS / INSTEAD) (Luxembourg)

**Abstract** All European national states are characterized by reasonable proportions of resident people not belonging to the national majorities in the countries. In Luxembourg about 37% of the resident population are foreigners; in Estonia about 20% are non-nationals; and the EU-25 average across the member states reports about 5% of the population are foreigners in the countries.

In comparative surveys, we are interested in the social stratification of the European countries and the social inequalities of their population. An increasing mobility of people across the national borders raises the probability to include non-nationals in samples for social surveys. Also, the position of foreigners in stratified social structure and the different chances of integration into the host society becomes an important issue in social sciences.

Therefore we need to identify nationality, citizenship, ethnicity and the migration experience of survey respondents for the comparison of the European countries.

We propose a survey instrument that covers four dimensions constituting the ethnical inequality in the societies: The first dimension is citizenship with the main information on the respondents right to participate in the social and political life of the state coded as the interviewees nationality. For non-nationals, this is completed by the residential status the respondent holds.

The second element covers race and ethnicity where we are interested in the self attributed membership of the respondent to the minority groups of the observed population.

The third dimension collects information on the origin of the surveyed person. The country of birth is the central information. In addition the year of arrival in the host society is of interest. In this section of the questionnaire we propose to ask about the naturalization.

The fourth useful variables inform about the use of languages at home; this can be interpreted as an indicator on assimilation to the host culture.

The fifth block on questions collects the information about the migration background.

### 10.3.4 Years of education or educational certificates? Construct validation of cross-national measures of educational attainment in seven European countries

Silke Schneider, Nuffield College (United Kingdom)

Manon de Heus, Tilburg University (Netherlands)

Ruud Luijckx, Tilburg University (Netherlands)

Annick Kieffer, École Normale Supérieure (France)

Ellu Saar, Tallinn University (Estonia)

Carlo Barone, Università degli Studi di Trento (Italy)

Ales Bartusek, Charles University of Prague (Czech Republic)

**Abstract** Cross-nationally comparable measurement of education is vital for comparative social stratification and mobility research. A number of international classifications of education have been designed throughout the years. Despite the wide utilisation of measures of education in cross-national comparative research, validation studies of such harmonised measures are scarce (but see e.g. Braun & Muller 1997, Kerckhoff, Ezell & Brown 2002). This paper intends to show the effects of harmonising country-specific measures of educational attainment in to different cross-national measures, using national labour force surveys from Germany, the Netherlands, Estonia, France, Italy, the Czech Republic and the UK.

The proposed paper firstly discusses the challenges of cross-national comparability and loss of predictive power that results from harmonisation. It then introduces the candidate measures of educational attainment. Among those, the complete ISCED 97 (UNESCO Institute for Statistics 2006 [1997]) is very difficult to implement in large-scale cross-national surveys, and too complex to be used in any statistical analyses. Therefore, different simplified versions of the ISCED-97 will be covered by this study, one of which was recently proposed (Schneider 2008). Here, different simplifications of ISCED 97, CASMIN (Konig, Luttinger & Muller 1988) and years of education will be compared in terms of their predictive power with respect to core social stratification outcomes.

A previous study by the first author (Schneider 2008) which used cross-national survey data (the European Social Survey) showed that most ways of harmonising education variables basically lead to non-comparability of regression outcomes. This study will use fewer countries but larger national data sets with more detailed country-specific education information. This will allow us to check the previous results for countries that provided only crude education variables in the ESS (like the UK and Italy), perform more detailed checks for all countries, and give an independent test of the alternative simplification of ISCED 97 proposed in that study.

The actual validation uses linear and (multinomial) logistic regression models, predicting social status (measured using ISEI scores), social class attainment (EGP or ESeC) and unemployment (ILO definition) by educational attainment. For the single countries, models including the national measure of educational attainment are compared in terms of model fit/explanatory power with models using the different simplified versions of ISCED 97. Regression coefficients will also be compared. It is expected that the use of the simplified ISCED 97 (EULFS and ESS versions) or years of education strongly attenuates the association between education and labour market outcomes in many countries, making cross-national comparisons highly problematic. It is furthermore expected that CASMIN and the alternative simplification of ISCED 97 performs much better.

### 10.3.5 The Value of Educational Certificates in European Countries: A Quantification of the ESS-categories

Heike Schröder, Vrije Universiteit Amsterdam (Netherlands)

Harry B.G. Ganzeboom, Vrije Universiteit Amsterdam (Netherlands)

**Abstract** Measures of educational attainment in cross-national surveys tend to be post-harmonized from indigenous educational classifications, using a common denominator approach such as the International Standard Classification of Education [ISCED]. Another common approach has been to use duration as a common standard. The European Social Survey (ESS) documents how countries have recoded their detailed national education classification into ISCED-97. Besides these two measures, ESS data contain also another indicator, the duration of the entire educational career. We exploit this presence of multiple education measures to assess the relative quality of three measures: (A) the indigenous classification, optimally scored, (B) ISCED and (C) duration. We develop an optimal scale for indigenous classification using a MIMIC approach, that estimates how the respective national education levels are scaled relative to each other with regard to a number of inputs (in particular parental characteristics) and a number of outputs (in particular the acquisition of occupation and partner). We investigate to what extent the increase in detail achieved by optimal scaling has yielded a noticeably better measurement of educational attainment relative to the cruder ISCED representation. We also test a widely held assumption that duration measures are a particularly bad representation of level of education, in particular in countries with tracked system. We conclude that improvement of cross-national measurement of education can best be obtained by using at least two independent indicators, and not by asking more detailed or better questions.

### 10.3.6 Income inequality of EU and Non-EU citizens in Germany. Results on the basis of the Micro-Census from 1973 to 2004

Peter Kriwy, University of Kiel (Germany)

**Abstract** This presentation shall analyze the development of the socioeconomic participation of the Non-German population in Germany during the past 30 years. Evaluations specific for nationality will be introduced for the period from 1973 to 2004 using the German Micro Census. For the analysis of the development of purchasing power differences and income inequality the Big Mac index as well as the Coefficient of Variation and the Gini Coefficient shall be applied.

The results highlight that in general the income inequality is steady in Germany over the past 30 years. Merely between the years 1999 and 2004 the Gini coefficient rises about three percent points. The evaluations concerning nationality prove an increase especially of the income inequality of the households from EU countries up to the level of German households, while households from non-EU countries maintain a clearly lower level of the income inequality. The average income of households from non-EU countries in 2004 is about 1620 euros, households from EU countries earn 1850 euros and German households achieve 2140 euros.

In addition to the presentation of the results, challenges concerning the data handling are discussed. A critical area is the procedure of data aggregation of large data sets using STATA and the calculation e.g. of the GINI coefficient using up to 8 Mio. cases simultaneously.

### 10.3.7 The Changing Economic Situation of Households in Germany

Peter Hartmann, University of Düsseldorf (Germany)

Inga Höhne, University of Düsseldorf (Germany)

**Abstract** Studying the changing economic situation of different types of households across longer time periods typically poses problems of comparability of the underlying household definition. The German Microdata Lab (GML) recently constructed a cumulative file based on German Mikrozensus data from 1962 to 2005. The file is a cumulation of the scientific use versions of the Mikrozensus data, it allows the flexible construction of various indicators of social change in the western part of Germany.

The relative welfare position of different household types is a matter of great political concern. An empirically valid answer to the question of how this position has been changing across time presupposes an output harmonization of the concept of household type. Due to the increase in the number of cohabiting couples, marital status cannot be used as a defining criterion. We have therefore developed a comparable typology with 6 categories based on the presence of children, the presence of a couple and total household size.

A descriptive analysis shows first a large reduction in the proportion of households with more than one adult containing children and secondly increases in the number of single adult and single parent households.

Using the typology we are able to demonstrate a massive historical loss of relative material welfare for single parent households from the 1960s until 2004. In the same time period, dual parent households with children experience some loss of welfare however to a significantly lesser degree.

Childless couples turn out to be the winners of the historical period, they are able to ameliorate their position in comparison to all other groups. These results do not only hold on a purely descriptive level but also controlling for the households labour force participation, education and community size.

## 10.4 Language, Communication, and Research in the Comparative Context

### Coordinator:

Janet Harkness, University of Nebraska, Lincoln (United States)

### 10.4.1 Translation evaluation in the team translation approach

Dorothee Behr, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** In cross-national survey research which is based on the ask-the-same-question model (ASQ) (Harkness et al. 2003) both the quality of the source text and the quality of the resulting translation are of utmost importance for data comparability. To ensure translation quality, special translation and assessment methods have been conceived over the past decades, the team translation approach being one of them. This approach includes, among others, draft translations by several translators and an ensuing team translation review which brings together the translators themselves and survey, design and subject matter experts. The goal of such a review team discussion is to pull together the different fields of expertise that are needed for producing an optimal translation (Harkness 2003).

The International Social Survey Programme (ISSP) is based on an ASQ-model. The source text is finalised in the English language and then translated into the languages of the ISSP participating countries. The ISSP 2007 questionnaire was translated into German according to the above described team translation approach. The review team discussion was recorded and thus could be used for translation process research.

Based on the recording of the team discussion, this paper aims at presenting the evaluation strategies as applied by the team and the challenges and trade-offs that the team experienced between the requirements of comparability across countries and the requirements of appropriate question design in the national context. Findings are related to question design, translation and assessment as well as further research issues in relation to questionnaire translation questionnaire translation.

### References:

Harkness, J. (2003). Questionnaire Translation. In J.Harkness, F. J. R. Van de Vijver & P. Ph. Mohler (Hrsg.), *Cross-Cultural Survey Methods* (35-56). Hoboken, NJ: Wiley.

Harkness, J., Van de Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire Design in Comparative Research. In J.Harkness, F. J. R. Van de Vijver & P. Ph. Mohler (Hrsg.), *Cross Cultural Survey Methods* (19-34). Hoboken, NJ: Wiley.

### 10.4.2 Scale Effects and Order Effects in a Foreign Language Questionnaire

Hilde Tobi, University of Wageningen (Netherlands)

Loes Maas, University of Wageningen (Netherlands)

Tina Glasner, University of Wageningen (Netherlands)

**Abstract** Previous research showed several scale and order effects in questionnaires that may differ between cultures (cf. Haberstroh et al., 2002; Hult et al. 2008). With the internationalization of university education, many universities in non-English speaking countries find themselves in situations where communication takes place in English. When students and staff experiences are investigated the data collection tool is in English as well. Also in international business studies often participants are required to answer an English questionnaire. In these situations English may be not native to any of the communication parties.

The aim of this study is to investigate scale and order effects in English self-administered questionnaires answered by non-native English speakers.

Master students enrolled in research methodology courses in fall 2008 and spring 2009 were randomly assigned to one of the 2 versions of the questionnaire. The questionnaires were administered in class and contained experimental questions and filler questions on student characteristics, course demands and language background.



In total more than 240 students filled in the questionnaire. These students were born in about 40 different countries and mentioned more than 30 different languages when asked what language they spoke with their parents or caretakers.

A scale effect was investigated by comparing scores on the experimental questions enquiring after How successful would you say you have been in your life so far with either an 11-point answer scale ranging from 0 to 10 or one ranging from -5 to +5. The second experimental question on time spent on study showed two sets of answer categories that overlapped only slightly. Questions on satisfaction with your life in general and satisfaction with your life as a student at WUR were asked in a different order to investigate an order effect. Overall results will be presented and different groups of students will be compared.

### 10.4.3 Telephone interviews, Swiss style: patterns of adaptation, for better or worse?

Isabelle Renschler, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Brian Kleiner, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Nicole Fasel, University of Lausanne (Switzerland)

Nicole Schöbi, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

**Abstract** Problem statement: Orally administered interviews conducted in Swiss German speaking regions of Switzerland often diverge from the written German questionnaire (there is no written form of the many Swiss German dialects), and this may compromise standardization, measurement equivalence, and data quality. Until now, there has been no systematic evaluation of the nature and effects of such dialectal adaptation on the data collected in these Swiss German speaking regions of Switzerland.

Purpose: In June and July 2005, a study was conducted in Germany and Switzerland on oral adaptation in telephone surveys. For the current study, we follow up on this previous work by focusing on yet unanalyzed Swiss interviews (based on the German text) to assess the extent to which, the ways in which, and the reasons for which Swiss German speaking interviewers deviate from the scripted German questionnaire at various linguistic levels (i.e., phonetic, morphological, syntactic, discourse). The study focuses on systematic patterns of adaptation and their influence on data quality. The aim is to determine if consistent "corrections" (i.e., patterns of adaptation) by interviewers provide lessons for survey question development and/or interviewer training in such linguistic contexts.

Approach: Twenty five interviews from the European Social Survey were recorded that had been carried out by five Swiss German interviewers following the written German questionnaire. The recordings were recently transcribed, and deviations from the scripted questionnaire were coded according to an elaborate coding scheme. Analyses will focus on variation across interviewers and within interviewers across interviews to identify consistent patterns of adaptation. We will examine the nature and extent of deviations from the script, apparent motivations for such departures, and assessment of whether the adaptations affected the intended meaning of survey items and survey responses.

### 10.4.4 Interpreted telephone surveys in Europe

Janet Harkness, University of Nebraska, Lincoln (United States)

Ana Villar, University of Nebraska, Lincoln (United States)

Yelena Kruse, Knowledge Networks (United States)

Laura Branden, Westat (United States)

Brad Edwards, Westat (United States)

Clarissa Steele, University of Nebraska, Lincoln (United States)

Ying Wang, University of Nebraska, Lincoln (United States)

**Abstract** The paper reports on findings from an exploratory investigation on interaction in interpreted telephone interviews in Russian and Spanish, using a commercial interpreting firm, Westat interviewers and respondents in Russia and Spain. Control telephone interviews with scripted translations of the questions that were interpreted were conducted by interviewers who were native speakers of Russian and Spanish. We focus our report on data from the interpreted interviews and on issues of burden, negotiation (interview metatalk) and of interpreting challenges found in the interviews. The research provides insight into the process and products of interpreted standardized interviews, the roles of participants and the various challenges that arise and may enable us to prepare better for conducting interpreted interviews when no other option is available.

## 10.5 Quality Monitoring Challenges in Cross-National Data Collection (I)

### Coordinator:

Beth Ellen Pennell, University of Michigan (United States)

### 10.5.1 Ex Ante Cross-National Data Collection: Challenges and Guidelines

Beth Ellen Pennell, University of Michigan (United States)

Kirsten Alcsér, University of Michigan (United States)

Janet Harkness, University of Nebraska, Lincoln (United States)

Sue Ellen Hansen, University of Michigan (United States)

**Abstract** This presentation will discuss guidelines for the design and implementation of cross-national data collections with an emphasis on quality assurance, control, monitoring and the development of quality profiles across the survey lifecycle. As the number and scope of studies covering many cultures, languages, nations, or regions have increased significantly over the past decade, there has existed a growing need to provide information on best practices across the multiple aspects of cross-cultural survey design and implementation to ensure the collection of high-quality comparative data. However, there is very little published material on the details of implementing surveys that are specifically designed for comparative research, including what aspects are to be standardized and when local adaptation is appropriate. The Cross-Cultural Survey Design and Implementation (CSDI) guidelines initiative was created to address this gap. Specifically, the aim of the initiative was to develop and promote internationally recognized guidelines that highlight best practice for the conduct of comparative survey research across cultures and countries. The intended audience for the guidelines includes researchers and survey practitioners planning or engaged in cross-cultural or cross-national research. The guidelines cover all aspects of the survey life-cycle and include the following chapters: (1) Study and Organizational Structure; (2) Tenders, Bids and Contracts; (3) Ethical Considerations; (4) Sample Design; (5) Questionnaire Design; (6) Translation; (7) Adaptation; (8) Survey Instrument Design; (9) Pretesting; (10) Interviewer Recruitment and Training; (11) Data Collection; (12) Harmonization of Survey and Statistical Data; (13) Data Processing and Statistical Adjustment; (14) Dissemination of Survey and Statistical Data; and (15) Assessing Quality for Cross-Cultural Surveys. The guidelines can be found at <http://ccsg.isr.umich.edu>.

### 10.5.2 Development of Standardized Documentation on Study Methodology for Cross-National Surveys

Ashley Bowers, University of Michigan (United States)

Rachel A. Orlowski, University of Michigan (United States)

**Abstract** Comprehensive documentation of survey methods is critical in order to appropriately analyze cross-national survey data, assess data quality, and replicate study findings, yet it is an area that has often been neglected (Harkness, 1999; Heath, Fisher, & Smith, 2005; Mohler, Pennell, & Hubbard, 2008; Mohler & Uher, 2003). Born out of a Comparative Survey Design and Implementation (CSDI) Initiative, the Survey Metadata Documentation System (SMDS) is a standardized web-based documentation tool that was developed by the University of Michigan's Institute for Social Research and Gesis-ZUMA, in an attempt to fill this void. This presentation begins with a review of the limitations in survey documentation that have been cited in the methodological literature and experiences with early documentation tools which led to the development of the SMDS. We then present an overview of the SMDS and describe in detail each of its 11 modules. The 11 modules gather metadata on each phase of survey activity as follows: General Project Information (Module 1), Ethics Review (Module 2), Sample Design (Module 3), Questionnaire Development (Module 4), Translation Process (Module 5), Computer Assisted Interview (CAI) Programming/Systems Development and Testing (Module 6), Pretesting (Module 7), Interviewer Recruitment and Training (Module 8), Data Collection (Module 9), Quality Control (Module 10), and Dataset Preparation/Final Report Information (Module 11). We report on initial efforts to implement the SMDS in large-scale, cross-national studies, such as the World Mental Health Survey Initiative. The potential to use this documentation tool to further methodological research in cross-national studies and to improve the design of later waves of repeated cross-sectional surveys and panel studies is explored as we consider the future of metadata documentation in cross-national surveys.

### 10.5.3 Components of Nonresponse and Experience of Interviewer Effects in the ESS

Oliver Lipps, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Alexandre Pollien, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

**Abstract** In the European Social Survey (ESS) response rates are significantly different across both countries and rounds. While between-country differences have been studied extensively elsewhere (Philippens and Billiet 2003, Philippens et al. 2003, Billiet and Philippens 2004, Billiet and Pleysier 2007), only few analyses investigate effects between the rounds (Symons et al. 2008), and none dealt with possible learning effects of survey agencies and in particular interviewers who participated during several ESS rounds.

In the current paper, we analyze possible performance gains from the use of experienced interviewers working for the same survey agency. We investigate the most important components of response (obtaining household contact, target person contact, target person cooperation) in the ESS according to whether they are produced by senior interviewers who are also used for refusal conversion calls, interviewers that have experience from previous rounds, and "usual" interviewers in their first round. We use data from the contact forms of four countries that participated during the three rounds conducted so far, used the same survey agency, and partially the same interviewers. To analyze interviewer effects, we use discrete two-level models.

### 10.5.4 Cross-National Data Collection Certification: A European Case Study

Kirsten Alcser, University of Michigan (United States)

Axel Börsch-Supan, Mannheim Research Institute for the Economics of Aging (Germany)

Grant D. Benson, University of Michigan (United States)

Heidi Guyer, University of Michigan (United States)

**Abstract** The growing prevalence of cross-national surveys has been accompanied by increasing interest in efforts to ensure the consistency of survey quality across countries. This includes proposals to impose rules for questionnaire design and translation, sample selection, and post-survey processing. Curiously, little attention has been paid to streamlining or accounting for the process of data collection itself. This presentation outlines steps taken on one large, longitudinal cross-national survey to measure and document the results of quality assessment on key dimensions of pre-production and production data collection processes. Evaluation of performance will be assessed based on the degree to which country-specific training is modelled upon the training provided via the projects centralized train-the-trainer model, including length of training and topics covered. Data collection processes evaluated will include: mode of data collection, length of field period, interview goal projections and obtainment; quality control measures such as verification of interviews; and sample management and monitoring conducted throughout data collection. Specific project results are compared with predetermined standards. Results will be integrated with the results of the projects ongoing monitoring of the accuracy and the timeliness of the survey data being collected for the current wave of data collection, to produce a Quality Profile for the project overall as well as for each of the member countries data collection operations.

## 10.6 Quality Monitoring Challenges in Cross-National Data Collection (II)

### Coordinator:

Beth Ellen Pennell, University of Michigan (United States)

### 10.6.1 Methods of Data Collection Quality Monitoring in a Cross-National Context

Grant D. Benson, University of Michigan (United States)

Hendrik Jürges, Mannheim Research Institute for the Economics of Aging (Germany)

**Abstract** While monitoring interviewer quality is of importance to all studies, it may be particularly difficult to accomplish consistently in a cross-national context. Cross-national studies often contract with different survey agencies for each country, each with a different incentive structure for interviewers. Thus, while one agency may provide interviewer compensation based on hours reported, another may provide payment based on the number of completed interviews an interviewer submits. Language and cultural constraints make it impractical if not impossible for a centralized agency to carry out traditional quality monitoring, typically in the form of call-backs to participating respondents. While the literature for detecting falsification is relative sparse, there are several recommendations available, including reinterviews / call-backs (Schreiner et. al. 1988), response and non-response rates (Schreiner et. al. 1988) use of time stamps to compare reported and actual completion dates (Bushery et. al. 1999), comparison of response patterns (Murphy et. al. 2004), and the application of Benford's law (Schafer et. al. 2004). Each of these tools is useful, but, except for reinterviews, requires the accumulation of falsified interviews before detection can take place. In this presentation, we suggest that the strategic implementation of specific item-level timings can help to uncover falsifications even at the beginning stages of data collection. We demonstrate the utility of this method with results from both Europe and the United States.

### 10.6.2 Survey Design and Operation in Areas of War Conflict: The Lebanon Wars Surveys Experience

Elie G. Karam, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

Zeina N. Mneimneh, University of Michigan (United States)

Alimee Nasser Karam, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

Caroline Cordahi Tabet, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

John Fayyad, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

Nadine Melhem, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

Mariana Salamoun, Institute for Development, Research, Advocacy and Applied Care (IDRAAC) (Lebanon)

**Abstract** Wars inflict tremendous damage on societies. While, there is a large body of literature dedicated to investigating the impact of war, little attention has been given to the methodological complexities encountered by researchers in conducting wartime surveys. Lebanon has been a scene of wars since 1975. The Institute for Development, Research, Advocacy and Applied Care (IDRAAC) was initially founded in 1982 to study the mental health consequences of war on the Lebanese population and to promote and conduct mental health awareness and community services in Lebanon. As part of its mission, IDRAAC has conducted several surveys at different phases of the war to investigate the short and long-term psychological effects of war on different target populations (adult, adolescents, and children). The series of surveys conducted by IDRAAC provides an opportunity for recognizing the challenges encountered in designing and operating surveys in war settings. Building on the collective experience of the research team, a preliminary framework is presented that describes the particularities of wartime research. The challenges imposed by these war conflict particularities on the survey lifecycle are also illustrated by giving examples from the Lebanon wars studies. Difficulties encountered during sample selection, questionnaire development, interviewers hiring and training, pre-testing, data collection, and post-production phases are discussed. Some future directions about conducting surveys in war settings are suggested. These include adapting wartime surveys to maintain scientific rigor while addressing feasibility concerns, building a

network of researchers and survey units specialized in conflict-related surveys, and establishing specific guidelines for adapting survey design and implementation in conflict settings.

### 10.6.3 Interview Privacy in Cross-Cultural Survey Research: The World Mental Health Survey Experience

Zeina N. Mneimneh, University of Michigan (United States)

Beth Ellen Pennell, University of Michigan (United States)

**Abstract** To encourage accurate responses, many surveys that address sensitive topics require that the interview take place in a private setting so that the interviewer is the only individual to hear the respondents answers. Maintaining such privacy is always a challenge but becomes more so in a cross-national study where there are wide cultural variations in the norms and patterns of social relations and economic and social development that could influence the ability to maintain interview privacy. This presentation focuses on the investigation of interview privacy measures across different cultures in a cross-national survey, the World Mental Health Survey Initiative (WMH). The WMH has conducted nationally or regionally representative surveys in 30 countries around the globe. These surveys are designed to measure the prevalence and correlates of mental health disorders as well as impairment and patterns of help-seeking. Using the WMH data, variations in interview privacy across different cultures are explored. These include measures of who was present during the interview and the duration of this presence. In addition demographic and socioeconomic determinants (including gender of respondent, age, marital status, income, and employment status) are explored. The impact of interview privacy on reporting mental health conditions (such as suicidality, and substance abuse and dependence) as well as other possible sensitive information are then investigated and compared. The presentation ends with practical strategies for addressing interview privacy across cultures.

## **Chapter 11**

# **Fieldwork Monitoring**

## 11.1 Fieldwork monitoring

### Coordinator:

Sarah Hughes, National Opinion Research Center, University of Chicago (United States)

### 11.1.1 Fieldwork Monitoring: the Devil is in the Details

Sarah Hughes, National Opinion Research Center, University of Chicago (United States)

**Abstract** Using survey data takes a certain amount of faith; faith that the questions were adequately tested, faith that the sample was selected correctly, and faith that data were collected using the specified and reported protocols. If all goes as hoped, researchers will have found the Holy Grail of High Quality Data. But, although survey researchers expend considerable time and resources designing protocols, developing instruments, fielding surveys and convincing respondents to participate, much can go amiss. This introduction to the session on Fieldwork Monitoring summarizes key issues in the process of data collection, including fieldwork management structure, hiring and payment of interviewers, the design and use of contact forms and systems as well as the interpretation of the data they produce. Papers in this session share a common focus on monitoring how the interviewer approaches the sample unit. Presenters discuss and evaluate the effectiveness of various tools used to gather or analyse contact data. Depending on their ease of use for interviewers, contact forms can provide process data that reveal patterns of non-response, problematic interviewer behaviour, or uneven sample management. Effective and timely analysis of contact data may allow mitigation or reduction of non-response, improve data quality and lower costs. Data from both single-country and cross-national surveys are discussed in this session.

### 11.1.2 Electronic Fieldwork Monitoring in the “Survey of Health, Ageing and Retirement in Europe”

Barbara Schaan, University of Mannheim (Germany)

**Abstract** The Survey of Health, Ageing and Retirement in Europe (SHARE) has developed an electronic sample management system (SMS) which allows for a realtime monitoring process which is standardized across all participating countries. The SMS was designed by CentERdata, MEA and SRC and programmed by CentERdata. The SMS enables the central management team of SHARE, but also the survey agencies, to analyse a huge variety of fieldwork progress indicators on different levels either in total, by country or by single interviewers. We are able to monitor not only current results of the fieldwork such as changes in the net sample, number of contacted households or number of completed cases, but we can also closely monitor interviewer activities as the SMS stores date, time and outcome of each single contact or contact attempt made by the interviewer. The SMS is installed on the laptop the interviewer uses in order to conduct the interview. The questionnaire itself can only be started from the SMS. Interviewers are supposed to send their data to their survey agency at least every other week. Survey agencies in turn collect the data from the interviewers and bi-weekly sent it to CentERdata. Every two weeks the collected data is uploaded to a central server, where the fieldwork coordination team downloads and analyses the data. Since the data entered into and exported from the SMS is standardized for all countries the SMS facilitates cross-country comparisons. Thus, a large variety of fieldwork indicators can be produced and possible problems can be detected timely which makes it easier to implement strategies in order to enhance the fieldwork quality.

But such a system also has pitfalls. The usage of the SMS requires intensive training with survey agency representatives as well as with interviewers. But still interviewers make mistakes when using the SMS. This can for example be shown by contacts which according to what the interviewer enters have taken place at very unlikely times (in the middle of the night). Interviewers also need to be trained to enter every contact attempt even if the attempt was unsuccessful. Further, the interviewer may additionally have to apply a paper and pencil system, since it is more difficult to electronically note down contact records at the doorstep. There are also adaptation problems (especially in the beginning) since the survey agencies have to work with a system that might differ a lot from the systems they usually use.

### 11.1.3 Fieldwork monitoring using R-indicators

Koen Beullens, Katholieke Universiteit Leuven (Belgium)

Geert Loosveldt, Katholieke Universiteit Leuven (Belgium)

**Abstract** The R-indicator is an instrument that evaluates to what extent a survey is representative of the population under investigation (Schouten & Cobben, 2007). It measures the variation of response propensities, predicted by a set of auxiliary variables that are available for the gross sample. This paper seeks to investigate the usefulness of such a monitoring tool to assess the quality of the contact process, as well as to give some typical examples. It is clear that many agents such as the fieldwork management, interviewers and sample units participate in the fieldwork process, so that many (unintended) decisions are made that possibly affect the representativity of the interviewed sample. Moreover, the contact process can be considered as multi-layered: first a sample unit should be attempted, then contact has to be made, etc. Of course, each of these fieldwork phases may have its own representativity issues. The presentation will focus (1) on the incorporation of paradata in the screening of the sample quality, (2) the development of best practices for the application to data collection monitoring and (3) development of graphical tools for the visualisation of representativity indicators. As an example one may think of the evolution of an R-indicator during the fieldwork. Sudden jumps in the curve may indicate a threat to the quality of the sample.

This paper is part of the RISQ project (Representativity Indicators for Survey Quality) and is financed by the 7th Framework Programme (FP7) of the European Union.

References:

Schouten, B., & F., Cobben (2007) R-indexes for the comparison of different fieldwork strategies and data collection modes.

#### 11.1.4 Investment in Quality: Monitoring Fieldwork on the Spanish Survey of Household Finances

Sarah Hughes, National Opinion Research Center, University of Chicago (United States)

Lauren Doerr, National Opinion Research Center, University of Chicago (United States)

**Abstract** Electronic fieldwork monitoring has become a fixture of survey research. On Spains Encuesta Financiera de las Familias (Spanish Survey of Household Finances), fieldwork monitoring in the third round included computerized sample management and a computer assisted personal interview programmed to include data quality measures. These systems, combined with daily wireless data transmission, permitted near real-time assessment of interviewer adherence to the contacting strategy and to the quality of data collected. This presentation reviews the monitoring strategies employed on the EFF 2008 and presents a case study of how fieldwork monitoring allowed researchers to identify a case of interviewer falsification during the field period. Other uses of the systems for sample management and data quality measurement are also discussed.



## 11.2 Interviewers as Agents of Data Collection (I)

### Coordinators:

Annelies Blom, GESIS - Leibniz Institute for the Social Sciences (Germany)

Achim Koch, GESIS - Leibniz Institute for the Social Sciences (Germany)

### 11.2.1 Interviewer performance and data quality

Geert Loosveldt, Katholieke Universiteit Leuven (Belgium)

**Abstract** With regard to the role of interviewers in the data collection process, some ambiguity can be observed. On the one hand they are considered as essential agents. Their task is complex and their responsibilities wide-ranging. Consequently, they can play an important role in the improvement of data quality. On the other hand, interviewers are also a potential source of systematic and/or variable error. This means that interviewers can facilitate good data quality or can cause error. This difference in point of view is also reflected in two different approaches to survey data quality: Total Quality Management (TQM) and Total Survey Error (TSE). It will be demonstrated that both approaches encourage different types of research regarding the impact of interviewer performance on data quality.

The first data quality approach is Total Quality Management (TQM). TQM is a comprehensive approach with strong emphasis on organizational performance and process quality being prerequisites for optimum quality of output. It will be shown that, in the context of survey interviews and interviewer performance, this approach considers interviewers as an instrument to improve data quality. TQM stimulates research into interviewing techniques, the effects of interviewer training and interviewer behavior during interviews.

The Total Survey Error (TSE) is the second and, in survey methodology, more dominant data quality approach. In this approach a negative definition of quality is used. It is the absence of error. All the different components, actors and activities necessary for survey research are considered as a source of error. Interviewer error is one component of the Total Survey Error. Evaluation of interviewer effects by means of analysis of interviewer variance perfectly fits within the TSE approach. A response and an interviewer oriented evaluation of interviewer effects will be discussed.

An integration of both quality approaches in a Total Survey Data Quality Matrix will be argued.

### 11.2.2 Interviewer Attitudes and Nonresponse in the European Social Survey

Edith De Leeuw, University of Utrecht (Netherlands)

Joop Hox, University of Utrecht (Netherlands)

Annelies Blom, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Interviewers play a key role in contacting and convincing potential respondents and recent research has focused on the role of interviewer attitudes and behaviour on nonresponse (Campanelli et al, 1997, De Leeuw et al, 1997, Groves & Couper, 1998; Hox & De Leeuw, 2002; Stoop, 2003). As part of this effort, several interviewer questionnaires were developed and analysed (for an overview see Hox & de Leeuw, 1998; Hox, de Leeuw, et al, 2002), and the need arose for one standardized questionnaire to stimulate international research. Therefore, IQUEST was developed, a questionnaire for both face-to-face and telephone interviewers. IQUEST incorporated the accumulated knowledge of the earlier interviewer questionnaires, omitting questions that did not work based on both psychometric and substantial analysis. Also new questions were added based on new theoretical and empirical insights. A master version was developed containing context information about the questions; English was used as lingua franca (Harkness et al, 2003). This master questionnaire was checked for clarity and intercultural translatability, and made available to countries participating in the European Social Survey (ESS) 2002. As the ESS 2002 used highly standardized questionnaires and data collection procedures, including strongly standardized field methods and non-response registration, this is an ideal situation to study interviewer effects cross-nationally.

In the present paper, we present the results of this effort. We summarize the results of psychometric analyses on the questionnaires cross nationally, and focus on the effects of interviewers on survey nonresponse cross-nationally.

### 11.2.3 Interviewers as Agents: Agency Theory Applied to Data Collection

Nicolas Sauger, Sciences Po (France)

**Abstract** Interviewers have a central place in collecting survey data. Yet, their role is often overlooked. This presentation proposes firstly to build on the Agency theory developed in economics to better understand how and to what extent interviewers have a potential impact on the quality of data. In economics, the principal-agent problem treats the difficulties that arise under conditions of incomplete and asymmetric information when a principal hires an agent. Various mechanisms may be used to try to align the interests of the agent with those of the principal, such as piece rates/commissions, profit sharing, efficiency wages, the agent posting a bond, or fear of firing. The principal-agent problem is found in most employer/employee relationships. Secondly, this presentation presents actual evaluation of the impact of interviewers on data collection. Besides a few general remarks coming from direct observation of interviewers (during interviews but also during briefings, debriefings and through discussion), variance in interviewers performance is assessed. We build on the possibility to bridge the results of four waves of the ESS, for France, where interviewers identification remains unaltered. This makes it possible to better analyse interviewers performance in terms of response rate, but also to measure the possible influence of interviewers on how people answer during the interview.

### 11.2.4 Monitoring interviewers: a safe way to good data quality?

Ann Carton, Research Centre of the Flemish Government (Belgium)

**Abstract** As Biemer & Lyberg demonstrated in their book *Introduction to survey quality* (2003) survey research has not missed the era of total quality management. Survey research can be viewed as a production process where each step should meet quality standards. Product quality data quality in survey research is achieved through process quality. Process quality depends on systems and procedures that are in place in an organisation.

In survey research using face-to-face interviews as mode of data collection the role of the interviewers is of crucial importance. The paper will focus on two different but linked types of interviewer evaluation. The first type is set up during the period of data collection and focuses in real time - on the main tasks of the interviewers: a) the administrative part of the job; b) contacting respondents and getting participation and c) data quality of registered answers. The second type is a post factum evaluation of the data quality once the fieldwork is completed (e.g. response, item-nonresponse, and interviewer effects on data). Special attention will be drawn on (1) the preconditions to implement the evaluation programme; (2) the roles of the different actors during data collection: interviewers, the organisation responsible for the data collection and the organisation responsible for the whole survey process and (3) cost implications. The different parts will be illustrated with recent data from the survey on Socio-cultural changes in Flanders.

The Research Centre of the Flemish Government has been conducting an annual face-to-face survey on Socio-cultural changes in the Flemish region and in Brussels since 1996 (N= about 1,500 Dutch-speaking respondents, age 18-85 years, stratified two-stage sample design, named individuals from National Register). The survey is a tool to follow-up systematically changes in attitudes, opinions, behaviours of the Flemish people. General attitudes and opinions as well as attitudes and opinions towards the policy of the Flemish government are surveyed. From 2002 on ISSP modules are attached as a drop-off questionnaire. The personal interview contains a fixed part of questions and a variable part with a rotation of attitude scales and themes and lasts about 1h15min. The fieldwork force consists of about 100 trained free-lance interviewers.

A commercial marketing bureau is in charge of the data collection (public tender). Since the survey of 2000 a programme to prepare and to monitor the fieldwork and to evaluate the quality of the data has been set up. The Research Centre of the Flemish Government became responsible for the training of the interviewers, for a close follow-up of the interviewers and for writing standard reports such as a methodological report and an evaluation of the output and process quality.

## 11.3 Interviewers as Agents of Data Collection (II)

### Coordinator:

Frauke Kreuter, University of Maryland (United States)

### 11.3.1 The different roles of interviewers: How does interviewer personality affect respondents' survey participation and response behavior?

Michael Weinhardt, DIW Berlin (Germany)

Frauke Kreuter, University of Maryland (United States)

**Abstract** In household panel surveys, such as the German Socioeconomic Panel Study (GSOEP), interviewers play a prominent role in securing cooperation—cooperation with the survey request itself and continued cooperation throughout the interview. However, only a few studies have focused on the influence of the interviewer's personality on response behaviour. Most studies of interviewer effects on non-response and measurement error have had to rely on data provided by fieldwork agencies to relate interviewer characteristics to respondents' data. In December 2006, a survey of all current interviewers of the GSOEP was conducted (N=586). 94% of all interviewers responded to the 10-page paper questionnaire, including self-rating measures of attitudes, values, beliefs, and personality characteristics in exactly the same question format used in the GSOEP respondent's questionnaires. The interviewer questionnaire also contained a 15-item measure of the 'Big Five' personality traits: openness, conscientiousness, extroversion, agreeableness, and neuroticism. With this data, it is possible to examine effects of the interviewers on survey participation and response behaviour by linking survey data from the interviewers with household and individual level information on respondents. In a multilevel logistic regression model predicting item nonresponse on income questions, extroversion in interviewers increased the chance of item non-response significantly. Further analysis will also look at the interaction effect between respondents' and interviewers' personalities on this measurement error, and will examine the effect of interviewer personality traits on response to the survey request. Results will be discussed in terms of their potential usefulness for recruiting and training of interviewers, highlighting the tension between ideal interviewer characteristics for recruitment vs. those ideal for the interview itself.

### 11.3.2 Does not following fieldwork strategy introduce bias?

Annemieke Luiten, Statistics Netherlands (CBS) (Netherlands)

**Abstract** Field interviewers for Statistics Netherlands are expected to adhere to a fieldwork strategy, that dictates that one of the first two visits to a sample unit should be in the evening, that the first visit should be in the first half of the fieldwork period, that non-contacts should be visited six times, and that visits should be spread over time of day and days of the week. It can be shown that when this strategy is followed, response rates can be up to 25 percentage points higher than when the strategy is not followed. Likewise, the number of visits needed is substantially less when interviewers adhere to the strategy.

Although this is an interesting phenomenon in its own right, more interesting is how this behaviour influences data quality. If there is a difference between addresses where the strategy is followed, with a high response as a consequence, and addresses where the strategy is not followed, with a high chance of nonresponse, chances are that bias is introduced. By linking paradata on fieldwork behaviour to registries with a rich variety of background variables (e.g., ethnic origin, household composition, job holders, income, etc.) and area properties (e.g., number of inhabitants per postal code, number of people of foreign ethnic origin, number of people without work, etc) I aim to study the relation between fieldwork behaviour and sample unit properties, to see if and how interviewers introduce bias by their fieldwork behaviour. The second question I will try to answer, concerns the direction of the causality between following fieldwork strategy, high response rates and costs.

### 11.3.3 The interplay between respondent and interviewer characteristics in predicting survey response

Joakim Dalen, Norwegian University of Science and Technology (NTNU) (Norway)

Øyvind Kleven, Statistics Norway (Norway)

Kristen Ringdal, Norwegian University of Science and Technology (NTNU) (Norway)

**Abstract** In a recent paper, Durrant, Staetsky and Steele (2008) notes that although survey experts have long recognized —the importance of the interviewers in considering surveys errors, little is known about the effects of interviewer characteristics on survey response (or nonresponse). They developed a multilevel logistic cross-classified model that allowed for testing both the effects of household characteristics and interviewer characteristics, as well as distinguishing between interviewer and area effects. This paper follows their lead and focus on the interplay between demographic characteristics of interviewer and of the potential respondents in the Norwegian sample of round 3 of the European Social Survey.

The study is based on the contact data from the round 3 combined with register data on the gross selected sample, and some interviewer characteristics from the administrative records of Statistics Norway. The respondents were selected from the population register by means of simple random sampling. Combined with register data, this allows us to improve upon the analysis of Durrant et al. by using characteristics of the individually sampled persons (sex, age, and education) rather than household characteristics. The information on the interviewers is, however, restricted to a few demographic variables (sex, age, education, and experience) available from the register of Statistics Norway. However, from the contact data, we will develop measures of contact patterns that may vary among the interviewers.

### 11.3.4 CATI Interviewers about their Job

Wojciech Jablonski, University of Lodz (Poland)

**Abstract** The aim of the presentation is to outline different aspects of the job of interviewers performing telephone surveys. Within the presentation I will refer to the results of two methodological studies carried out by myself.

The first study was based on a standardized self-completion questionnaire and was conducted (in 2005) among CATI interviewers working for major Polish research centre. The questions measured the interviewers opinion on following topics: advantages and disadvantages of CATI interviewer job, predispositions to do the job, telephone interviewer training and dealing with respondents refusal-related manners.

The second study is going to be conducted in spring 2009 and will cover CATI interviewers employed in all Polish research agencies which have professional CATI facilities (about 15 companies). The aim of this research is to characterize CATI interviewers as a job category and to point out the limitations of use of telephone interview. The standardized self-completion questionnaire will consist of questions grouped into following modules:

- 1 interviewers motivation for work in a CATI call-centre and their opinions on this job (advantages and disadvantages);
- 2 interviewers opinion on general and project-specific trainings as well as on management system (e.g. communication with supervisors, employment status, pay and benefits, workplace settings and routines);
- 3 problems with conducting interviews using CATI mode, considering:
  - a different types of CATI surveys (e.g. business-to-business & business-to-consumer),
  - b different methodological solutions implemented in CATI instruments,
  - c different stages of CATI interview calling (e.g. scheduling callbacks), introducing the interview (e.g. respondent selection, survey introduction, refusals and refusal conversion) and asking questions (e.g. probing).

In the study I intend to include questions identifying socio-demographic features, psychographic features and chosen psychological variables (e.g. those connected with stress and professional burnout). Additionally, I plan to prepare supplementary questionnaires for CATI supervisors as well as for CATI units managers.

The presentation will be based on the results of mentioned two studies. My intention is to present the most interesting conclusions about CATI interviewers work.

### 11.3.5 Estimating the Value of Project-Specific and Respondent Specific Interviewer Experience: Evidence from Longitudinal and Repeated Cross-Section Surveys

Rupa Datta, University of Chicago (United States)

Ting Yan, University of Chicago (United States)

**Abstract** Interviewer experience can be operationalized in many different ways. One definition is interviewers prior experience working on the same survey; that is, the project-specific experience. Interviewers with such project-specific experience are often given preference in staffing for panel surveys because, at a minimum, these interviewers require less training and are more familiar with project procedures than brand-new hires. In addition, interviewer experience also includes prior experience working with the same respondent/household on the same survey in the context of panel surveys. Presumably, interviewers with respondent-specific experience have an advantage in gaining cooperation from the same respondent/household. In repeated cross-sectional surveys, project-specific experiences are relevant. On panel surveys, however, both project-specific and respondent-specific experience are relevant. This paper investigates the effect of project-specific and respondent-specific experience on interviewer productivity and data quality.

We will examine three surveys conducted by the National Opinion Research Center (NORC) at the University of Chicago. The National Longitudinal Surveys of Youth 1997 Cohort (NLSY97) is a panel survey. The other two the General Social Survey (GSS) and the Survey of Consumer Finance (SCF) are repeated cross-section surveys. By comparing these two types of surveys, the paper will generate estimates of project-specific versus respondent-specific interviewer experience.

Available data permit assessment of interviewer productivity on these surveys in terms of cases completed and costs per case. In addition, data quality comparisons are possible for the NLSY97. For example, we can estimate the effects of project-specific and respondent-specific interviewer experience on item non-response, self-report to sensitive questions, and consistency across rounds.

## 11.4 Non-response bias in cross-national surveys: an evaluation of Designs for detection and adjustment (I)

### Coordinators:

Jaak Billiet, Katholieke Universiteit Leuven (Belgium)

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

### 11.4.1 Nonresponse bias in cross-national surveys: an evaluations of designs for detection ad adjustment

Jaak Billiet, Katholieke Universiteit Leuven (Belgium)

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

**Abstract** Among the range of approaches proposed to estimate nonresponse bias and correct for it (Groves 2006) several of these are applied in past rounds of ESS. These approaches are: comparing sample statistics with population statistics and poststratification weighting; using information from reluctant respondents based on converted refusals; asking a small set of crucial questions at occasion of first contact (and refusal) or in a period after the main survey; and collecting information of the live environment of the sampling units by means of observation; a follow up survey among nonrespondents. Other approaches may be used in other cross-nation surveys. The benefits, costs, challenges, and weaknesses of these approaches in single country surveys or in surveys with a small number are known. What are the specificities, the requirements, and weaknesses in cross-nation surveys covering a large number of countries and aimed to produce data that are useful for cross-nation comparisons and models that include all of the participating countries. We are now in a stage where the theoretical discussions about opportunities and challenges can be completed by real cross-nation data. The contributions of this section are expected to provide critical reflection of approaches to nonresponse bias used in cross-nation surveys and to find solutions for the problems that are met.

### References:

Billiet, J., Phillipens, M., Fitzgerald, R. & Stoop, I. (2007). Estimation of Nonresponse Bias in the European Social Survey: using Information from Reluctant Respondents. *Journal of Official Statistics*, vol. 23 (2), pp. 135162.

Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70, 646675.

### 11.4.2 Nonresponse Adjustments: What can Contact Data Contribute?

Annelies Blom, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** The techniques used for nonresponse bias adjustments in household surveys typically vary from country to country. This is largely due to differences in the availability of auxiliary data for nonresponse adjustments. Whilst some countries have detailed frame data, others have to resort to aggregate information of varying quality. In cross-national surveys these differences in availability of auxiliary information pose a problem for a coherent nonresponse adjustment strategy.

Recently, researchers have started exploring the virtues of contact and neighbourhood data in nonresponse analyses. For these data to be suitable for nonresponse bias adjustments they need to be related to both response propensity and key survey estimates. The European Social Survey (ESS) collects standardised contact data in each participating country, thereby documenting the processes leading to nonresponse across countries. The advantage of such data is that they can be collected in almost every country for both respondents and nonrespondents. Their suitability for nonresponse adjustments, however, is yet to be proven.

This paper looks at whether the ESS contact data can be useful when adjusting for nonresponse bias in key survey estimates. Specifically, it examines whether contact data can contribute to nonresponse adjustments over and above the effect of traditional post-stratification weights for gender, age and education. As a baseline the paper uses the age, gender and education distributions available from frame data in a selection of countries. To this baseline it compares the contribution of weights derived from the ESS contact data with the contribution of preliminary ESS post-stratification weights.

Keywords: nonresponse bias, nonresponse adjustment strategies, auxiliary data, cross-national surveys

### 11.4.3 Measuring and correcting bias based on the European Social Survey (ESS) Round 3 2006-2007 and non-response survey

Hideko Matsuo, Katholieke Universiteit Leuven (Belgium)

Jaak Billiet, Katholieke Universiteit Leuven (Belgium)

Geert Loosveldt, Katholieke Universiteit Leuven (Belgium)

**Abstract** In 2006-2007, non-response survey (NRS) was implemented in the context of European Social Survey (ESS) Round 3 in 4 countries (BE, CH, NO, PL). Its study design was unique making use of the same sampling frame as in the main survey collecting additional information of non-respondents. Focusing analyses on Belgium and Poland, this paper documents measurement of non-response bias through number of socio-demographic characteristics and attitudinal variables between two different groups: cooperative and reluctant respondents on one hand, and cooperative respondents and non-respondents on the other hand. On the basis of core socio-demographic and attitudinal variables between cooperative respondent and nonrespondents, propensity scores are obtained through logistic regression model where these scores are transformed into weights.

Comparing unweighted and weighted survey responses among cooperative respondents, we evaluate the model (propensity scores) by offering different possibilities in calculating these weights.

The paper ends with some discussion and reflection on nonresponse survey design, bias and use of weights in correcting this bias by making references to other possible ways in obtaining weights (ex. observable type of information or additional response units from refusal conversion activities).

Key words: Non response survey, European Social Survey, nonrespondents, propensity scores, nonresponse bias, correction of bias.

## 11.5 Non-response bias in cross-national surveys: an evaluation of Designs for detection and adjustment (II)

### Coordinators:

Jaak Billiet, Katholieke Universiteit Leuven (Belgium)

Ineke Stoop, Social and Cultural Planning office (SCP) (Netherlands)

### 11.5.1 Comparing contact sequences in the ESS and mapping the impact of survey methodology

Alexandre Pollien, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Oliver Lipps, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Dominique Joye, University of Lausanne (Switzerland)

Sylvie Leuenberger-Zanetta, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Marlene Sapin, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

Nicole Schöbi, FORS - Swiss Foundation for Research in Social Sciences (Switzerland)

**Abstract** In this research we examine the sequences of all contacts attempts that interviewers conduct on households to obtain contact and finally cooperation with the target person. Each contact attempt within a sequence is meaningful with respect to the result of the survey request. It's likely that different survey and fieldwork specific characteristics determine the structure of typical contact sequences to a high degree. E.g., countries which sample households rather than individuals are liable to the gatekeeper phenomenon, or certain sampling frames tend to lead to many ineligible cases. Where interviewers are paid per completed interview, the number of appointments is smaller. In addition, survey culture in each country plays a role in the interaction process. That is, common practices to contact respondents and, for potential respondents, manners to react to such requests prevail. The aim of this research is therefore to understand the effects of certain survey specificities on contact sequences rather than to measure the chance of success for an interview.

We use the contact forms of the ESS 2004 and 2006. One focus is to find a suitable measure of similarity of contact sequences. Optimal matching analysis is a technique which allows computing a matrix of dissimilarities between each sequence. Contrary to other classical longitudinal methods this perspective takes into account that actors engaged in the process have knowledge about usual contact sequences and anticipate them.

A matrix of dissimilarities converted by factorisation can map the distribution of the different countries according to their survey specificities. Clustering produces typical sequences for each country and can show the link between the countries fixed survey issues and the interviewers strategies in the fields. Such a map can help to reveal the effective specificities of the fieldwork in different context.

### 11.5.2 The Role of Cognitive Ability in the Relationship between Respondent Reluctance and Response Quality

Olena Kaminska, University of Nebraska, Lincoln (United States)

Allan McCutcheon, University of Nebraska, Lincoln (United States)

Jaak Billiet, Katholieke Universiteit Leuven (Belgium)

**Abstract** Declining response rates and increasing efforts to convert reluctant respondents have led to research interest in the relationship between reluctance to respond and quality of survey answers. According to Krosnick (1991), respondents with lower motivation are more likely to engage in satisficing, i.e. skipping cognitive stages of response process, resulting in lower data quality. Previous research (e.g. Fricker, 2007) has supporting evidence of a negative relationship between the quality of answers and response propensity. The focus of the current study is on whether the relationship between respondent reluctance and satisficing remains significant after controlling for cognitive ability. The study is based on the European Social Survey data (third round) conducted in 2005 / 2006. The comparison of four countries (Belgium, the Netherlands, Norway and Sweden) is implemented. The quality measure is developed using latent class analysis with four indicators of measurement error, including "don't know" responses, "extreme and middle" responses, straight-lining and inconsistent responses. Among two classes found, a class of satisficers is consistently more likely than the other class, optimizers, to provide answers with measurement error. A structural equation model is developed with reluctance and cognitive ability measures



based on call record information, interviewer evaluation indicators and respondent's demographic characteristics. The findings suggest that a negative relationship between reluctance and response quality is wholly mediated by level of cognitive ability. The structural invariance of the model indicates that the negative effect of cognitive ability on reluctance and the positive effect of cognitive ability on response quality have identical magnitudes across the four countries. The findings have important practical implications suggesting that data quality may be improved by decreasing the task difficulty for reluctant respondents.

### 11.5.3 Reasons for refusals and their collection: Lessons learned from content analysis of interviewers' notices in contact protocols in Germany

Natalja Menold, GESIS - Leibniz Institute for the Social Sciences (Germany)

Cornelia Zuell, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Refusals are an important source of non-response in national and cross-national surveys. Strategies to reduce refusals (e.g. appropriate methods to communicate survey purpose, quality of instruments of data collection or refusal conversion strategies) are of higher importance than ever. During field work some surveys collect reasons for refusal in call record data. The European Social Survey even provides such data for secondary research. There is more than one good reason to collect data regarding reasons for refusal: in addition to providing interesting information for research they can also reveal some information about non-respondents which could be used to reduce refusals. But, some questions should be posed on the collection of this information: Do people who refuse state valid reasons or only excuses? Is there some usable information, e.g. for field work coordination? Is there a valid collection of the data? Here two aspects are included: validity of categories used by interviewers and accuracy of data collection done by the interviewer. These questions are handled in this paper. Notices by interviewers regarding reasons for refusals were analysed with the help of content analysis. The data used were provided in call record protocols for the Germany General National Survey (ALLBUS 2008). Prior to this we analysed reasons for refusals in ESS and found that the numbers of undifferentiated outcomes ranged broadly between 0% and 50% across the rounds and countries. Additionally, there were only marginal numbers of several reasons of refusals in numerous countries. ESS applied a category list which included 10 to 13 categories for the collection of reasons for refusals in different rounds. The categories of ESS (blue print and categories applied in Germany) were used as a starting point for the content analysis of interviewers notices in ALLBUS. We found numerous new categories, while the percentage of other was 1.4%. Additionally, less occupied ESS-categories were marginally occupied in ALLBUS data also. On the basis of these results a schema of categories was developed and applied for the further analysis of ALLBUS 2008 data. Here, correlations with socio-demographics of interviewers as well as interviewer variance were taken into account to find any evidence regarding the question of interviewer impact on data collection. Finally, the results were discussed in light of the posed research questions, the usability of developed schema in surveys, as well as in relation to further research on the topic.

Key words: contact protocols, interviewer, instruments, reasons for refusals

### 11.5.4 The EU Statistics on Income and Living: the Problem of Output Harmonization

Hans Schmeets, Statistics Netherlands (CBS) (Netherlands)

Bart Huynen, Statistics Netherlands (CBS) (Netherlands)

**Abstract** The mission of Eurostat the statistical office of the EU is to provide the European Union with high-quality statistical information. For that purpose, it gathers and analyses figures from the national statistical offices across Europe and provides comparable and harmonised data for the European Union to use in the definition, implementation, and analysis of Community policies.

Based on an EU regulation, the member states were requested to collect information on social and cultural participation in an additional module in the EU-survey Statistics on Income and Living (EU-SILC). According to the guidelines in the regulation, only the concepts (definitions) were defined. Consequently, the questionnaires varied substantially between the EU-states. In addition, several designs were used to conduct the information. In many new member states a new survey was introduced, whereas in other countries existing samples, often based on registers, served as a base for collecting the data on social and cultural participation. At least three different interview modes were adopted: face-to-face, by telephone, and by mail. Other aspects varied widely, such as the response rates and the percentage of proxy-interviews.

The presentation deals with the comparison of the 2006 EU-SILC data on social and cultural participation in all EU-member states. It will be demonstrated that:

- a The questions are differently phrased between the countries; in addition, the response rates and number of proxy-interviews differ substantially.
- b The combination of register-countries and telephone interviews results in higher participation rates compared to the combination of survey-countries and face-to-face interviews;
- c Overall, the results show that Eurostat did not meet its own criteria as to highly comparable and harmonised data.



## **Chapter 12**

# **Special Issues related to Survey Research**

## 12.1 Analysing children born of war across time, nations and disciplines (I)

### Coordinator:

Ingvill C. Mochmann, GESIS - Leibniz Institute for the Social Sciences (Germany)

### 12.1.1 Developing a methodology for the research field of “Children born of war”

Ingvill C. Mochmann, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** Children born of war are children who have been fathered by members of occupational or peacekeeping forces and local women. Although it may be assumed that such children are born in every war and conflict that has taken place throughout the history of mankind, only little research on this topic exists. Knowledge available indicates that these children have often been exposed to discrimination and stigmatisations in their families, communities and societies, though. From the methodological perspective this research field comprises several challenges. Firstly, no theoretical model has been developed yet, which includes all aspects relevant to the analysis of children born of war. So far theories and explanations from other research fields have been used to analyse social, psychological, medical, juridical etc. aspects of this particular group. Secondly, it is rather difficult to obtain data as children born of war often belong to the hidden and marginalised population groups. Thus, an estimation of the population size of this group is difficult and representative samples thus difficult to achieve. In addition, the topic is very sensitive. This implies that in the process of acquiring knowledge on the dimension of the problem researchers have to be careful not to uncover mothers and children thereby increasing their stigmatisation and discrimination in family and society. In this presentation a general overview of the state of the art will be given and suggestions for a methodological approach which facilitates the analysis and comparison of children born of war across time and space will be elaborated.

### 12.1.2 A Conspiracy of Silence

Monika Diederichs, Netherlands Institute for War Documentation (Netherlands)

**Abstract** In the Netherlands about 12000-16000 children were born during World War II by Dutch women and German soldiers. After the war stigmatization was the reason that many children fathered by Germans were rejected by their mothers and society. As a result, the children of members of the Wehrmacht in the Netherlands still have the feeling of not being accepted and suffer from serious identity crises. As in the other occupied European countries this topic was taboo for decades. However, in contrast to countries such as, for example, Norway, Denmark and France where it has received wide attention both from media and academia, this topic has found very little attention in the Netherlands so far. Information on the Dutch children born of war is sketchy, and even after newspaper interview with several of these children in the Dutch media between 1990 and 1995 this did not have the same effect such interviews have had in the other countries. This does not mean that the children do not exist and do/did not suffer, however. In this presentation, the different sources available on Dutch children born of war will be presented and challenges related to data collection discussed. Furthermore, the suitability of applying an anonymous questionnaire survey in order to obtain information on the lives of children born of war will be discussed based on experiences with survey among Dutch children fathered by German soldiers in 2001.

### 12.1.3 Sources to cryptic knowledge of paternity and maternity during and after the German occupation of Denmark

Arne Øland, Danish War Child Association (DKBF) (Denmark)

**Abstract** From time immemorial ignorance of maternity and particularly of paternity has been commonplace in European societies. Ancestry was always associated with blood. In the first half of the 20th century the introduction of the classification of blood groups was seen as a promising tool of evidence in paternity cases. The political reasons were equivocal. In Denmark e.g. the eminent social democrat politician and later Minister of Justice, K.K. Steincke, got seven new laws passed in 1937 concerning illegitimate children in order to diminish their mortality, a mortality which allegedly was the double of the mortality among legitimate children. On the

one hand the manoeuvre succeeded, because the maintenance payments secured a certain living standard for the children, while it failed on the other hand, because the identities of the children were neglected later on. In this presentation different sources of knowledge on paternity and maternity in the Danish case will be presented and discussed: 1. Traces of concealment. 2. Gossip 3. Traces, marks and clues 4. Testimonies 5. Evidence in the parish register 6. Evidence in Statstidende 7. Evidence in the police archives 8. Interrogations by Modrehjlpn 9. Paternity-cases in the Vergleich Kommission 10. Notifications in Standesamt I in Berlin or at local courts in Germany 11. General negotiations between Danish and German authorities 12. Evidence by dna-fingerprinting.

### 12.1.4 The forgotten generation: Children of Soviet Soldiers and Austrian/German Women

Barbara Stelzl-Marx, Ludwig Boltzmann Institut für Kriegsfolgen-Forschung (Austria)

**Abstract** In all four zones in Germany and Austria children of occupation were born after the Second World War: as a result of voluntary sexual encounters between local women and foreign occupation-soldiers, but also as a consequence of rape. Often they were regarded as children of the enemy, who were together with their mothers frequently discriminated against. After all they had the wrong father. Thus many of them were confronted with racial, ideological and moral prejudice. "Russenkind" ("Russian child") or "Russenbalg" ("Russian bastard") were common abusive words up to the 1960s.

Many of the children of occupation and their children have been in search for the father for several decades. This is linked to the wish of exploring ones own identity and looking for ones personal roots. Even children who had been born as a consequence of rape have dealt with this vital issue. In some cases, on the other side, former soldiers of occupation have tried to find their children and the mothers in the country where they had been stationed. In the Russian Federation the euphemism children of liberation has emerged in this context. In the case of a family reunion, the drama, emotions and joy can hardly be exceeded.

## 12.2 Analysing children born of war across time, nations and disciplines (II)

### Coordinator:

Ingwill C. Mochmann, GESIS - Leibniz Institute for the Social Sciences (Germany)

### 12.2.1 Case Study: Children born of American soldiers and local mothers: Comparison between the UK and Germany

Sabine Lee, University of Birmingham (United Kingdom)

**Abstract** During and after the Second World War, American soldiers were stationed in Europe, most notably in Great Britain during the period of preparation for the opening of the Second Front in 1944, and in Germany during the post-war occupation and later the stationing of so-called support troops. In both cases, the situation of stationed troops was that of friendly (or at least not enemy) soldiers with more or less frequent contact with the indigenous population. In both cases, contact between American soldiers and local women resulted in relationships of varying degrees of intimacy; these relationships ranged from entirely voluntary relations to those involving different degrees of coercion; and in both cases a significant number of these liaisons resulted in the birth of children.

The fate of children varied, and the aim of this research is to investigate differences and similarities of the children's experiences among others depending on

- Where and when they were conceived
- Whether the parents married, or whether they grew up in single-parent families, or were adopted either by third parties or one parent's family
- Whether the children were of single-race or bi-racial descent
- Whether the children stayed in their mothers or their fathers' home country
- Whether the children knew about their fathers' background or not

The proposed presentation will outline the state of research in this broader field of field and it will present the work in progress relating to this particular comparative project with focus on working hypotheses, preliminary results and specific methodological problems encountered in the research.

### 12.2.2 Children of the East African great lakes: War breeds

Eunice Apio, Repu Consultancy (Uganda)

**Abstract** In Uganda, since 1988, a new hitherto unknown phenomenon of child birth has been taking place, courtesy of the unfortunate two decade long war orchestrated by the barbaric, cultish Ugandan rebels of the Lords Resistance Army (LRA). The LRA orchestrated over 60,000 child and adult abductions on the unprotected communities forcing them to fall in rank as sex slaves, porters and soldiers, often subjecting them to untold levels of sufferings to break their spirits and consolidate allegiance to the rebels' cause. The LRA have, for the last two years, expanded their cross border bases, continuing to swell its military and sex slave population through ruthlessly harvesting the children of the Democratic Republic of Congo, Southern Sudan and The Central African Republic, where more than 1,000 children have been abducted and subjected to the same situation as Ugandan young girls, perhaps a near-familiar phenomenon befalling the Civilian Darfur population in the hands of the Janjaweed militia of Sudan, except for this latter case, the perpetrators are not kin and kith.

With extremely limited literature and an attempted documentation that started more than ten years into the war, a lot of hard facts remain unknown and unconsolidated, with no known deliberate attempts to feature such in the near future. But perhaps one of the major consequences has been the birth of thousands of children in to the war itself in at least four categories:

- Born of thousands of abducted girls within the LRA encampment returned or still holed up with the LRA
- Conceived of the LRA and born upon return by abducted girls
- Born of rape within Internally displaced persons camps
- Born of Ugandan military at war fronts

This paper will examine the status of children born of thousands abducted by the LRA against the background of forced relationships with men responsible for their abduction and the complete destruction of whole communities from where the abducted and rebels hail: never before had the various tribes involved confronted with a similar phenomenon. Of importance will be an examination of the extent to which qualitative and quantitative documentation on this category has been undertaken.

### **12.2.3 Bringing peace and babies: Children born by relationships between peacekeepers and local women. Examining the UN strategy for assistance for acceptance**

Kai Grieg, United Nations Association of Norway (Norway)

**Abstract** On December 21, 2007 the UN General Assembly adopted a resolution with the long title: United Nations Comprehensive Strategy on Assistance and Support to Victims of Sexual Exploitation and Abuse by United Nations Staff and Related Personnel. Some of these victims are victims in a very peculiar way; their victim status is a result of them being born as a result of relationships or sexual exploitation by personnel that the United Nations brought into the country. The resolution calls for a detailed report to be made on this issue for the 64th General Assembly, in 2009. The report will then be the basis for the follow up by UN on this issue. Children born by different actors of war - and peace - bring forward questions of both political and academic interest: Are children born of rape or relationships really victims? Are children of mixed origin subject to stigmatization due to their origin? What are the factors minimizing the risk of stigmatizations? Could the UN strategy be used as a blueprint for other children born of war and their rights? Is the United Nations having success in pursuing claims of paternity as is planned in the strategy? How can such a sensitive issue be addressed by researchers and victim support facilitators? These and other issues will be addressed and discussed in the presentation.



## 12.3 Climate Change Survey: substantive and methodological issues

### Coordinator:

Eugene Kritski, GlobeScan Incorporated (Canada)

### 12.3.1 Patterns of Climate Change Perceptions and Individual Behavior

Eugene Kritski, GlobeScan Incorporated (Canada)

**Abstract** GlobeScan Inc. has completed two waves of Climate Change Monitor, an annual survey of 1000 respondents surveyed in each of more than 20+ countries of the world. Using the data collected in 2008 we have attempted to build a taxonomy that contextually describes key paradigms or mindsets based on respondents attitudes to different issues related to climate change as well as their personal strategies and behaviors aimed at mitigating global warming or at adapting to the changing natural and socio-economic conditions.

The input variables that included attitudes towards climate change, personal empowerment and frequency of specific actions have helped identify 5 distinctive patterns of public perception of global warming: Mitigating Enthusiasts, Self-Contented Activists, Conventional Activists, Potential Activists, and Indifferent.

The presentation will focus on methodology of classification and on segment profile across respondents cultural and socio-demographic characteristics.

### 12.3.2 Climate Change Perception and Mitigation Behavior in Switzerland within a Structural Equation Modeling Approach

Aysel Tikir, Swiss Federal Institute of Technology Zurich (ETH) (Switzerland)

**Abstract** According to Kearney (1994 cited in Rebetez 1996) climate change cannot be perceived directly by individuals because of the temporal scales associated with it. In addition, human perception of climate is strongly influenced by expectations, which may have little relationship to the true nature of climate as provided by the instrumental record (Rebetez 1996). Moreover, responses to climate change are diverse and thus difficult to focus on certain behaviors to mitigate. Altogether climate change is a rather complex issue and needs a comprehensive analysis. Some recent studies give evidence that public perception of climate change may also be affected by different cultures. For instance, US-Americans have a moderate perception of climate change as a risk (Leiserowitz 2006) while British people are rather concerned about the same issue (Poortinga, Pidgeon, Lorenzoni 2006). In contrast, Semenza et al. (2008) found out that almost all respondents (USA) have heard about climate change or global warming, are rather concerned (women more than men) and some 50% reported behavior changes like decreased energy usage at home, reduced gasoline consumption, increased recycling and some other behaviors.

Therefore the objectives of this study are a) to assess the perceptions of Swiss people regarding climate change and influencing factors towards mitigation behavior; b) to identify and quantify the influencing factors of mitigation behavior within a behavioral model; c) to test the Theory of Planned Behavior (Ajzen 1991) within the climate change context and extend the theory with additional variables.

To achieve these goals an online survey among all university members (students and staff of ETH, about 19000 people) is conducted in November 2008. The response rate is about 18% (3514 respondents). Data is analyzed within a Structural Equation Modeling (SEM) approach to model the influencing factors of mitigation behavior (e.g. public transport, recycling and heating).

The application of SEM requires a well-defined theoretical framework, as it takes a confirmatory approach to the analysis of a given structural theory. The Theory of Planned Behavior (ToPB), developed by Ajzen (1985), is taken as the theoretical basis. It encompasses three theoretical constructs, which influence the intention to perform a given behavior, viz. the attitude towards behavior, the subjective norm, and the perceived behavioral control. The more favorable the attitude toward a given behavior and the subjective norm, and the greater the perceived behavioral control, the stronger should be the persons intention to perform the behavior in question. The survey contained socio-demographic variables as well as indicator variables since all theoretical constructs need to be operationalized through indicators. All indicators are graded on a five-point scale.

First results show that Attitudes have the biggest effect (.48) on Intentions followed by Subjective Norms (.22); these effects are significant at the .1% significance level. In contrast, Perceived Behavioral Control has no

effect. The explained variance in Intentions is 36%. The commonly used fit indices RMSEA (.07) and NFI (.95) indicate a good fit of the model.

Future work will consider risk perceptions and world views within the SEM model.

References:

- Ajzen, I. (1985, Reprinted): Attitudes, personality, and behavior. Milton Keynes : Open University
- Leiserowitz, T. (2006): Climate Change Risk Perceptions and Policy Preferences: The Role of Affect, Imagery and Values. *Climatic Change* 77: 45-72.
- Rebetez, M. (1996): Public Expectation as an Element of Human Perception of Climate Change. *Climatic Change* 32: 495-509
- Semenza, J.C., Hall, D.E., Wilson, D.J., Bontempo, B.D., Sailor, D.J. George, L.A. (2008): Public Perception of Climate Change Voluntary Mitigation and Barriers to Behavior Change. *American Journal of Preventive Medicine*, 35(5): 479-487.
- Poortinga, W., Pidgeon, N.F., Lorenzoni, I. (2006): Public Perceptions of Nuclear Power, Climate Change and Energy Options in Britain. Summary Findings of a Survey Conducted during October and November 2005. Technical Report (Understanding Risk Working Paper 06-02), Norwich: Centre for Environmental Risk.

### 12.3.3 Democratic Smog? An Empirical Study on the Correlation between Social Class and Environmental Pollution

Reto Meyer, ETH (Switzerland)

Andreas Diekmann, ETH (Switzerland)

**Abstract** For years the public and scientific debate about Environmental Justice was mostly confined to the U.S. Only recently, the question about the existence and strength of the social gradient of environmental pollution has entered the European debate. Early research finds inconclusive results in the German speaking world, primarily because this preliminary research simply records subjective perceptions of pollution and correlates them with indicators of social status. Objective measures of environmental quality are seldom used, and even then only in studies of small geographic areas. In contrast, the present study uses various objective measures of air pollution (NO<sub>2</sub>, particulate matter, ozone) and road traffic noise (day, night) and assigns them to the respondents of the Swiss Environmental Survey 2007 (N=3369) using a geographic information system (GIS). The combination of objective GIS coded data with subjective measures allows for a new approach in Environmental Sociology that takes spatial context into account, which is often neglected in sociological studies. Using bivariate and multivariate statistics this objective data on pollution is related to indicators of social stratification such as income, education and nationality as well as the subjective perception of pollution. Surprisingly, and contrary to the expectations, the indicators of social status (income and education) are not significantly correlated with the environmental pollutants considered. In the multivariate analysis, however, a significant correlation with income is observed, although this effect is fairly weak. Differences between urban areas and the countryside are more pronounced. Concerning the subjective perception deviations from the objective exposure can be observed for ozone.

## 12.4 Cognition in survey research (I)

### Coordinators:

Bregje Holleman, University of Utrecht (Netherlands)

Naomi Kamoen, University of Utrecht (Netherlands)

### 12.4.1 Cognition in surveys

Bregje Holleman, University of Utrecht (Netherlands)

Naomi Kamoen, University of Utrecht (Netherlands)

**Abstract** In recent years, various models that describe the cognitive processes underlying question answering have been proposed. A lot of research is guided by the model of question answering by Tourangeau, Rips and Rasinski (2000). This model distinguishes four stages in question answering: (1) comprehension of the question, (2) retrieval of information, (3) deriving a judgement, and (4) formulating a response.

In addition, there are dual-process models, such as the satisficing model proposed by Krosnick (1991). In this model, two groups of respondents are distinguished: those who satisfice, and try to do just enough to give a plausible answer versus those who optimize, and do their best to give a good answer.

Cognitive models such as the two described above, have many applications. For example, they help in understanding what is measured when administering surveys, and they provide a point of departure in explaining the wide range of method effects survey researchers observe. Also, cognitive theory in surveys is used by psychologists, linguists and other scholars to obtain a deeper understanding of, for example, language processing, the nature of attitudes, and memory.

In this talk we will discuss some cognitive models, focussing especially on the models by Tourangeau, Rips and Rasinski (2000) and Krosnick (1999). Furthermore, we will talk about some recent research on cognition in surveys. Part of this research uses increasingly advanced cognitive measures, such as eye-tracking or ERPs (brainwaves). Different fundamental issues are touched upon, such as: to what extent is research on cognition in surveys useful for survey methodology? And to what extent do the explanations for specific method effects contribute to the development of cognitive theory?

### 12.4.2 Empirical test of a generic cognitive dual-process model to explain attitude-behavior relations and occurrence of response effects

Jochen Mayerl, University of Stuttgart (Germany)

**Abstract** According to common dual-mode models of social cognition, two modes of information processing are distinguished: answering to survey questions in an automatic-spontaneous mode in contrast to answering in a deliberative-controlled manner. Based on a generic cognitive dual process model of attitude-behaviour research, it is argued according to the MODE model that attitude accessibility acts as a moderator of attitude-behavior relations in the spontaneous mode only.

Several hypotheses are derived regarding a) the predictive power of attitudes towards behavioral intentions and actual behavior and b) the occurrence of response effects while answering to attitude questions in surveys. The statistical test is conducted on money donations to charities using data of a 2005 nationwide two-wave German CATI-survey with 2002 respondents. The included measurement of response latencies allows the operationalization of the mode of information processing while answering to survey questions.

As a result of several structural equation models, the theorems of the generic dual process model are confirmed empirically. Therefore, attitudes are most predictive towards information processing and behavioural intentions when processing takes place in the spontaneous mode with high chronic attitude accessibility. Additionally, direct effects of attitudes on reported behaviour appear in the spontaneous mode with high accessible attitudes even if controlling the intention as a mediator variable. In contrast, no moderator effects of the chronic accessibility are found in the deliberative mode of information processing. As expected, simple judgement heuristics e.g. acquiescence only have a meaningful influence on answering to attitude questions in surveys when information is processed spontaneously with low attitude accessibility. Furthermore, deliberative judgments (i.e. judgments with longer response latencies) are based more on beliefs than spontaneous judgements.

### 12.4.3 Identifying Nonattitudes: Comparing Response Latencies and Behavior Coding

Piet Sellke, University of Stuttgart (Germany)

**Abstract** Why do respondents answer questions in surveys they do not know anything about? Fazio interpreted this behaviour in the context of attitude strength (Fazio 1986, 1989). Thereby, Fazio defined attitude strength as the strength of the cognitive association between an attitude object and its evaluation. Further, if there is no association between attitude object and its evaluation, the answer given is defined as nonattitude.

The research question of this paper is how nonattitudes can be identified. To answer this research question, a methodological comparison of the identification with response latencies on the one side and with behaviour coding on the other side is undertaken. A German-wide representative CATI-survey, conducted in 2005, serves as data basis. Within this survey, attitudes towards known and unknown (i.e. fictive) attitude objects were measured, assuming that substantial answers to unknown attitude objects represent nonattitudes. Response latencies have been measured for all items of the survey and about 50% of all respondents were additionally tape recorded during the interview. Three basic steps were undertaken to answer the research question: First, substantial answers to attitude questions with fictive attitude objects have been analysed according to their response latencies. Second, behaviour coding was undertaken for these fictive questions. Aim of those first two steps was to identify response latency patterns and behaviour coding patterns of respondents while answering nonattitudes. As a third step, patterns found from response latencies and behavior coding were applied to attitude questions with non-fictive attitude objects. Items found to be a possible nonattitude in reference to response latencies and/or behaviour coding were cross-checked regarding their attitude strength, measured with meta-criteria. Further, attitude stability of possible nonattitudes is checked with multi-group analysis within structural equation modelling.

Results indicate that Nonattitudes can in fact occur with long and short response latencies. Further, respondents with long response latencies can be differentiated through behaviour coding into those with strong attitudes and those holding Nonattitudes. Thus, in the case of long response latencies behaviour coding can add additional information one would not have gained through the analysis of response latencies alone. However, in case of short response latencies behaviour coding is not applicable and response latencies identify more reliable Nonattitudes and strong attitudes.

### 12.4.4 Cognitive processes in answering contrastive survey questions: the use of eye-tracking

Naomi Kamoen, University of Utrecht (Netherlands)

Bregje Holleman, University of Utrecht (Netherlands)

Huib van den Bergh, University of Amsterdam (Netherlands)

**Abstract** A well-known advice on survey design is that balance between positive and negative wordings should be obtained. The assumption behind this advice is that respondents should give the same answers to positive and negative questions. However, a large body of research shows that this assumption is incorrect; people express their opinions more positively when the question is worded negatively.

By relating wording effects to the cognitive process underlying question answering, it can be understood why the asymmetry between positive and negative questions arises. Tourangeau, Rips & Rasinski (2000) describe four stages in answering attitudinal questions. With respect to the validity of survey questions, it is fruitful to cut down their model into two stages: (1) interpretation of the question and retrieving an evaluation (2) mapping the evaluation onto the response options. If the question wording affects the interpretation stage, positive and negative questions measure different attitudes and differ in their validity. If question wording affects the mapping stage, contrastive questions are equally valid, but the same attitudes are expressed differently onto the response options.

In previous research, Holleman (1999) observed a high correlation between answers to positive and negative questions. This implies that question wording does not affect the interpretation stage. Because the answers to positive and negative questions do differ, the question wording must have become relevant in the mapping stage. More direct evidence comes from Chessa & Holleman (2007). In their experiment, question and answering options were offered separately to respondents, and the times for reading the question (interpretation) and giving an answer (mapping) were recorded. The results show that there is no difference in interpretation times for positive and negative questions, but that the mapping stage takes longer for negative questions.

Based on the evidence available we would conclude that wording effect arise in the mapping stage. Yet, the correlational study gives only indirect evidence for this conclusion, and a drawback of the reaction time study

is that the natural process of question answering may have been distorted by separating question and answering options. The aim this research is to explore whether eye-tracking can be used as a non-distorting method to measure cognitive processes.

In an experiment, respondents ( $N = 28$ [1]) answered survey questions about smoking policies, while their eye movements were being measured. Two versions of the survey were compared: questions that were phrased positively in version 1 were worded negatively in version 2. The answering data show that negative questions are answered more positively. Preliminary analyses of the eye-tracking data show that respondents spend more time looking at the question (and not at the answering options) when it is worded negatively. This seems to contradict previous results, because time spent at the question can be interpreted as an interpretation process. However, respondents also switch more often between question and answering options when the question is worded negatively, so the larger time spent at the question might also reflect a mapping process. Various measures that can be distilled from the rich eye-tracking data will be presented and related to the cognitive process of question answering.

## 12.5 Cognition in survey research (II)

### Coordinators:

Bregje Holleman, University of Utrecht (Netherlands)

Naomi Kamoen, University of Utrecht (Netherlands)

### 12.5.1 Bipolar versus Unipolar Scale Format in Fully versus Endpoint Verbalized Scales

Dagmar Krebs, University of Giessen (Germany)

Jürgen H.P. Hoffmeyer-Zlotnik, GESIS - Leibniz Institute for the Social Sciences (Germany)

**Abstract** The application of cognitive theory to survey methodology uncovered that answering survey questions is a cognitive process consisting basically of four tasks: question interpretation, memory retrieval, judgement formation, and response editing. This paper deals with the third of these tasks. In the process of formatting a response, respondents have to map the selected answer into response categories offered by the researcher. In this process respondents use every bit of information they can infer from the questionnaire. According to the conversational logic, characteristics of response categories are a main source of information because respondents derive subjective theories from these characteristics which function as orientation guides in locating themselves on the response scale into one of the offered response categories.

Acknowledging the impact response categories can have on the cognitive task of formatting an answer, this paper deals with two specific aspects of response format: First, fully versus partly verbalized response scales, and second, unipolar versus bipolar format of response scales.

Responses on an endpoint verbalized response scale are expected to be more extreme in either positive or negative direction than responses on a fully verbalized scale. Verbal labels at the extreme points might have anchoring effects thus offering raw orientation for respondents in formatting their answers. Fully verbalized scales, in contrast, direct respondents attention to each single scale point, thereby supporting the choice of less extreme response categories.

The second hypothesis is that response behavior is affected by the polarity of the response scale. Unipolar scales run from, for example, very important to not at all important while bipolar scales runs from very important to very unimportant. It is expected that responses on the bipolar response format are more positive than on the unipolar format because in formatting their answers, respondents not only shy away from the extreme negative pole but they also avoid the negative area of the scale altogether. While the first hypothesis tests for effects across types of verbalization within polarity type, the second hypothesis tests for effects across polarity of the response scale within verbalization type.

Based on a panel design combined with split-ballot forms of the questionnaire the postulated effects of response scale format on response behavior are tested. Respondents were students. Questionnaires were distributed three times with a weeks distance between the waves. Randomization of split forms of the questionnaire was achieved by distribution technique. Split groups of respondents did only participate once, panel members participated either twice or three times. Question content was anomie, authoritarianism, attitudes toward immigrants, achievement motivation, and extrinsic as well as intrinsic job motivation. For all response formats a 7 point scale was used.

By comparing responses across split groups (a subgroup of respondents participating and responding to the questionnaire only once) differences in test statistics (proportions, means, standard deviations and reliability coefficients) are described. Additionally, based on the repeated measures with different response formats, reliability and validity of indicators as well as constructs are tested by analysing an MTMM model. By using the MTMM model it is also possible to describe the different impact (method effect) of each response format on measurement quality.

### 12.5.2 Presuppositions and Filter Questions in Surveys

Ting Yan, University of Chicago (United States)

**Abstract** Survey literature has documented that filter questions reduced the percentage of people who gave opinions and increased Dont know responses when compared to questions directly asking people to give opinion on issues (the so-called corresponding direct questions). Various rationales have been proposed to account for

this phenomenon. One recent explanation is that direct questions carry a presupposition that can influence respondents subsequent responses. By contrast, filter questions cancel the presupposition behind the corresponding direct questions and allow for people the choice to claim that they don't know about the topic of interest or the topic doesn't apply to them. This paper extends the previous work on presupposition in survey by examining the effects of presupposition in web surveys, a self-administered mode. Two experiments were embedded in web surveys. The first experiment tested the effects of filters on questions asking people about the perceived importance of various issues. For a random half of the sample, respondents were asked directly how important they consider an issue is (the direct question condition) and the other random half received a filter first asking whether they consider the issue important or unimportant (the filter question condition), followed by a second question asking them how important (unimportant) they think the issue is. The second experiment studied the effect of filter questions on behavioral/factual questions. Again a random half of the respondents were asked directly how many times they've witnessed a crime (or have been victimized) in the direct questions condition and the other half were given a filter asking whether they have witnessed a crime (the filter question condition). Both experiments showed that filter questions change the response distributions by cancelling out the presupposition carried by the corresponding direct questions. Furthermore, some respondents are more prone to the filter effect than others. I will examine how respondent characteristics interact with the presence/absence of filters in survey responses.

### 12.5.3 What causes Don't Know Responses of Children in Interviews? Models of Cognitive Processes in Answering Questions as Explanation

Susanne Vogl, Katholische Universität Eichstätt-Ingolstadt (Germany)

**Abstract** Although children are increasingly seen as competent informants we know very little about their cognitive and interactive skills in interview settings. To tackle problems of constructing adequate research instruments for children (of different age groups) we need to take those skills into account. A specific problem in any research is item-nonresponse. Don't know (DK) as well as substantive response based on weak information or mere guessing undermine the data quality in social research. By applying cognitive information-processing models I try to explain children's DK answers in interviews. Only with a deeper understanding of children's cognitive information processing we can construct research instruments that are adequate for children.

The analysis is based on results from a semi-structured approach to interview children aged 5 to 11. Each of the 56 children was interviewed twice, once on the telephone and once face-to-face. Altogether 112 interviews resulted. By this qualitative procedure the context of the given DK answers can be analysed and inferences about reasons for DK answers within the question-answer process can be made. Research Questions are: What causes DK responses in interviews with children? Where in the cognitive question-answer process can reasons be allocated? Does age as a correlate to cognitive abilities have an effect on either the total number of DK responses or the cognitive reason? What do response latencies tell us about the cognitive state of the interviewee and the validity of his (DK) response? Are there specific interactional patterns that accompany DK responses?

The qualitative approach can be used as a vehicle also for a better understanding of question-answer processes in more standardised interviews.

## 12.6 Data quality in surveys among the elderly

### Coordinator:

Marek Fuchs, University of Kassel (Germany)

### 12.6.1 Elderly in an Internet panel, the quality of the data

Corrie Vis, CentERdata (Netherlands)

**Abstract** According to literature some groups are often underrepresented in survey research. For instance people with a lower economic status, people in urbanized areas and lonely people. Internet panel research forms an extra threshold. People without Internet access and people who don't want to join for a longer period are less inclined to participate. In the Netherlands in 2007 a probability based Internet panel of 5000 households was recruited and all recommendations from literature to get an optimal response have been implemented; Longitudinal Internet Study for the Social Sciences (the LISS panel). We will discuss how older people, mostly without Internet access, were persuaded to participate in the panel. Furthermore we will consider the quality of the data of the older panel members (wave nonresponse, item nonresponse) over a year and compare the quality of the answers on some questions from the face-to-face survey SHARE (Survey of Health, Aging and Retirement), which were also asked in the LISS panel.

### 12.6.2 Item non-response in the first wave of SHARE. The impact of question characteristics and of the respondents' cognitive resources

Marek Fuchs, University of Kassel (Germany)

**Abstract** The elderly population is growing consistently across most European countries. Usually, standard procedures derived from survey methodology are adapted for surveys in the elderly population. However, several studies have indicated that older respondents are prone to survey error to a larger degree. For example, item non-response, non-differentiation, extremity bias, and other indicators are more pronounced among the elderly compared to younger adults.

In the literature various causes for the reduced data quality of responses obtained from the elderly are discussed. Among others, the respondents cognitive resources are held responsible for item non-response. Since older respondents are assumed to have a reduced working memory capacity as well as other cognitive limitations they are expected to have more pronounced problems with answering survey questions. In light of these studies, the problems are especially prevalent in behavioral questions and factual questions. By contrast, other scholars assume that older respondents have higher expectations regarding the quality of their behavioral and factual reports which in turn reduces their willingness to provide unreliable responses. Thus, they assume that higher levels of item non-response among older respondents do not necessarily indicate poor data quality. Instead, item non-response should be treated as sign of optimizing response behavior.

In this Paper we will focus on the item non-response in the data obtained from elderly respondents in the first wave of the SHARE study. We will test whether the occurrence of item non-response can be correlated to the available measure of the respondents cognitive resources. We will present findings from a multilevel logistic regression (STATA gllamm) assessing the driving factors for the occurrence of item non-response. In addition to question characteristics and socio-demographic characteristics of the respondent we will assess the impact of the respondents cognitive resources.

Results confirm the hypothesis according to which question properties as well as the socio-demographic characteristics of the respondents have a significant impact on the occurrence of item non-response. In addition, a reduced working memory capacities leads to more item non-response. Similar effects are observed for measures of reasoning and fluency. Findings are discussed in light of the contradictory explanations of item non-response among the elderly.

### 12.6.3 Does Age Affect Quality? Evidence from SHARE

Mathis Schröder, University of Mannheim (Germany)

Sabrina Zuber, University of Mannheim (Germany)



**Abstract** The demographic development is one of the greatest challenges for Europe in the 21st century. As older people constitute larger and larger proportions in the population, not only does their political clout grow, but also the interest for this group from various other sources, like the media, advertising or medicine. Social research can play a key role in understanding the challenges of demographic change in its various facets, including health, economics and social developments. However, it is not quite clear how the growing proportion of the elderly influences particularly the collection and quality of survey data and thus empirical research. While increasing age is generally associated with declining cognitive ability, it is not quite clear in the literature whether survey quality is indeed affected by this decline or to what extent (see for example Andrews and Herzog, 1986, and Calahan, 1969).

The goal of this paper is to investigate how quality of survey data is influenced by age and whether any differences found are large enough to warrant differences in data collection depending on age. We use the Survey of Health, Aging and Retirement in Europe, SHARE. SHARE concentrates on respondents 50 years old and older in 15 European countries, and is constructed similar to other aging surveys, like the Health and Retirement Study (HRS) in the US or the English Longitudinal Study on Aging (ELSA) in the UK.

The task of measuring quality is approached mainly in considering fractions of item non-response, i.e. the amount of questions a respondent answers with Dont know and / or Refusal in relation to the overall number of questions (see for example Guadagnoli and Cleary, 1992). This will be split by different topics in the questionnaire. In a preliminary analysis, we found that between the ages of 70 and 90, the item non-response almost doubles in a linear fashion from 6.3% to 11.9% in SHAREs employment and pension module. Other measures of quality we consider include response scale effects and variance analyses. The analyses will be elaborated upon by controlling possibly confounding variables, like gender, education, health and other demographic variables but also paradata, like the duration of the interview when the question was asked to cover effects of fatigue. Finally, SHARE allows assessing country specific differences in these effects, as the ex-ante harmonized survey ensures that all interviews are conducted in the same manner.

## 12.7 Demonstrations of new resources

### 12.7.1 Presentation of ESS EduNet, an Internet-based analysis-training resource

Atle Jaastad, Norwegian Social Science Data Services (NSD) (Norway)

**Abstract** The students and teachers of today are living in an era of possibilities. Never before have the amount of data been richer, never before have the barriers for access to empirical data been lower and never before have the tools for processing data been more powerful.

But it is still difficult to be a student or a teacher. The student will always face the problem of how to utilise a dataset in a way that makes sense, and the teacher is obliged to lead the students on the path from ignorance to competence. These tasks can't be solved by any external development: The students have to internalise facts and knowledge, and the teachers need to reflect upon how and where they would like to lead their students.

However, the recent technological development has made life easier for students and teachers. Several services on the Internet have been established to facilitate analysis of real world data. Some services offer just data, others have added texts and exercises and some have also included tools for data analysis. Compared to the old days, it ought to be possible to save a lot of time and effort in the process of data handling and analysis.

In this presentation, I would like to focus upon one of the services available on the Internet: European Social Survey Education Net, or ESS EduNet, <http://essedunet.nsd.uib.no>.

ESS EduNet is a training resource mainly developed for use in higher education. The resource aims to be a social science laboratory where theoretical questions can be explored using high quality empirical data. The pedagogical idea is to create an interactive learning environment that combines the elements of social science to an integrated whole: theory, data and methodology.

Through this presentation, my objective is to give you some ideas regarding what kind of web site ESS EduNet is, about its contents and functionality, how it is designed to make it appealing to the users, and finally about how it is to work with this site.

### 12.7.2 Survey Research in Virtual Worlds: Second Life® as a Research Platform

Craig A. Hill, RTI International (United States)

Elizabeth Dean, RTI International (United States)

**Abstract** Second Life, a 3D virtual world with over 15 million online residents, was created in 2003. At any one time, an average of over 50,000 residents are online, employing an avatar (a digital representations of self), that allow users to explore the virtual world, socialize with each other, learn, and transact business. Second Life is a global phenomenon (approximately 60 percent of residents live outside the United States).

RTI International operates a prototype virtual research facility in Second Life. RTI's virtual research facility was built to investigate and evaluate the potential for collecting high quality survey data in virtual worlds; not coincidentally, our facility is used to communicate with, and train, a globally-dispersed field staff. The virtual building features open meeting space and four private interview rooms. Two of the private rooms are large enough to seat 415 avatars for a focus group. The other two private rooms are designed for one-on-one interviews. All rooms allow additional space for observers.

Our first experiments with these virtual interviewing capabilities and modes, including both survey kiosks and avatar-to-avatar interviews, will be discussed. At present, we are interested in discovering which modes work best in-world, whether or not, and what kind of, incentives work to gain the cooperation of inworld avatars, and whether or not we can get avatars to answer survey questions as their real-life human counterpart, as divorced from their inworld avatar. Our first interviews are being conducted during January - March 2009, and will be able to present preliminary findings in Warsaw. Finally, we conclude with a discussion of next steps and plans for further research in-world.

## 12.8 Special Issues

### Coordinator:

George Petrakos, Agilis SA (Greece)

### 12.8.1 Quality assurance and quality assessment in the European Working Conditions Survey

Agnes Parent Thirion, European Foundation for the Improvement of Living and Working Conditions (Ireland)

Michalis Petrakos, Agilis SA (Greece)

Greet Vermeylen, European Foundation for the Improvement of Living and Working Conditions (Ireland)

Branislav Mikulic, European Foundation for the Improvement of Living and Working Conditions (Ireland)

**Abstract** The quality of statistical data is a primary concern and substantial effort has been devoted by organisations involved to ensure high quality data is collected and disseminated. European surveys, implemented simultaneously on many countries pose significant challenges for statisticians and survey managers. Ensuring comparability in different legal, cultural and societal backgrounds is compounded by the fact that data collection is performed by independent organizations in each country while available budgets are constrained. A quality assurance framework is required to ensure that data produced meet the needs of users for high quality, comparable data.

The presented plan has been developed based on the needs of the European Working Conditions Survey (EWCS). It aims at ensuring that survey data collected and provided to the EU social policy debate are of the highest quality within the budgetary constraints of the Foundation. Data quality is based on Eurostat framework for data quality widely used in official statistics. The main aspects of the quality assurance plan are outlined in this presentation. It includes a thorough mapping of survey operations by the main survey actors. The plan identifies the key process variables that affect the quality of the statistical process and establishes a system for monitoring and control.

At the same time the dissemination of detailed quality information along with the estimates and the statistical data is also important as it allows users to make the best use of the data provided for their purposes. Survey quality characteristics are presented based on the six quality components defined by Eurostat along with the methodology used for their computation and the related information needs.

### 12.8.2 Using SDMX standard for assessing Quality in Statistical Frames

Spyros Liapis, Agilis SA (Greece)

George Petrakos, Agilis SA (Greece)

**Abstract** Quality in Statistical Surveys is a complex (multidimensional) concept, which incorporates attributes and characteristics such as timeliness, relevance, accuracy, coherence, comparability, clarity and accessibility addressed to statistical data which is the main product of the various statistical systems and surveys. The quality of these products is defined and measured in accordance with the customer (data user) needs. In a similar way, we consider a statistical frame (or a register) as a product and we define quality with its major attributes and characteristics, focusing on the user, which is the statistical survey itself, needs. In this paper we attempt to define quality components addressed to any statistical frame, using the same attributes and characteristics redefined in context of statistical frames and registers. These components can be expressed in a descriptive and measurable manner through a metadata repository, organized in a SDMX (Statistical Data and Metadata eXchange) standard aiming at the harmonization and evaluation of quality in large surveys repeated in time and space (Official Statistics Surveys, multinational social surveys etc.). SDMX is a standard that facilitates the statistical data and metadata exchange. It offers artefacts for representing structures for reference metadata that can be used to produce metadata sets in XML (i.e. SDMX-ML the XML implementation of the SDMX standard). The artefact for creating such structures is called Metadata Structure Definitions (MSD) and can define all metadata and their definitions. This artifact has been used for representing the metadata for the statistical frames presented in this work. Therefore these metadata values can be represented with SDMX-ML metadata sets and processed by any SDMX tool (e.g. related with storage, transmission, dissemination). This opens a broad area of application providing registers and frames with structural metadata repositories.

Keywords: Metadata; Statistical frames; Quality; SDMX standard; XML

### 12.8.3 Socio-economic inequalities and electoral conduct for local Romanian elections in 2008

Corneliu Iatu, University of Iasi (Romania)

Elena Daniela Viorica, University of Iasi (Romania)

Adrian Lucian Lupu, University of Iasi (Romania)

**Abstract** The article will examine the effects of socio-economic variables on the parties results in local Romanian elections in 2008. Econometric model of voting was simply filled with socio-economic variables. Moreover, studies (Kramer, 1971) on the link between electoral success and economic conditions (especially macroeconomic) or between their policies and election results up to the penalty / reward voting mode (Key, 1966; Hibbs, 2006) were based on the causal links between the electoral conduct and socio-economic status around elections. We follow that through multivariate analysis to determine whether mayors who belonged to a particular party or that party (if it has changed election candidate) have managed to win elections, or instead were penalized by the electorate. Among political variables may include: the share of votes obtained by the mayors (differentiate here those who were re-elected, those belonging to the same party and those who were not re-elected), the number of seats that re-elected and not elected mayors had, the parties that the mayors who have won the election are part of, etc. Among the socio-economic variables are mentioned: the unemployment rate, the rate of migration, the natural increase rate etc.

The results show that in Romania the election is partisan, somewhat atypical against consolidate democracies and the socio-economic component plays an important role in some cases but not determining in others.

### 12.8.4 Real Data?, Real Software? Experiences with redesigned Teaching of Statistics and Survey Methodology

Torsten Harms, FU Berlin (Germany)

**Abstract** Many lecturers use surveys in order to teach elementary statistics or survey methodology - due to various reasons, there are often typical limitations associated with these efforts: No use of "real" data but rather a multitude of pre-cleaned samples to fit certain topics, uncoordinated use of tools/software, introduction to the actual sampling process late in the curriculum. In our experience, this has often led to students being unable to properly conduct surveys or use real data from official surveys when faced with this task.

The aim of this talk is to give an overview of the redesigned statistical education at the Freie Universitat Berlin with the particular aim to overcome the above limitations: The program consists of 2 main blocks: Elementary statistics with extensive use of a software called "statistical lab" (see also: ([www.statistical-lab.com](http://www.statistical-lab.com))) that was specifically designed to support ease of use but also professionalism through being based on the language R. The 1-year elementary-curriculum also heavily builds on survey data collected during the course thus exposing students to the challenges of survey statistics early on. At the advanced level the students naturally transition from the "statistical lab" software to R and SAS and from simpler to more complex surveys including the use of the full but anonymized German Microcensus provided via a Campus agreement.

A key part of the whole program is that almost all exams (including those for elementary statistics with over 300 students per class) are based on the use of statistical software and electronic submission of the results during timed tests. In addition the whole course could be followed over the internet. Both has been a particular administrative challenge but also a key driver for the success of this program and we will give details on the tools and processes used to make this possible.

Given that the first cohort of students has fully completed this program we would also like to reflect on typical pitfalls and challenges encountered as well as our key insights that might be of use for other statistical programs.

### 12.8.5 Statistical Survey of the "Challenges and Projections of the Basel II International Capital Framework in the Field of..."

Veselka Pavlova, University of National and World Economy (Bulgaria)

**Abstract** Complete title: Statistical Survey of the "Challenges and Projections of the Basel II International Capital Framework in the Field of the Accounting of Credit Institutions".

The purpose of Basel II, which was initially published in June 2004, is to create an international standard to be used by banking regulators when creating regulations about how much capital banks need to put aside in order

safeguard against various financial and operational risks they face. The three essential requirements of Basel II are: 1. Mandating that capital allocations by institutional managers are more risk sensitive; 2. Separating credit risks from operational risks and quantifying both; 3. Reducing the scope or possibility of regulatory arbitrage by trying to align the real or economic risk precisely with regulatory assessment.

A specialized study was carried out among the Bulgarian credit institutions in September 2008 in regard to the opinions expressed by experts for the criteria, approaches, advantages and deficiencies as well as the successful management of credit risk, credit protection etc. It was based on a small sample of 54 institutions and 152 experts. They were divided in three basic segments - managers, inspectors and accountants. The questionnaire includes 16 topics. Statistical programs like SPSS and STATA, that offer very good tools for data analysis, were used.

On the basis of the surveys results and their analysis the following main conclusions can be drawn:

- 1 About  $\frac{2}{3}$  of the people interviewed are of the opinion that the necessity of using the credit rating is determined by the information advantages and the favourable correlation between costs and benefits.
- 2 The minimal applied criteria for recognition of the credit evaluation are as follows; objectivity and independence, ongoing review and relevance.
- 3 The interviewed are equally. 50% are using the credit agency Fitch Ratings and the rest 50% - Moodys Investors Service.
- 4 In Bulgaria is in force regulation No 8 of the Bulgarian National Bank for the capital adequacy. According to 58, 8% of the interviewed this Regulation has imposed new principles and concepts for announcement of the information in the statements with the application of new evaluation bases.
- 5 Over 82% of the credit institutions are applying the standardized approach for reporting the credit risk. About a half of them consider that it is rather applicable for the purpose of the regulatory reporting while about one third for both regulatory reporting and the internal risk management. As a main deficiency of the method 56, 5% of the interviewed have indicated the fact, that it does not completely cover the potential losses in future critical events.
- 6 There are substantial differences between the type of credit institution and the preferred by it credit defence. The credit institutions of the first group are using mainly guarantees issued by a regional or local government authorities (24,3%) and guarantees of the central government, international banks and organizations. The credit institutions of the second group preferably rely (27,7%) on financial provisions, insurance policies, and real estate.
- 7 As a whole the statistical survey reports the fast introduction of the Basel II capital framework among credit institutions operating in Bulgaria.

## 12.8.6 Unit-Nonresponse in the IAB Establishment Panel

Florian Janik, Institute for Employment Research (Germany)

Susanne Kohaut, Institute for Employment Research (Germany)

**Abstract** The paper proposed focuses on the unit-nonresponse in the IAB Establishment Panel, a comprehensive data set describing the employer side of the labour market in Germany. Every year since 1993 (1996) the IAB Establishment Panel has surveyed the same establishments from all branches and different size categories in western (eastern) Germany. Data are collected in personal interviews with the owners or senior managers of the establishment. Meanwhile more than 15.500 firms participate in the survey. The questionnaire covers a variety of questions on important aspects of the employment. Although great efforts are taken to convince the owner/manager to take part in the survey there are always firms that do not answer the questionnaire. If this unit-nonresponse does not occur at random serious selection problems can arise. Therefore it is important to know about the selection process in order to develop effective methods to reduce nonresponse or to construct useful postsurvey compensation schemes. In this paper we try to find out why some establishments are not willing or able to respond to the questionnaire. We are especially interested in the influence of the interviewer on the survey participation. We find support for the thesis that the mutual trust between interviewer and the firms representative is crucial for the decision to participate again in the survey. We also hit upon some evidence that interviewers who carry out a lot of interviews are more successful in avoiding nonresponse.

## 12.9 Surveying sensitive subjects

### Coordinator:

Paula Devine, Queen's University Belfast (United Kingdom)

### 12.9.1 Surveying sensitive subjects: religious and national identity in Northern Ireland

Paula Devine, Queen's University Belfast (United Kingdom)

**Abstract** This paper will explore issues involved when including sensitive topics in questionnaires, using the example of religious and national identity in Northern Ireland. Although many violent political conflicts actually relate to tensions over territory, power or resources, the identity of different groups also plays a crucial role. Whilst any individual may hold several identities at any one time (for example, father and Catholic; woman and Asian), single identities are often talked about in isolation. The aim of this research was to explore the nature of national and religious identities in a more textured way than previous surveys had done, and so included a set of questions in the 2007 Northern Ireland Life and Times (NILT) survey. The research team set out to uncover the extent of diversity evident within the main national and religious groups. In addition to the strength of attachment or loyalty people showed to national and religious groups labels, the manner in which the identities combined and their impact on attitudes and emotional responses was also explored.

The paper will consider the methodological issues relating to the inclusion of sensitive topics, and how the 2007 NILT survey incorporated innovative methods to tap into emotions relating to different forms of identity. Using these survey data, this paper will examine the extent to which national and religious identities that have underpinned difference and division in Northern Ireland still remain, nearly a decade after the signing of the Good Friday Agreement. The paper will conclude with an evaluation of how well the survey was able to incorporate the complexities and sensitivities of identity, and evaluate if the research team achieved their aim.

### 12.9.2 Surveying non-EU immigrants in the Czech context

Yana Leontiyeva, Czech Academy of Sciences (Czech Republic)

**Abstract** Immigration is a relatively new issue for the CR, which was predominantly an emigration country until 1989. During the latest decade the country is experiencing an intensive inflow of population of foreign origin. Immigration is still a relatively new topic for Czech sociology. However a series of academic research projects appeared within the last decade. They are focused on different target groups, studying different aspects of migrants' life and naturally they are using quite different methodological approaches. What is worse, sometimes there is an obvious lack of serious methodological background. Majority of quantitative researches of migrants in the CR applies questionnaire survey method. Obviously the major pitfall for applying this method in the CR (as well as in many other countries) is the lack of the credible sample base as the main obstruction for gathering representative random sample data. Researchers applying non-probability quota sample in the Czech context are facing the dilemma of insufficient official statistics necessary for setting up the quota characteristics.

The aim of my presentation is to discuss two methodologically comparable quantitative surveys of non-EU foreigners, who stay in Czech Republic on the basis of long-term residence permit for the purpose of employment. The surveys to be discussed are quite unique in terms of representing certain part of immigration community. The first one was held by the Research Institute for Labor and Social Affairs in 2001. The second survey was conducted in 2006 by the Center of Public Opinion Research, which is a department of the Institute of Sociology, Academy of Sciences of the Czech Republic. Both addressed surveys applied the quota sample with quota characteristics derived from the official statistics on foreign migrants. The sample was based on the database of the registered foreign holders of work permits from the Ministry of Labor and Social Affairs and Czech Foreign Police.

In my short presentation I would like to discuss main methodological difficulties and successes of named surveys and to give some ideas for the future development of empirical research on immigrants in the Czech context. I believe that my presentation on Czech surveys dedicated to work and life conditions of labor migrants could be useful and close to the topic of the Special Issues Related to Survey Research session focused on Surveying Sensitive Subjects.

### 12.9.3 Is there anything you can tell me about people who are Catholics? Ethical and questionnaire design issues relating to the evaluation of Sesame Tree...

Emma Larkin, Queen's University Belfast (United Kingdom)

Paul Connolly, Queen's University Belfast (United Kingdom)

**Abstract** Is there anything you can tell me about people who are Catholics? Ethical and questionnaire design issues relating to the evaluation of Sesame Tree on young childrens social attitudes and awareness in Northern Ireland.

Sesame Tree the Northern Ireland version of Sesame Street has been produced in Northern Ireland to support the Personal Development and Mutual Understanding area of the Revised Northern Ireland Curriculum. It aims to promote a range of positive outcomes among young children including a better understanding of themselves and others, more positive relationships with others, a greater appreciation of and respect for diversity and an awareness of how to live as part of a community.

The effectiveness of the television series is being evaluated through a series of studies that will be described in this paper: The first of these was a cluster randomised controlled trial involving 20 primary schools and 440 children aged 5-6 years. The intervention consisted simply of the children in the P2 classes of these schools being shown three episodes of Sesame Tree a week for a period of 10 weeks. The second was a naturalistic study tracking childrens viewing habits (specifically in terms of their naturalistic exposure to Sesame Tree on television) and attitudes over an eight month period. 34 primary schools were selected randomly from across Northern Ireland and 697 young children (aged 5-6) took part in this second study. Their attitudes and awareness were tested at the beginning of the study period and then again at the end and three different measures of their exposure to Sesame Tree were captured at four time points.

The effectiveness of Sesame Tree was assessed in relation to four main outcomes: 1) whether it increased childrens willingness to be inclusive of others in general; 2) whether it increased childrens willingness to be inclusive of someone from a different racial background; 3) whether it increased childrens positive attitudes towards cultural activities associated with their own and other communities; 4) whether it reduced Catholic and Protestant childrens tendency to see themselves as different to one another; and 5) whether it increased childrens environmental awareness, particularly in relation to the need to recycle household waste. Multilevel modeling was used to analyse the data from both from cluster randomised controlled trial and the naturalistic study.

A third cluster randomised controlled trial involving 28 primary schools and children aged 5-6 years is ongoing. This aims to evaluate an Outreach Intervention whereby children are shown 6 episodes of Sesame Tree and then asked to engage in related school and home based activities. The effectiveness of Sesame Tree will again be assessed in relation to the four main outcomes described above and will also include a focus on childrens awareness of and understanding of emotions.

This paper will explore the practical and ethical issues of researching young childrens social attitudes and awareness, particularly in relation to assessing general willingness to be inclusive of others; attitudes towards differing cultural activities; racial attitudes and the attitudes of Catholic and Protestant children towards one another.

### 12.9.4 Methodological Challenges of Surveying the Roma: The Issues of Minority Sampling and Sensitivity of the Questionnaire Items

Cosima Rughinis, University of Bucharest (Romania)

Irina Tomescu-Dubrow, Polish Academy of Sciences (Poland)

**Abstract** This paper discusses methodological challenges researchers face in collecting and analyzing quantitative data on the Roma minority in Romania. Specifically, we are interested in two issues: (1) the operational definition the Roma minority a priori in the sample frame and post hoc in the actual data sets, and (2) the sensitivity of questionnaire items for this minority in comparison with the majority, or the general population. Estimates of the Roma minority in Romania range from 2.5% of the total population (2002 Census) to up to 7.9% (the World Bank). This discrepancy is primarily driven by the definitions used in identifying the target population: self-identification gives much lower estimates than hetero-identification, as Roma are reluctant to express their identity in a society that strongly discriminates against them. Using various survey data we analyze the consequences of four forms of delineating the Roma minority: (1) self-identification, (2) interviewers identification, (3) institutional identification, and (4) community identification. In particular, we demonstrate that demanding on the definition of Roma, the basic demographic characteristics of this minority such as age of

marriage, number of children, or household composition have different distributional properties. In addition, we analyze whether persons from the Roma minority tend to respond to different questions through the same type of answers. Specifically, we are interested in the extent to which response-set poses a measurement problem. We assess this through the proportion of don't know and yes answers across different questions for self-identified Roma in relation to non-Roma and hetero-identified Roma.

### 12.9.5 Surveying children: Methodological issues involved in the development and implementation of the Kids' Life and Times survey

Katrina Lloyd, Queen's University Belfast (United Kingdom)

**Abstract** Large-scale quantitative surveys of children are valuable as they can be used to compare data from across different age-groups and countries, and over time. However, the issues involved are complex. Children are not simply small adults; rather they are a special group whose intellectual abilities and attention span differ from those of adults. Quantitative research designed to be carried out with children needs to take account of these disparities along with a range of other factors such as ethics, consent and the use of questions on sensitive topics. The flexibility and confidential nature of computer-assisted self-administered interviews (CASI) are especially suitable for children and can be designed in ways that overcome some of these issues. Nevertheless, there are drawbacks associated with using CASI; for example children's responses may be affected by the knowledge that the information they provide is held locally on the computer and could potentially be accessed by parents or teachers. Kids Life and Times (KLT) is a large quantitative survey of Primary 7 children (age 10 and 11 years) in Northern Ireland which was carried out online and in schools in June 2008. Approximately 24,000 children within 918 primary schools were invited to take part. This presentation will discuss the issues involved in designing survey research for use with children in general, and examine the development and implementation of the KLT survey in particular.

### 12.9.6 Following Youth Out of Care, Observation Versus Interference: The Ethics of a Follow-Up

Géraldine Vivier, Institut National d'Etudes Démographiques (France)

Isabelle Frechon, Institut National d'Etudes Démographiques (France)

Pernelle Issenhuth, Institut National d'Etudes Démographiques (France)

**Abstract** This paper presents and discusses the ethical and methodological issues involved when conducting a prospective survey among a vulnerable population: young people (including minors) in state-run child care and on a sensitive topic: transition from care to adulthood.

Since previous studies have shown that youth leaving care are at high risk of social exclusion, this project responds to a strong social demand to shed light on the critical period following leaving care, when young people have to fend for themselves. The objective is also to overcome the limits and biases which affect the available data on this subject. Most leaving care studies are based on retrospective data collected among people who have left care for several years. Besides some memory and reconstruction biases, this kind of approach tends to leave out the most and the less integrated people. Thus, the ambition of this new project experimented in France is to follow up 1000 youth from age 17 to 23 through 3 data collection rounds, with several contacts in between.

Of course, such a prospective survey raises many critical methodological and ethical questions. In order to test its feasibility and design the conditions under which it should be undertaken (or not), qualitative interviews have been conducted with youth in care and a follow up of 100 young people over a six month period with two data collection rounds (before and after their leaving care) has been planned. If the sample can be drawn from the Child Welfare database, to be fully representative of the population, the first inclusion wave has to be conducted just before the leaving care, when most of the youth are still minors and live in foster families or residential settings. Since the great majority of these youth are under their parents' legal authority, a parental agreement is required (before proposing any interview to any child). This is a first key issue: depending on the parental agreement rate, selective sampling biases can be expected. From an ethical perspective, how does one deal with the parents legal rights and that of young people who have often broken off with their families? This protocol also implies cooperation from the care structures (where youth live) to get in touch with the youth drawn while keeping a correct distance with social workers (who should not interfere) and ensuring the confidentiality of responses.

Then, from a follow up perspective, the main difficulty remains the attrition. How to stay in touch with a population both very mobile and vulnerable? More specifically, how to avoid the selective attrition expectable



among young people who want to cut themselves off from their institutional past, and among youth in very precarious situations? How to reconcile the necessity of frequent contacts with people with no interference nor intervention in their lives, especially if very difficult moments occur?

Finally, the implications of the sensitivity of the topic and population on the questionnaire will also be presented.

# Index of authors and coordinators

- Abreu Lopes, Claudia , 29  
Alcser, Kirsten , 189, 190  
Alda, Holger , 26  
Alexandrowicz, Rainer , 157, 158  
Alisch, Lutz-Michael , 150  
Anheier, Helmut , 134  
Apio, Eunice , 210  
Asan, Zerrin , 53  
Astudillo, Javier , 38  
Ayhan, Öztas , 53, 56  
Bachner, Yaacov G. , 85  
Balluerka, Nekane , 175  
Barone, Carlo , 184  
Bartolomé Peral, Edurne , 45  
Bartsch, Simone , 115  
Bartusek, Ales , 184  
Bashkirova, Elena , 15  
Bassi, Francesca , 94  
Batista-Foguet, Joan Manuel , 12, 13  
Bauer, Gerrit , 112  
Baumgartner, František , 42  
Bechert, Insa , 178  
Beedham, Hillary , 76  
Behr, Dorothée , 187  
Bekkers, Rene , 45, 63, 134  
Belchior, Ana Maria , 38  
Benítez, Isabel , 84, 93, 169, 174  
Bensing, Jozien , 152  
Benson, Grant D. , 190, 191  
Bergh, Huub van den , 215  
Berglund, Frode , 61  
Bergseth, Gyrid Havag , 76  
Bergsma, Wicher , 159, 160  
Berkowitz, Susan G. , 172  
Berry, E. Helen , 148  
Berzelak , Nejc , 89  
Best, Henning , 112  
Betti, Gianni , 127  
Beuckelaer, Alain De , 6  
Beukenhorst, Dirkjan , 120  
Beullens, Koen , 194  
Beurs, E. de , 69  
Bialowolski, Piotr , 136  
Billiet, Jaak , 201–203  
Blake, Margaret , 84  
Blasius, Jörg , 178  
Blohm, Michael , 61  
Blom, Annelies , 196, 201  
Boeije, Hennie R. , 168  
Bonet, Eduard , 153  
Börsch-Supan, Axel , 190  
Bowers, Ashley , 189  
Boyatzis, Richard , 12, 13  
Boyd, Jason , 120  
Bradford, Ben , 140  
Branden, Laura , 188  
Braun, Michael , 144  
Braziene, Ruta , 102  
Breugelmans, Oscar , 68  
Brites, Rui , 10, 137, 138  
Bühlmann, Marc , 33  
Bulman, Joanna , 85  
Busse, Britta , 118  
Bynner, John , 71  
Cabeza, Laura , 34  
Calderwood, Lisa , 71  
Cambré, Bart , 10  
Cappelli, Maurizio , 11  
Cardenal, Ana Sofia , 38  
Cardoso, Inês , 138  
Carton, Ann , 197  
Castiglioni, Laura , 98  
Castillo, Miguel , 84, 169, 174  
Chanvriil, Flora , 136  
Chen, Tao , 96  
Chiodini, Paola Maddalena , 52  
Christine, Marc , 54  
Coelho, Bernardo , 137, 138  
Coenders, Marcel , 2–4, 165  
Collins, Debbie , 84  
Comsa, Mircea , 19  
Connolly, Paul , 226  
Conrad, Frederick G. , 92  
Consolini, Paolo , 127  
Cools, Mario , 40, 42  
Coromina Soler, Lluís , 31, 138  
Cote-Lussier, Carolyn , 141  
Couper, Mick , 92, 113  
Cousteaux, Anne-Sophie , 94  
Croon, Marcel , 159  
Czesnik, Mikolaj , 179  
D'Arrigo, Julia , 112  
Dalen, Joakim , 198  
Dalla Valle, Luciana , 54  
Dallinger, Ursula , 28, 43  
Dalton, Sue , 102  
Das, Marcel , 65  
Datta, Rupa , 199  
Davidov, Eldad , 6, 9, 10, 144, 151  
Davis, Arwenna , 120

- de Cledat, Bernard , 53, 68  
 Dean, Elizabeth , 221  
 Deboeck, Pascal R. , 149  
 Delhey, Jan , 43  
 Delsing, Marc , 150  
 DeMaio, Theresa J. , 168  
 Demarest, Stefaan , 22, 60  
 Devine, Paula , 225  
 Diederichs, Monika , 208  
 Diekmann, Andreas , 213  
 Dijkstra, Wil , 91  
 Dijkum, Cor van , 152  
 Doerr, Lauren , 195  
 Donatiello, Gabriella , 127  
 Draisma, Stasja , 69, 91  
 Drasch, Katrin , 99  
 Drieskens, Sabine , 60  
 Dubrow, Joshua K. , 147  
 Dülmer, Hermann , 6  
 Durrant, Gabriele , 71, 112  
 Edlund, Jonas , 178  
 Edwards, Brad , 181, 188  
 Eikelenboom, Merijn , 91  
 Elcheroth, Guy , 161  
 Elliot, Mark , 37  
 Ely, Margaret , 60  
 Engel, Uwe , 115  
 Eva, Gillian , 119  
 Faivre, Sébastien , 54  
 Farrall, Stephen , 140  
 Fasel, Nicole , 188  
 Fayyad, John , 191  
 Fazzi, Gabriella , 101  
 Ferligoj, Anuška , 35  
 Ferreira, Mafalda , 14  
 Ferreira de Almeida, João , 10  
 Ferrín Pereira, Mónica , 34  
 Fitzgerald, Rory , 86  
 Font, Joan , 138  
 Ford, Robert , 2  
 Fortin, Jessica , 178  
 Fox, Katherine , 25  
 Fraile, Marta , 138  
 Frechon, Isabelle , 227  
 Freire, André , 38  
 Frick, Joachim R. , 67  
 Fuchs, Marek , 117, 118, 219  
 Fumagalli, Laura , 72  
 Funke, Frederik , 88, 89  
 Gabler, Siegfried , 48–51, 56  
 Galesic, Mirta , 92  
 Gamrot, Wojciech , 67  
 Ganninger, Matthias , 50  
 Ganzeboom, Harry B.G. , 185  
 Gaspar de Matos, Margarida , 14  
 Gerber, Monica , 141  
 Gesthuizen, Maurice , 20  
 Getka-Wilczynska, Elzbieta , 65  
 Giesen, Deirdre , 26  
 Gisle, Lydia , 60  
 Glasner, Tina , 98, 187  
 Glorieux, Ignace , 40–42  
 Goldstein, Harvey , 71  
 Gómez Benito, Juana , 174  
 Gómez Fortes, Braulio , 34  
 Gordoni, Galit , 122  
 Gordoni, Yiftach , 122  
 Göritz, Anja , 88  
 Gorostiaga, Arantxa , 175  
 Graaf, Paul M. de , 15, 17, 19, 36  
 Grabka, Markus M. , 67  
 Graff-Iversen, Sidsel , 22  
 Gray, Michelle , 84  
 Grieg, Kai , 211  
 Groh-Samberg, Olaf , 67  
 Grotvedt, Liv , 22  
 Groves, Robert M. , 71  
 Guglhör-Rudan, Angelika , 154  
 Guillén, Laura , 12, 31  
 Gundelach, Peter , 16  
 Guo, Yan , 148  
 Guyer, Heidi , 190  
 Häder, Sabine , 117  
 Hagenaars, Jacques , 106, 159  
 Halman, Loek , 15  
 Hanquinet, Laurie , 171  
 Hansen, Sue Ellen , 181, 189  
 Haraldsen, Gustav , 25  
 Harkness, Janet , 181, 182, 187–189  
 Harms, Torsten , 223  
 Harris-Kojetin, Lauren , 181  
 Harrison, Eric , 80  
 Hartmann, Peter , 185  
 Hatemi, Peter K. , 43  
 Häuberer, Julia , 35  
 Hausstein, Brigitte , 78  
 Heldal, Johan , 22  
 Hempel, René , 150  
 Heus, Manon de , 184  
 Heyden, Johan Van der , 60  
 Hicks, Wendy , 181  
 Hidalgo, María-Dolores , 174  
 Hill, Craig A. , 221  
 Hlebec, Valentina , 35  
 Hoekstra, Aafke T. , 156  
 Hoekstra, Chantal , 80  
 Hoelscher, Michael , 134, 135  
 Hoffmeyer-Zlotnik, Jürgen H.P. , 78, 183, 217  
 Hohl, Katrin , 162  
 Höhne, Inga , 185  
 Holleman, Bregje , 214, 215, 217  
 Hoogendoorn, Adriaan W. , 80  
 Hopt, Oliver , 102  
 Hough, Mike , 140  
 Hox, Joop , 196  
 Huang, Lihong , 164  
 Hufken, Volker , 57  
 Hughes, Sarah , 194, 195  
 Huismans, Sipko , 7  
 Huizi, P. , 175

- Hulliger, Beat , 127  
Huynen, Bart , 204  
Iatu, Corneliu , 223  
Iedema, Jurjen , 109  
Isasi, Xabier , 175  
Issenhuth, Pernelle , 53, 68, 227  
Jaastad, Atle , 221  
Jablonski, Wojciech , 199  
Jäckle, Annette , 99  
Jackson, Jonathan P. , 140, 141  
Jagodzinski, Wolfgang , 15, 17, 19  
Jakubowski, Maciej , 106, 157  
Janik, Florian , 224  
Jansen, Margo G.H. , 156  
Jelonek, Magdalena , 132  
Jerónimo, Paula , 108  
Johnson, Timothy , 144  
Jones, Jacqui , 25  
Joye, Dominique , 115, 161, 203  
Jürges, Hendrik , 191  
Kalgraff Skjåk, Knut , 178  
Kaminska, Monika E. , 20  
Kaminska, Olena , 203  
Kamoen, Naomi , 214, 215, 217  
Kankarash, Milos , 144  
Kappelhof, Joost , 145  
Karaganis, Milana , 113  
Karam, Alimee Nasser , 191  
Karam, Elie G. , 191  
Kaspar, Hanna , 61  
Keler, Karolina , 132  
Ketende, Sosthenes , 71  
Kettner, Anja , 181  
Khroul, Victor , 109  
Kieffer, Annick , 184  
Kiersztyn, Anna , 147  
Kiesl, Hans , 26  
Kleiner, Brian , 188  
Kleven, Øyvinn , 61, 198  
Koch, Achim , 61, 196  
Kogovšek, Tina , 35  
Kohaut, Susanne , 224  
Kolsrud, Kirstine , 74, 76  
Koponen, Päivikki , 22  
Kordos, Jan , 48  
Korkala, Siru , 36  
Koster, Ferry , 20  
Kotnarowski, Michal , 179  
Kowalczyk, Barbara , 69  
Krebs, Dagmar , 85, 217  
Kreuter, Frauke , 112, 113, 198  
Krieger, Ulrich , 98  
Kriesi, Hanspeter , 33  
Kritski, Eugene , 212  
Kriwy, Peter , 185  
Krumpal, Ivar , 97  
Krupnik, Seweryn , 171  
Kruse, Yelena , 188  
Kunz, Ruth , 33  
Kutylowski, A. Jan , 139, 154  
Kuulasmaa, Kari , 22  
Laaksonen, Seppo , 48, 51, 56, 145  
Laflamme, François , 113  
Lam, Niek , 152  
Landsheer, Johannes A. , 150  
Lange, Cornelia , 59  
Langmeyer, Alexandra , 154  
Larkin, Emma , 226  
Latcheva, Rossalina , 131  
Laurent, Raphael , 53, 68  
Le Hay, Viviane , 64  
Lee, Sabine , 210  
Leeuw, Edith De , 196  
Leicester, Andrew , 72  
Lensvelt-Mulders, Gerty , 63, 65  
Leontiyeva, Yana , 225  
Leuenberger-Zanetta, Sylvie , 115, 203  
Liapis, Spyros , 222  
Lippe, Elena von der , 59  
Lipps, Oliver , 190, 203  
Lloyd, Katrina , 227  
Lobe, Bojana , 171  
Londen, Marieke van , 4  
Longford, Nick , 129  
Loosveldt, Geert , 194, 196, 202  
Lopes, Rui , 5  
Lozar Manfreda, Katja , 89  
Ludwig, Volker , 98  
Lutig, Peter , 98, 99  
Luijckx, Ruud , 15, 17, 19, 106, 107, 184  
Luiten, Annemieke , 198  
Lupparelli, Monia , 160  
Lupu, Adrian Lucian , 223  
Lynn, Peter , 72  
Lysa, Olha , 130  
Maas, Loes , 187  
Macchia, Stefania , 94, 101  
Macri, Ester , 181  
Maggino, Filomena , 131  
Magun, Vladimir , 7  
Manabe, Kazufumi , 18  
Manca, Anna Rita , 9  
Marasini, Donata , 52  
Marchand, Miquelle , 65  
Marchetti, Giovanni M. , 160  
Mariotti, Luca , 151  
Markou, Efi , 53, 68  
Marquart-Pyatt, Sandra T. , 147, 148  
Martí, Mónica , 56  
Martin, Nicholas G. , 43  
Martini, Maria Cristiana , 136  
Martire, Fabrizio , 101  
Mascherini, Massimiliano , 9  
Matschinger, Herbert , 157, 158  
Matsuo, Hideko , 202  
Matthes, Britta , 70, 99  
Mauer, Reiner , 75  
Maxwell, Christopher , 77  
Mayerl, Jochen , 214  
McCutcheon, Allan , 203

- McFarland O'Rourke, JoAnne , 77  
 McHugh, Sarah , 80  
 Medina, Lucia , 39  
 Melgar, Natalia , 179  
 Melhem, Nadine , 191  
 Méndez, Mónica , 132, 138  
 Menold, Natalja , 204  
 Merkys, Gediminas , 102  
 Meuleman, Bart , 10, 161, 163  
 Meulemann, Heiner , 28, 29  
 Meulen, Bieuwe F. van der , 156  
 Meyer, Reto , 213  
 Mikulic, Branislav , 222  
 Minnen, Joeri , 40, 41  
 Mneimneh , Zeina N. , 191, 192  
 Mochmann, Ingvill C. , 208, 210  
 Moon, Nick , 80  
 Moons, Elke , 40  
 Moors, Guy , 144  
 Morren, Mattijn , 26  
 Mortensen, Anne Katrine , 183  
 Moss, Abbie , 181  
 Mueller, Gerrit , 113  
 Muñoz, Jordi , 153  
 Murgia, Manuela , 101  
 Naef, Michal , 44  
 Näher, Anatol-Fiete , 97  
 Nandi, Alita , 84  
 Németh, Renáta , 160  
 Nicol, Simon , 117  
 Nicolini, Giovanna , 54  
 Nolen, W.A. , 69  
 Noordhuizen, Suzanne , 36  
 Oberski, Daniel , 165  
 Oirschot, Lonneke van , 45  
 Øland, Arne , 208  
 Oldfield, Zoë , 72  
 Ongena, Yfke P. , 91, 93  
 Oravec, Zita , 151  
 Orłowski , Rachel A. , 189  
 Orriols Galve, Lluís , 39  
 Orten, Hilde , 183  
 Oud, Han , 149, 151  
 Padilla, Jose Luis , 84, 93, 169, 174  
 Palacios Brihuega, Irene , 34  
 Paniotto, Volodymyr , 37  
 Parent Thirion, Agnes , 222  
 Park, Alison , 81  
 Paulinger, Gerhard , 32  
 Pavlova, Veselka , 223  
 Pavsic, Rita , 101  
 Pedraza, Pablo de , 58, 63, 90, 123  
 Pennell, Beth Ellen , 189, 191, 192  
 Pereira, Cícero , 4, 5  
 Perek-Bialas, Jolanta , 75  
 Petrakos, George , 222  
 Petrakos, Michalis , 222  
 Petrenko, Tatyana , 37  
 Pforr, Klaus , 98  
 Pfuhl, Nadja , 154  
 Phalet, Karen , 3  
 Phillips, Miranda , 81  
 Pichler, Florian , 2  
 Pitrone, Maria Concetta , 101  
 Plagnol, Anke , 126  
 Platt, Lucinda , 84  
 Plewis, Ian , 71  
 Pokropek, Artur , 50  
 Polak, Regina , 21  
 Pollien, Alexandre , 115, 190, 203  
 Popping, Roel , 95  
 Poulton, Richie , 60  
 Putte, Bart Van de , 116  
 Quandt, Markus , 76, 180  
 Quatto, Piero , 52  
 Raeside, Robert , 96  
 Ramos, Alice , 4  
 Razafindratsima, Nicolas , 53, 68  
 Read, Sanna , 43  
 Régnier-Loilier, Arnaud , 96  
 Reinecke, Jost , 151  
 Reips, Ulf-Dietrich , 88, 89  
 Renner, Katharina , 21  
 Renschler, Isabelle , 188  
 Revilla, Melanie , 119  
 Riele, Saskia te , 31  
 Ringdal, Kristen , 61, 198  
 Roberts, Caroline , 122  
 Robins, Cynthia S. , 168  
 Ródenas, Carmen , 56  
 Rohrbach, Daniela , 26  
 Rojas Gonzalez, Gara , 127  
 Rossi, Máximo , 179  
 Rudas, Tamás , 160  
 Rudnev, Maksim , 7  
 Rughinis, Cosima , 226  
 Ruland, Michael , 70  
 Rush, Sarah H. , 77  
 Rusu, Horatiu , 19  
 Ruviglioni, Elena , 133  
 Rybkowska, Anna , 127  
 Saar, Ellu , 184  
 Sala, Emanuela , 96  
 Salamoun, Mariana , 191  
 Sapin, Marlene , 115, 203  
 Saris, Willem E. , 12, 13, 31, 119, 122, 136, 138, 164  
 Sauger, Nicolas , 197  
 Schaan, Barbara , 194  
 Scheepers, Peer , 4  
 Schlaud, Martin , 23  
 Schlüter, Elmar , 6, 144, 163  
 Schmeets, Hans , 31, 33, 204  
 Schmich, Patrick , 59  
 Schmid, Veronika , 171  
 Schmidt, Peter , 6, 9, 122, 151  
 Schnabel, Christiane , 115  
 Schneider, Silke , 184  
 Schneiderat, Götz , 117  
 Schnell, Rainer , 102  
 Schöbi, Nicole , 115, 188, 203

- Schröder, Heike , 185  
Schröder, Mathis , 22, 23, 219  
Schupp, Jürgen , 44  
Schoor, Wijnbrandt van , 7, 153, 156  
Schwartz, Shalom , 9  
Sciarra, Caterina , 11  
Scott, Jacqueline , 126  
Segatti, Paolo , 153  
Sellke, Piet , 215  
Serlavos, Ricard , 12, 13  
Sieben, Inge , 36  
Sieggers, Pascal , 18, 28  
Simões, Celeste , 14  
Sindall, Katy , 140  
Singer, Eleanor , 113  
Sipahi Dantas, Alaíde , 12, 13  
Skovgaard Christensen , Rikke , 16  
Slomczynski, Kazimierz M. , 79, 147  
Slot, Jeroen , 45  
Smit, Johannes H. , 69, 80, 91  
Smith, Tom W. , 115, 179  
Smits, Jeroen , 163  
Snijkers, Ger , 25, 26, 168  
Spengler, Norman , 134  
Spini, Dario , 161  
Staetsky, Laura , 71  
Stares, Sally , 134  
Stark, Tobias , 92  
Steele, Clarissa , 188  
Steele, Fiona , 71, 112  
Steinkopf, Leander , 112  
Steinmetz, Stephanie , 58, 63, 90, 123  
Stelzl-Marx, Barbara , 209  
Stempfhuber, Max , 102  
Stenger, Horst , 49  
Stepanova, Elena , 172  
Stieger, Stefan , 88  
Stocké, Volker , 92  
Stoop, Ineke , 108, 109, 201, 203  
Stops, Michael , 181  
Stoykova-Kanalieva, Andreana , 129  
Sturgis, Patrick , 43, 45, 71  
Suerdem, Ahmet K. , 87  
Swyngedouw, Marc , 3  
Szalaniec, Henryk , 157  
Tabet, Caroline Cordahi , 191  
Tafforeau, Jean , 60  
Tarnai, Christian , 154  
Tart, Indrek , 8  
Taylor, Jane , 25  
Taylor, Rebecca , 71  
Teney, Celine , 3  
Terwey, Michael , 74  
Thiessen, Victor , 178  
Thomas, Randall K. , 89  
Thorsdottir, Fanney , 156  
Tijdens, Kea , 58, 63, 90, 123  
Tikir, Aysel , 212  
Tobi, Hilde , 187  
Todosijevic, Bojan , 106  
Tolonen, Hanna , 22  
Tomczyk, Emilia , 69  
Tomé, Gina , 14  
Tomescu-Dubrow, Irina , 147, 226  
Tormos, Raul , 148  
Torres, Analia , 10, 108, 137, 138  
Tourangeau, Karen , 181  
Tourangeau, Roger , 92  
Tranmer, Mark , 37  
Trappmann, Mark , 113  
Trivellato, Ugo , 94  
Trull, Tim , 43  
Tuerlinckx, Francis , 151  
Tufis, Paula Andreea , 17  
Uhl, Hanna , 171  
Uhrig, Noah , 96  
Ünalán, Turgay , 56  
Väisänen, Paavo , 41  
Vala, Jorge , 4, 5  
Vandezande, Veronique , 3  
Vehovar, Vasja , 67, 69, 89  
Vehre, Helen , 115  
Veld, William van der , 164  
Verbakel, Ellen , 20, 106, 107  
Vercruyssen, Anina , 116  
Verheul, William , 152  
Verma, Vijay , 126  
Vermeylen, Greet , 222  
Vermunt, Jeroen , 144  
Veykher, Andrey , 57  
Vezzoni, Cristiano , 153  
Villar, Ana , 188  
Vinopal, Jiri , 38  
Viorica, Elena Daniela , 223  
Vis, Corrie , 219  
Vivier, Géraldine , 227  
Voelke, Manuel C. , 151  
Vogl, Susanne , 91, 218  
Voicu, Malina , 17  
Wallis, Ruth , 102  
Wang, Ying , 188  
Warner, Uwe , 78, 184  
Weaver, Bryce , 70  
Weber, Wiebke , 39  
Weinhardt, Michael , 198  
Wenemark, Marika , 173  
Wenninger, Agnieszka , 78  
Wets, Geert , 40  
Wetzels, Willem , 120  
Weziak-Bialowolska, Dorota , 136, 157  
Widdop, Sally , 101  
Wiegand, Gordon , 127  
Wiepking, Pamala , 134  
Willis, Gordon , 168  
Willmann, Johanna , 31  
Wolf, Christof , 183  
Wright, Margaret J. , 43  
Wroblewski, Angela , 131  
Wywiał, Janusz , 51  
Yan, Ting , 92, 199, 217

Zabal, Anouk , 80  
Zagorski, Krzysztof , 42  
Zandbelt , T. , 69  
Zandberg, Izabella , 172  
Zhu, Gu , 43  
Zouwen, Johannes van der , 91  
Zuber, Sabrina , 219  
Zuell, Cornelia , 108, 204  
Zwingenberger, Anja , 102