Abstracts

July 4, 2005

Contents

1 Comparative Research (Aula 002)

1.1 Comparative voting analysis: analysis of pooled data, multiple group or multilevel

Session chair:
  Hanspeter Kriesi; University of Zürich, Switzerland

This session takes place on:
  Tuesday, July 19; 9:00 to 11:00
  Tuesday, July 19; 11:30 to 13:30

Presentations:

- Invited presentation: *Heuristic short-cuts and argument-based voting in direct-democratic choice* - Hanspeter Kriesi (Switzerland)

Contributed presentations:

- *Unusual election result explained by usual models for subgroups: The Dutch 2002 elections* - Daniel Oberski (Netherlands), Willem Saris (Spain), Abdelhafid Ballafkih, Peter Lugtig, Linda Bos (Netherlands)

- *Class voting in eight West European countries: trends and explanations* - Oddbjørn Knutsen (Norway)

- *The Vlaams Blok vote in Antwerp: a multilevel approach* - Koen Abts, Mark Swyngedouw (Belgium)

- *From unstable preferences to unstable choices: Macro-micro interactions and the explanation of electoral volatility* - Romain Lachat (Switzerland)

- *Economics, Ideology and Vote: Southern Europe, 1985-2000* - André Freire, Marina Costa Lobo (Portugal)

- *Differences between electorates in old and new EU member states* - Wouter van der Brug (Netherlands)

- *Individual and Geographical variations in the electoral participation of young people: evidence from a multilevel analysis of the European Social Survey* - Edward Fieldhouse, Mark Tranmer (United Kingdom)

- *Retrospective Voting in 70 Countries - Linking Individuals to Context* - Mark Bühlmann (Switzerland)

1.1.1   Heuristic short-cuts and argument-based voting in direct-democratic choice

Hanspeter Kriesi; University of Zürich, Switzerland

Abstract.   After some short introductory remarks on the different ways to analyze data from heterogenous populations, I shall present the design and key results from my own research on the direct-democratic choice of Swiss voters. Following the dual-process theories in social psychology (Chaiken and Trope 1999, Eagly and Chaiken 1993: 306f.), I distinguish between two qualitatively different paths of individual opinion formation: a heuristic and a systematic path. The distinction between these two fundamentally different ways of opinion formation is based on the role played by arguments. Systematic opinion formation is essentially argument-based, while heuristic opinion formation is essentially based on shortcuts, which do not make any reference to substantive arguments. In my presentation, I shall discuss the relative importance of the two paths for reaching a voting choice in Swiss direct-democratic decisions. My analysis is based on the combination of two unique data-sets covering the direct-democratic Swiss votes during the period 1981-1999: an individual level data-set containing the result of surveys with samples of Swiss citizens held after the 49 federal votes that took place during this period, and an aggregate-level data-set containing information on the 148 proposals and the related issue-specific campaigns which have been submitted to the vote on these 49 occasions. The analysis will use two-level models estimated by HLM.

In direct-democratic votes, the choice of the individual voters is prestructured by the issue-specific coalitions among the political elite that form preceding the vote and by the mobilization efforts made by these coalitions during the issue-specific campaign. These coalitions and their campaigns provide the relevant partisan cues as well as the issue-specific arguments for the individual voters. In my presentation, I shall first discuss evidence that Swiss voters heavily rely on heuristics: they use partisan cues and and vote as a function of their trust in the government. Next, I shall present evidence that Swiss voters tend to be less minimalist than is generally assumed: argument-based voting turns out to be rather widespread among them. The extent of argument-based voting depends on the context: I shall show that it is most important, when campaigns are intense and projects are familiar.

Third, I shall discuss the relative importance of the two types of making voting choices for different kinds of voters: the results suggest that voters who have a strong opinion on a given issue do not need to rely on heuristic shortcuts. This is not to say that voters with strong opinions do not use heuristic cues at all, but it is to suggest that such voters are rather more likely to resist inconsistent cues: They tend to use heuristic cues to invalidate messages that threaten their opinions or to validate messages that further strengthen them. In whatever ways such voters arrive at their issue-specific opinions, the very strength of their opinions serves as a clear guide for their voting choices. By contrast, voters who do not quite know what to think about the issue in question, who are ambivalent or undecided about them or who simply ignore the arguments in favor or against the project to be decided are more likely to rely on heuristics. In fact, such voters do not seem to have an alternative to a heuristic choice. The analysis suggests that a) intensive campaigns and familiar projects develop their effect by reducing the relevance of heuristic strategies for the voters with weak opinions and b) that it is above all the voters with weak opinions in the opposition's camp who benefit from the clarifying effect of intensive campaigns and familiar projects. We may conclude that systematic voting is generally enhanced by clearly structured situations where the voters know the stakes at hand.

1.1.2   Unusual election result explained by usual models for subgroups: The Dutch 2002 elections

Daniel Oberski; University of Amsterdam, Netherlands

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abdelhafid Ballafkih; University of Amsterdam, Netherlands

Peter Lugtig; University of Amsterdam, Netherlands

Linda Bos; University of Amsterdam, Netherlands

Abstract.   The 2002 elections in the Netherlands were very exceptional with respect to the volatility of the voters. This suggests that unusual explanations are needed. We show that this is not necessarily the case. Using standard variables from voting studies and standard theories about voting behavior we can explain what happened in these

elections quite well. The major difference with earlier studies of these elections is that we split the sample up in sub-samples and show that in the different sub-samples different commonly accepted voting theories apply. Using a Structural Equation Modelling (SEM) approach with correction for measurement error, we have evaluated the differences in effects that occur when conditioning on voters' political sophistication, external political efficacy and politcal cynicism. Furthermore the correction for measurement error contributes to a higher explained variance than is normally obtained.

1.1.3 Class voting in eight West European countries: trends and explanations

Oddbjørn Knutsen; University of Oslo, Norway

Abstract. The paper will use Eurobarometer data from eight West European countries and examine trends in class voting from the midst 1970s to the late 1990s by means of three ways of treating the social class and party choice:

1) Total class voting which is the difference in voting between all parties in the party system. I use a variant of the Erikson/Goldthorpe class schema to tap social class and Cramer's V to tap total class voting.

2) Overall or total left-right class voting which is the left-right voting based on all social classes. I use the so-called kappa-index based on log-odds ratios to tap this type of class voting.

3) Traditional left-right class voting based on a dichotomous class variable (manual versus non-manual classes. I use the Thomsen index based on log-odds ratios and also the Alford index based on percentage differences to tap this type of class voting.

I then go on to examine possible explanations of the cross-national and cross-time patterns in the various measures of class voting. I discuss a series of variables which have been used to explain changes in and comparative strength of class voting. These variables are economic structural variables such as economic prosperity level, unemployment level, employment in industry, size of the working class, income inequality and union density. I also include religious structural variables such as religious fragmentation and the size of the secular segment in the population. Political factors such as the overall left-right polarization of the party system and the distance between the larger left and right party are also considered. Finally I discuss the impact of new politics value orientations on the party system as a possible cause of change in class voting. I also consider the relationship between economic left-right value and class voting.

I then perform analyses of causes of changes in class voting within each of the eight countries and analyses of cross-national differences in class voting within the various time periods where class voting is studied. Finally I collapse the correlation measures for the various time periods in order to get a larger number of cases and to present an overview of the relationships between the various explanatory variables and the levels of class voting. This is a pooled cross-sectional time-series data set which comprises 40 units, each unit being one country for a given time period.

1.1.4 The Vlaams Blok vote in Antwerp: a multilevel approach

Koen Abts; University of Leuven, Belgium

Mark Swyngedouw; University of Leuven, Belgium

Abstract. In this paper we set out to improve our knowledge of voting for extreme-right parties, in this case the Vlaams Blok, by taking into account individual and contextual characteristics simultaneously. We will test diverse theories that provide possible explanations as to why certain social categories are more likely to vote for the Vlaams Blok. The hypotheses will be tested by means of multilevel analysis, using data from a sample collected in 1999 in the city of Antwerp. Between districts, variances exist in the extent to which respondents voted for the Vlaams Blok. In public opinion, these variances are always explained in terms of the number of ethnic minorities that live in the district or in terms of unemployment level. Nevertheless, in Belgium there is never done scientific analysis of the contextual basis of the extreme right success at city level.

Antwerp, the cradle of radical Flemish nationalism, developed itself as a bulwark of right-wing extremism. Since the elections of 1985, support for the Vlaams Blok has only been on the increase. In 1994 the party became even the biggest party at city level. In the local elections of 2000, the Vlaams Blok increased its share of votes to 33%, with peaks of roughly 38% in three of the nine boroughs of Antwerp. Since Antwerp is the center of Vlaams Blok support in Flanders, it is on the one hand crucial to develop a theoretical framework addressed

at the contemporary urban context, in which theoretical perspectives of late modernity and of urban sociology are combined with insights of the literature of extreme right-wing parties. To test the hypotheses it is, on the other hand, necessary to test contextual characteristics simultaneously with individual characteristics, within an administrative, political and discursive unit.

First of all, the general election studies has shown that certain social categories were more apt to vote for the extreme right-wing party: non-manual workers, lower educated people, non-religious people and voters younger than 25, in particular, were more likely tot do so. Furthermore Vlaams Blok voters were found to be more ethnocentric, authoritarian and politically alienated. Contrary to some studies we want to see the Vlaams Blok vote not as an isolated behaviour. Therefore, the vote is linked to the dialectical structure of (changing) structural context and urban life world. At the individual level, the structural positions within the different social systems are linked to the subjective perceptions of the structural context. The question is how the extreme right vote can be explained by social structure and life world. To what extent have structural individual characteristics direct effects to the Vlaams Blok vote or are there especially indirect effects trough feelings of insecurity, perceptions of neighbourhood changes, and general values. The point of departure is that the Vlaams Blok vote in an urban setting is a result of two related factors. On the one hand a very dynamic threatening and always less understandable and controllable setting of high complexity. On the other hand a population, which is –as a result of their social, economic and cultural living conditions and the drastic neighbourhood changes –very receptive to those insecurity stimuli; very receptive to the supplied anti-immigrant and anti-politics discourses in a transformed discursive space. With respect to the empirical model this framework has the consequence that the neighbourhood context as well as the (specific) perceptions/feelings of urban life has to be included simultaneously in the analysis.

The improvement of the model would be achieved partly because we also take into account characteristics of districts, therefore going beyond micro-level. We set out to test whether contextual characteristics, which are considered to be important in shaping the extreme right-wing vote, are contributory in explaining Vlaams Blok voting behaviour. Our test will be based on multilevel analysis, a technique that allows simultaneous tests of individual and contextual effects. We make use of the two-stage sampled Election Study 'Antwerp and their citizens' of 1999 from ISPO (Institute of Social and Political Opinion research, University of Leuven), in which characteristics and attitudes of 938 respondents of 65 different districts of Antwerp are stored. At the individual level the data is a combination of social background variables (age, occupation, education, church involvement, job insecurity, association membership), feelings of everyday life world (operationalised by (1) experiences of powerlessness, uncertainty and isolation, (2) feelings of relative deprivation like discrimination, deterioration and marginalisation; (3) evaluations of neighbourhood changes; and (4) feelings of social trust and insecurity), and general value orientations like individualism, authoritarianism and conservatism. At the district level we have collected data concerning the (changes in the) unemployment, number of immigrants, income, housing and criminality in each of these districts.

We are looking for as well multinomial analysis where the dependent variable contains all electoral significant parties in the cities. However if this approach fails because of technical reasons we will continue with the binomial case. In the binomial analysis the dependent variable is (the intention to) vote for the Vlaams Blok versus not (the intention to) vote for the Vlaams Blok. So, we perform a logistic regression, within the program MlwiN. By introducing characteristics of context, we want to find out whether the variance can be explained by the theoretically proposed characteristics of the districts, whereas the introduction of characteristics at the individual level makes it possible to control for composition effects at the contextual level. Furthermore, multilevel deals with dependent observations, which result from a two stage sample. More substantially, multilevel could give further insight in the influence of neighbourhood composition in regard to the success of Vlaams Blok in an urban context.

We will end the paper with a critical review of the advantages and disadvantages of the use of multilevel analysis for concrete political survey data.

1.1.5 From unstable preferences to unstable choices: Macro-micro interactions and the explanation of electoral volatility

Romain Lachat; University of Zürich, Switzerland

Abstract. This paper analyses individual-level and contextual-level determinants of electoral volatility in Swiss elections. In the last decades, a large body of literature in electoral research has focused on the dynamics of electoral dealignment and realignment. The factors that once contributed to the stability of voters' party preferences

and of party systems have weakened or changed. Voters today are expected to be more independent from parties and/or to rely less strongly on traditional loyalties or conflicts. Following this literature, one of the major consequences of these evolutions should be a higher level of instability in voters' decisions, as reflected among others in levels of electoral volatility. Furthermore, one of the main characteristics of the literature on electoral dealignment/realignment is that it is expected to be a general phenomenon. These processes are not rooted in social and political evolutions that are specific to different countries. On the contrary, it is argued that this phenomenon is a consequence of a general process of social modernization and value change that affects all democracies.

In this paper, we address the latter aspect of the dealignment argument. We consider that while a tendency towards a higher level of volatility may be observed in many electoral contexts, its extent should be related to contextual level factors. Most important among these are the characteristics of voters' decision context. The likelihood that unstable preferences actually translate into unstable choices will be related to the fractionalization and polarization of the party system. This means, also, that the strength of the relationships between individual-level factors (party identification strength, political sophistication) and electoral volatility will be conditioned by contextual characteristics. In other words, it is only by accounting for both types of factors, at the micro and macro level, that we can get a better understanding of the causes and of the process of electoral change.

To explore these interactive effects between micro and macro factors, we analyse in this paper the determinants of electoral volatility in Swiss national elections, over the last three decades. Switzerland is a very appropriate case, as it allows us to contrast, for each national election, twenty-six constituencies which present much variation in the characteristics of their party systems. On this basis, we test a series of multilevel models of electoral volatility, including both direct effects and cross-level interactions. At the individual-level, we rely on our own model of the instability of voting choices, combining social-structural predispositions (party identification, cleavage structure) and political sophistication. At the macro level, we consider three characteristics of voters' decision context: constituency size, and the fractionalization and polarization of the party system.

1.1.6 Economics, Ideology and Vote: Southern Europe, 1985-2000

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

Marina Costa Lobo; Higher Institute for Labour and Business Studies (ISCTE), Portugal

Abstract. This paper analyses the role of economics in voting behaviour in Greece, Portugal and Spain, using both individual level data (on economic perceptions, social cleavages, and left-right self-placement) from the Eurobarometers (1985-2000), and objective macroeconomic indicators in pooled datasets. First, the objective economic conditions in which countries found themselves are described, and the degree to which they reflect the electorates' economic perceptions is investigated. Second, a multi-level model of voting behaviour for Greece, Portugal and Spain is tested, to determine if, as expected, economics is more important than social and political cleavages. This approach permits assessing the effects of both individual voters' economic perceptions and objective economic conditions to determine which are more important, and to compare their effects with social and political cleavages. Except for the Greek case, the economic voting model will also be tested under different political conditions, i.e. type of government in each country.

1.1.7 Differences between electorates in old and new EU member states

Wouter van der Brug; University of Amsterdam, Netherlands

Abstract. Research question

On the basis of their analyses of the 1989 and 1994 European Election Studies, Van der Eijk and Franklin (1996: 364) concluded that in some important respects the electorates of the 12 member states of the EU could be considered to form one single electorate. This conclusion was based on the fact that "the relative effects of cleavages, issues, ideology, and approval of the government on the choices made by individual voters are quite similar in the various countries of the European Union (...)". Surely, there are important differences in the outcome of the electoral processes in the various member states. Such differences in election results are, however, not caused by differences among the electorates in the effects of the determinants of party choice. Rather, they are due to differences in the distributions of the independent variables, differences at the supply side (parties and their programs), and institutional differences. The proposed paper addresses the question whether the general

conclusion of Van der Eijk and Franklin (1996) is still valid in the enlarged Union in 2004. Is this conclusion still valid for the old member states, and is it valid for the new ones?

Method

Survey data from the European Elections Studies 2004 will be employed, which were organized in 24 European member states. For all of the significant parties in each of the EU countries these surveys contain the question "how likely it is that a respondent will ever vote for it". The answers from these questions will be used to arrange the data in a so-called 'stacked' form, in which the party/respondent combination becomes the new unit of analysis (e.g., Van der Eijk & Franklin 1996). This way of rearranging the data enables us to do a pooled analysis across all EU countries. Since it combines data at different levels of aggregation (more than 20,000 individual respondents, more than 100 parties and across 24 contexts), this design yields sufficient variation at the micro- as well as at the macro-level to estimate their effects. The basic elements of our research design are explained in more detail elsewhere (e.g., Stimson, 1985; Macdonald, Listhaug and Rabinowitz, 1991; Van der Eijk and Franklin, 1996; Van der Eijk, Franklin and Van der Brug, 1999; Van der Brug, Fennema and Tillie, 2000).

References

Brug, W. van der, M. Fennema and J. Tillie (2000). Anti-immigrant parties in Europe: Ideological or protest vote? European Journal of Political Research 37:77-102.

Eijk, C. van der and M. Franklin (1996). Choosing Europe? The European Electorate and National Politics in the Face of Union. Ann Arbor: The University of Michigan Press.

Eijk, C. van der, M. Franklin, and W. van der Brug (1999). Policy Preferences and Party Choice. In: H. Schmitt and J. Thomassen (eds.), Political Representation and Legitimacy in the European Union. Oxford: Oxford University Press.

Eijk, C. van der, M. Franklin, and E. Oppenhuis (1996). The Strategic Context. In: C. van der Eijk & M. Franklin (eds.), Choosing Europe? The European Electorate and National Politics in the Face of Union. Ann Arbor: The University of Michigan Press.

Macdonald, S.E., O. Listhaug, and G. Rabinowitz (1991). Issues and Party Support in Multiparty Systems. American Political Science Review 85:1107-31.

Stimson, J.A. (1985). Regression in Space and Time: A Statistical Essay. American Journal of Political Science 29: 914-47.

1.1.8 Individual and Geographical variations in the electoral participation of young people: evidence from a multilevel analysis of the European Social Survey

Edward Fieldhouse; University of Manchester, United Kingdom

Mark Tranmer; University of Manchester, United Kingdom

Abstract. Turnout at general elections across Europe is in decline. For example in the U.K. in 2001 turnout declined to 59%, the lowest level since 1918. Reported turnout in the ESS across the 21 countries for which data are currently available is 74% (though this is likely to be an over-estimate due to non-response bias and misreporting of vote). Rates for young voters, those aged 18-24 tends to be significantly lower than this. This paper examines individual and national variations in turnout amongst young people, and uses multilevel models to understand the extent to which the propensity of young people to vote reflects different attitudes and characteristics, or simply mirrors the general level of participation in particular elections.

Previous studies have provided various explanations for individual level variations in turnout. Thus we turn to three major theories to explain these variations; rational choice, social capital and civic voluntarism. We include a wide range of covariates to represent various aspects of these theories. In addition we explore how young people vary in their participation across different countries, and how well these differences between such groups are accounted for by these models. The results show that some of variation by country simply reflects the level of participation in countries concerned, whilst some reflects common attributes and attitudes of young people across all countries. Finally, we look at the extent to which young people are more likely to turn to alternative forms of participation such as political protest

1.1.9 Retrospective Voting in 70 Countries - Linking Individuals to Context

Mark Bühlmann; University of Zürich, Switzerland

Abstract. The economic political theory provides us with a very simple idea of voting behavior: Voters, evaluating the performance of their government, are capable of rewarding political incumbents for good times and of blaming them for bad times. Since the origin Downsian (1957; see also Key 1966) model many studies not only showed the importance of retrospective evaluation for voting behavior but also contributed to a greater complexity of the model (the literature is extensive, for an overview see MacKuen et al. 1992): we have to distinguish between sociotropic and egocentric vot-ing (Alvarez and Nagler 1995), and between prospective and retrospective evaluations (Fiorina 1981). We also have to take into account abstention behavior. Satisfaction as well as dissatisfaction with the own or/and the overall national situation can lead to abstention decisions (Radcliff 1992). Context plays the key role in the models providing voters with different issues, above all economic, to evaluate the government's performance. Additionally, the institutional context has much impor-tance as a pre-condition for voting. Institutional settings influence the ability of the electorate to assign responsibility for good or bad situations (Powell and Whitten 1993).

However, most of the studies only analyse either aggregate or individual levels of retrospective voting. The standard problems associated with these two approaches are well known. Aggregate-level approaches assume that macroeconomic situations influence all individuals in the same way, re-gardless of their particular characteristics. Individual-level analyses do not take into consideration contextual influences at least not in a methodologically appropriate manner. We do not have clear results which enable us to know whether it is the individual evaluation of the individual or the na-tional economic situation or whether it is the objective economic conditions of the nation influenc-ing individual behavior. Additionally, we do not have empirically supported explanations on how objective economic circumstances influence individual's retrospective evaluations.

The aims of the planned paper are fourfold: (1) a simultaneous analysis of the influence of individ-ual sociotropic and egocentric factors as well as contextual characteristics on retrospective voting, (2) an analysis of cross-level interactions between individual and contextual levels. With this two analyses not only the influence of different contexts on the individual voting and participation deci-sion but also the influence of the context on the different individual retrospective evaluations can be measured. (3) I argue that economic characteristics are only one important possibility of govern-ment evaluation. Individual well-being, life satisfaction, or protest behavior on the individual level and different institutional settings on the context level also influence retrospective voting. The purely economic models, therefore, will be extended by sociological and institutional factors. (4) The analyses will be conducted with the multi-level method. In a short part of the paper, the advan-tages of this method compared to purely individual or economic analyses will be discussed.

The analyses are based on the individual side on the 1999/2000 World Value Survey providing us with individual data of 70 to 80 different countries and on the contextual side on different macro-economic and institutional data from the corresponding countries.

1.2   Study of interaction in business and media research

Session chair:
    Josep Bisbe; ESADE, Universitat Ramon Llull, Spain
    Germà Coenders; University of Girona, Spain

This session takes place on:
    Tuesday, July 19; 15:00 to 17:00

Presentations:

- Invited presentation: *Simple, Efficient and Distribution-free Approach to Interaction Effects in Structural Equation Models with Indirect Effects* - Germà Coenders, Joan Manuel Batista-Foguet, Willem Saris (Spain)

Contributed presentations:

- *Interaction effects in Structural Equation Models: the problem of dichotomous variables* - Irene Jonkers (Netherlands)

- *Moderating Effects of Budget Systems and Innovation on Performance: Correcting Measurement Error in Tests of Interaction Models with Small Samples* - Josep Bisbe, Germà Coenders, Willem Saris, Joan Manuel Batista-Foguet (Spain)

- *A Non-Linear Approach to Psychological Well-Being in Adolescence: Major Strenghts and Methodological Issues* - Mònica González, Ferran Casas (Spain)

1.2.1 Simple, Efficient and Distribution-free Approach to Interaction Effects in Structural Equation Models with Indirect Effects

Germà Coenders; University of Girona, Spain

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. Structural equation models with non-linear constraints make it possible to estimate interaction effects. This paper first makes a review of the various alternative approaches, with emphasis on those which estimate the measurement and structural parts of the model together by means of maximum likelihood, namely those of Jaccard and Wan (1995), of Jöreskog and Yang (1996), of Algina and Moulder (2001), of Marsh et al. (2004) and of Schermelleh-Engel et al. (1998).

We try to combine the best features of each approach by eliminating the mean structure, using non-overlapping pairs of indicators and all the constraints but one, which leads to a more easily handled model that is more robust to non-normality and more general as it can accommodate endogenous interactions and thus indirect effects.

Our approach is compared to the alternatives with a Monte Carlo simulation and is found to be equally efficient under normality and less biased under non-normality. An empirical illustration about indirect effects on consumer behaviour is included.

1.2.2 Interaction effects in Structural Equation Models: the problem of dichotomous variables

Irene Jonkers; University of Amsterdam, Netherlands

Abstract. The representation of complex social realities in a testable model often is a difficult task. The possibility to introduce interaction effects to regression models with the use of a product between the variables in question is a great improvement. While this option has been available in OLS regression for some time, the use of product variables in Structural Equation Modelling has been limited, due to some specific complications. Thanks to the accumulative efforts of several scientists, restrictions have been formulated that make the estimation of interaction effects with SEM possible. This is a great achievement, since SEM has a lot of benefits over OLS regression. The logics behind the restrictions, however, are based on metric variables. This is a criterion that is not always met in the social sciences. Many variables are dichotomous. In the worst case nominal variables are represented by a series of dummy variables. This kind of data cause specific problems in the estimation of interaction effects with SEM. In my research into media effects on political attitudes, I have compared the audiences of several newspapers within a country on their attitudes towards the European integration. I was especially interested in several interaction effects, among which political interest and frequency of exposure. The product term that represents the interaction effect is the product of a dichotomous and a metric variable. I encountered three major problems with my analyses:

1. All the respondents that do not read a certain newspaper, and thus get scored with 0, will end up in the same category. For the majority of the newspapers, the number of non-readers exceeds the number of readers. Since the same elements end up in one category, the result is extremely high collinearity between the dummy variable representing the readers of a newspaper and the product term.

2. The value that gets appointed to the group of non-readers is an arbitrary value. This value, however, has a major influence on the form of the relationship between the product term and the dependent variable.

3. The restrictions - as first formulated by Kenny & Judd (1984) and simplified among others by Jöreskog & Yang (1996), Batista-Foguet e.a. (2004) –that should make the estimation of interaction effects with SEM possible, don't apply to the product of a dummy and a metrical variable. The restrictions represent the created dependency between the variance of the error terms and the latent variables, when two variables are multiplied. The variance of a product of a dummy and a metric variable, however, is not equal to the product of the variance of the two variables, since the large group of non-readers will end up in the same category.

My presentation will mainly be focused on identifying these problems. In my research I have chosen some ad hoc solutions, which are not entirely correct. I would like to share my solutions and discuss the problems in the hope to find a better approach.

1.2.3 Moderating Effects of Budget Systems and Innovation on Performance: Correcting Measurement Error in Tests of Interaction Models with Small Samples

Josep Bisbe; ESADE, Universitat Ramon Llull, Spain

Germà Coenders; University of Girona, Spain

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Abstract. In the accounting literature, interaction effects are often assessed by means of ordinary least squares regression and summated rating scales are constructed to reduce measurement error bias. Structural equation models and two-stage least squares regression could be used to completely eliminate this bias, but large samples are required for these procedures to be applicable. Partial Least Squares are appropriate for small samples but do not correct measurement error bias. This paper discusses disattenuated regression as an alternative for correcting measurement error in small samples. Disattenuated regression is illustrated on a model that examines the interaction effect of innovation and style of use of budgets on performance. Disattenuated regression is recommended for future research as a method for correcting the effects of measurement error when the aim of the analysis is parameter estimation or theory testing and sample size is small.

1.2.4 A Non-Linear Approach to Psychological Well-Being in Adolescence: Major Strenghts and Methodological Issues

Mònica González; University of Girona, Spain

Ferran Casas; University of Girona, Spain

Abstract. Psychological well-being in adolescence is an increasing field of study. Deepening in its knowledge during this period of life can be of a lot of help to the designing of more adjusted prevention programs aimed to avoid or reduce the problems adolescents might be experiencing. Complexity theories can be a productive alternative to the important limitations explanations about psychological well-being in adolescence have nowadays.

Answers to a questionnaire have been obtained by a sample of 968 Catalan adolescents from 12 to 16 years old including 29 psychological well-being indicators measuring 8 dimensions related to satisfaction with specific life domains, self-esteem, perceived social support, perception of control and values.

A structural equation modeling approach to complexity that focuses on the non-linearity property has been followed. Given the large number of dimensions, the model has been estimated in two steps. First a confirmatory factor analysis model has been fitted to the 29 indicators and appropriate factor scores have been saved. Then all possible products and squared terms of the factor scores have been computed and have been used as predictors of the dependent variable using an ordered logit model.

The results show that a non-linear model including interaction effects among the 8 dimensions has a higher explanatory power to predict satisfaction with life as a whole, compared to a linear model estimated from those same indicators.

This work must be understood as a first step, basically a methodological one, to the future elaboration of new models of psychological well-being in adolescence to be based on the principles defended by complexity theories.

1.3   Models for the analysis of data from several groups of the same or different populations

Session chair:

    Albert Satorra; University Pompeu Fabra, Spain

This session takes place on:

    Tuesday, July 19; 17:30 to 19:30

Presentations:

- Invited presentation: *Models for Multigroup and Heterogeneous Data* - Albert Satorra (Spain)

Contributed presentations:

- *Stability of household income in European countries in the 1990's* - Nick Longford (United Kingdom), M. G. Pittau (Italy)

- *Mean comparison across countries when factor indicators do not exhibit measurement invariance. Some practical experiences* - Alain De Beuckelaer (Netherlands)

- *Context Heterogeneity in Ongoing Relationships: Country and Market Climate effects in the Relationship Between Customer Behavioral Intentions and antecedents* - Jagdip Singh (United States), Edwin J. Nijssen (Netherlands), Hartmut Holzmueller (Germany), Clara Agustin (Spain)

1.3.1   Models for Multigroup and Heterogeneous Data

Albert Satorra; University Pompeu Fabra, Spain

Abstract.   In the practice of structural equation models, data is often composed of several samples of observations from the same or different populations. We may have data sets from several countries and we are interested whether countries share the values of a particular subset of parameters or want to characterize the differences among them. When there are many countries we may represent them by random effects. Then these questions can be described in terms of the between-country variance matrix. When the groups (countries) are not recorded, the heterogeneity in the amalgamated dataset leads to finite-mixture models. In this talk, we survey the different modeling approaches for data with heterogeneity arising from different perspectives. A connection will be established with SEM for multiple-group, multilevel, and mixture distribution data.

1.3.2   Stability of household income in European countries in the 1990's

Nick Longford; SNTL, Leicester, United Kingdom

M. G. Pittau; University of Rome 'La Sapienza', Italy

Abstract.   This paper explores the patterns of change in the annual household income in the countries of the European Community during the years 1994-1999. The household income is modelled by mixtures of multivariate log-normal distributions, and the mixture components are interpreted as representing one subpopulation with steady increments and others with various levels of volatility. Models are fitted by the EM algorithm in which the group membership is regarded as the missing information. The method is extended to models for a combination of log-normal and categorical variables. An index of income stability is defined for the countries. Throughout, we emphasize graphical summaries of the results.
    Keywords: EM algorithm; Household income; Kernel estimation; Log-normal distribution; Mixture models.

1.3.3   Mean comparison across countries when factor indicators do not exhibit measurement invariance. Some practical experiences

Alain De Beuckelaer; Radboud Universiteit Nijmegen, Netherlands

Abstract.   One of the greatest challenges facing cross-cultural survey researchers is the development of multi-item scales that measure a factor (or
    construct) across multiple cultures (or across multiple countries). To make sure that valid cross-country comparisons based on (estimated) factor means can be made, a thorough assessment of measurement invariance of factor indicators (across countries) is required. Factor mean comparisons are likely to be biased if one or more factor indicators do not exhibit measurement invariance across countries.
    This paper shows how such a 'bias assessment' was made in a recent multi-country job satisfaction survey. Data from an employee opinion survey study collected in 16 countries have been used for this purpose.
    All data have been analysed by means of Mean- And Covariance Analyses (MACS), a special type of Confirmatory Factor Analysis (CFA).
    Keywords: Cross-cultural applicability of scales; Confirmatory Factor Analysis; Mean and Covariance Structure Analysis; Employee survey; Job satisfaction.

1.3.4   Context Heterogeneity in Ongoing Relationships: Country and Market Climate effects in the Relationship Between Customer Behavioral Intentions and antecedents

Jagdip Singh; Case Western Reserve University, United States

Edwin J. Nijssen; Radboud Universiteit Nijmegen, Netherlands

Hartmut Holzmueller; University of Dortmund, Germany

Clara Agustin; University Pompeu Fabra, Spain

Abstract.   Marketing scholars and practitioners interested in relationship marketing have mainly focused on the attitudes and behaviors of consumers and firms' contact personnel in ongoing relational exchanges between individual consumers and firms. Despite of the growing attention on contextual variability in the practitioner literature and the empirical evidence (Jones & Sasser 1995; Zeithaml et al. 1996; Tax et al. 1998), little theoretical work has been initiated to both explain this contextual variability and systematically develop its implications for individual consumer-firm relationships.

In this study, we take a step forward in partialling out the heterogeneity due to context-level effects testing and describing the situational context heterogeneity in the relationship between loyalty, complaining and retaliation intentions and their antecedents in the real estate industry across three different countries–United States, Germany and the Netherlands in a cross-sectional survey-based study. Our approach on the treatment of contextual variability differs from commonly used approaches in the sense that we do not consider contextual heterogeneity as a source of random error that interferes in obtaining generalizable conclusions, neither describing heterogeneity without providing theoretical explanation concerning contextual variability.

Rather, we conceptualize the exchange context in theoretical terms by developing the notion of 'market climate' (Alderson and Halbert, 1968; Hirschman 1970) drawing from different literatures including the exchange paradigm in marketing theory, consumer attitudes toward businesses, and market orientation. Thereafter, we conceptualize the 'market climate' construct, examine its psychometric properties with data from the three countries, and identify its key four dimensions (cloudy, dark, stormy and sunny) using Latent Class Analysis. Then, we provide an empirical test of its validity by exploring the influence of context effects on individual level consumer-firm relationships comparing across the four different market climates found using Structural Equation Modeling. Sensitive to measurement error and other sources of bias, we use latent linear and non-linear constructs, and controlled for common method and response style bias. Moreover, to facilitate substantive inferences and generalizability we use Multigroup Analysis to test for construct equivalence (factorial similarity and factorial equivalence) across countries and model equivalence (structural coefficients equivalence) across the four market climates.

1.4   Multilevel analysis

Session chair:
  Marc Saez; University of Girona, Spain

This session takes place on:
  Wednesday, July 20; 9:00 to 11:00

Presentations:

- Invited presentation: *Multilevel analysis* - Marc Saez (Spain)

Contributed presentations:

- *Comparative Poverty Dynamics* - Marc Callens (Belgium)

- *Multilevel measurement error models for estimation of neighbourhood effects* - Jouni Kuha (United Kingdom)

- *The indirect impact of social stratification determinants for poverty transitions in European countries: a multilevel event history approach* - Leen Vandecasteele (Belgium)

- *Multilevel modeling of survey data and small-area estimation* - Milorad Kovacevic, Rong Huang (Canada)

1.4.1   Multilevel analysis

Marc Saez; University of Girona, Spain

Abstract. It is widely known that in areas such as sociology, medicine, psychology and many others disciplines, data typically exhibit a more or less complex hierarchical structure. A lower level, usually individuals, tends to grouped together in an upper level, say communities, institutions and so on. These grouping effects affect not only those upper but also, and moreover, lower level's behaviour. In this sense, multilevel analysis has now become the standard statistical tool to analyze such data.

   In this introduction, I first comment how multilevel analysis is known to depend on both the particular discipline and on the nature of the dimension, for example temporal or not. Second, I comment the two perspectives multilevel analysis could take, marginal or conditional, depending on the type of inferences one is looking for. Finally, and using applications to different disciplines, I show the different statistical models used depending on: i) the perspective of analysis and; ii) the nature of the dependent variable.

1.4.2   Comparative Poverty Dynamics

Marc Callens; University of Leuven, Belgium

Abstract. We review and discuss the relative merits of three major contextual regression strategies in the context of cross-national poverty dynamics: separate regressions, analysis of covariance and multilevel analysis. Compared with other analysis methods, multi-level modelling proves to be a general and flexible, but complex method. In an application based on individual panel data of the European Community Household Panel linked with a regional time series database, we illustrate the use of multilevel discrete-time recurrent hazard analysis to simultaneously model the impact of life cycle events and structural processes on poverty entry and exit across European Regions. Main findings are that men's poverty dynamics is dominated by employment-related events, while for women demographic events also play a role. Regional structural factors only have a slight or no influence on poverty transitions, but the welfare regime turns out to be highly significant for poverty entry.

1.4.3   Multilevel measurement error models for estimation of neighbourhood effects

Jouni Kuha; London School of Economics, United Kingdom

Abstract. Many attitudes and behaviours may be influenced by the characteristics of the neighbourhood in which a person lives, absorbed through personal observation and conversations with neighbours. Such neighbourhood effects have been reported in various domains, including public health, poverty, employment, fear of crime, educational attainment, voting behaviour, and tolerance between ethnic and religious groups. Increased understanding of these effects is important for a variety of academic and policy debates about individualism, community cohesion and polarization, political mobilization and engagement, and the dynamics of public opinion.

   A typical statistical analysis of neighbourhood effects involves standard multilevel modelling of individual-level survey data, with neighbourhoods as the higher-level clusters of individuals. Specific neighbourhood effects are then operationalised as the effects of corresponding neighbourhood-level explanatory variables, while their interactions with individual-level variables indicate differential sensitivity of individuals to neighbourhood effects. A methodological problem with such modelling is the difficulty of measuring many neighbourhood characteristics. Although some measures can be derived from census and economic data, these sources provide no information on quantities such as average opinions or average levels of crime experienced by individuals in a neighbourhood. An obvious approach is then to use sample averages derived from survey data to represent these variables. This, however, introduces a measurement error problem because sample results are only estimates of the true averages, even when the corresponding individual-level variables are measured with no error. The magnitude of the measurement error, i.e. the sampling error of the sample means, is typically large as social surveys may have few respondents in any particular neighbourhood.

   This talk will describe methods for adjusting for the measurement error in neighbourhood averages in multilevel linear and logistic models. Such methods are needed because the naïve approach of using the sample means instead of the true averages as explanatory variables will yield biased estimates of all of the model parameters. The adjusted estimators are extensions of methods of measurement error modelling for standard generalized linear regression models. The methods will be illustrated by analyses of voting data from the British Election Panel

Studies, and of data on attitudes towards the police from the Policing for London Survey. The results are also used to provide guidelines for the design of studies of neighbourhood effects, especially for efficient choices of sample sizes within and between neighbourhoods

1.4.4   The indirect impact of social stratification determinants for poverty transitions in European countries: a multilevel event history approach

Leen Vandecasteele; University of Leuven, Belgium

Abstract.   In the context of the last decade's poverty research, the life course is often referred to. Temporary poverty spells are studied within a person's biographical experiences. Life course events like divorce, the birth of a child, or the loss of a job, can temporarily influence the chances of poverty entry. Less attention has been focused on the structural context of these life course events. Poverty spells are often seen as belonging to an individual's personal biography. In that view, poverty is principally a temporary phenomenon associated with certain life course events, and less influenced by traditional social characteristics (Leisering & Leibfried, 1999).

In this paper, the focus is on social stratification determinants as mediators in the process of life events leading to poverty entry/exit. Household composition changes and household employment situation changes have found to be predictors of poverty dynamics, but not all persons experiencing the triggering events do consequently experience a movement into or out of poverty. For which social groups leads the experience of a critical life event consequently to a poverty transition? Which inequalities can be found and are these inequalities structured according to traditional social stratification determinants? Previous research has shown that the negative income effects from union dissolution are larger for women than for men in the USA and Germany (Di Prete & McManus, 2000). In this paper, the mediating effects of social class, gender and cultural capital will be investigated by using interaction effects in the random effects model for event data.

The empirical analyses in this paper are based on data from the European Community Household Panel Survey. The technique of multilevel discrete-time event history analysis will be used. An advantage of this technique is that multiple spells of poverty per person can be taken into account. A random component on the level of the individual makes it possible to model individual-specific factors influencing multiple poverty transitions. Advantages and drawbacks of the chosen technique will be described. The countries under study are Germany, Spain, Denmark and the United Kingdom. Differences and similarities between the countries will be linked to the respective welfare regimes.

1.4.5   Multilevel modeling of survey data and small-area estimation

Milorad Kovacevic; Statistics Canada, Canada

Rong Huang; Statistics Canada, Canada

Abstract.   When fitting a multilevel model to survey data one has to account for the sample design, i.e., the unequal probabilities of selection, the clustering in the population, as well as the stratification. The point estimates of the model fixed parameters are affected by the unequal probabilities of selection, the point estimates of the model random parameters are sensitive to the weight scaling, and the standard errors of these estimates, as well as, the tests and the interval estimates depend on the amount of clustering in the data. Several methods for fitting the linear mixed models to survey data suggested in literature are reviewed and compared. These methods include the probability weighted iterative generalized least squares method of Pfefermann et al. (1998) and the weighted estimating equation method of You et al.(2003), then the pseudo-empirical best linear unbiased estimation of You and Rao (2002), pseudo maximum likelihood of Kovacevic and Rai (2003), method of moments of Korn and Graubard (2003), and the modified Henderson's method III of Huang and Hidiroglou (2003). Their application to small area estimation is discussed. The finite population properties of the methods are studied in a small simulation study designed according to the Canadian Work Place and Employee Survey.
References:

Huang, R. and Hidiroglou, M. (2003). Estimators for a mixed linear model on survey data. Proceedings of the Section on Survey Research Methods, 2003 JSM, 1897-1904

Korn, E.L. and Graubard, B.G. (2003) Estimating variance components by using survey data, Journal of Royal Statistical Society, B 65, 175-190.

Kovacevic,M.S. and Rai, S. (2003). A pseudo maximum likelihood approach to multilevel modeling of survey data. Communications in Statistics: Theory and Methods 32, 103-121.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. Journal of Royal Statistical Society B 60: 23-40.

You,Y. and Rao, J.N.K. (2002). A pseudo EBLUP approach to small area estimation using survey weights. The Canadian Journal of Statistics, 30, 431-439

You, Y, Rao, J.N.K. and Kovacevic, M.S. (2003). Estimating fixed effects and variance components in a random intercept model using survey data, Proceedings of Statistics Canada Symposium 2003

1.5   Variance estimation with an emphasis on small area estimation

Session chair:
  Nick Longford; SNTL, Leicester, United Kingdom

This session takes place on:
  Wednesday, July 20; 11:30 to 13:30

Presentations:

- Invited presentation: *On standard errors of model-based small-area estimators* - Nick Longford (United Kingdom)

Contributed presentations:

- *Diagnostics for the use of random effects models in small area estimation* - Li-Chun Zhang (Norway)

- *Linear Mixed Models with Restrictions in Small Area Applications* - A.F. Militino, M.D. Ugarte, Tomas Goicoa (Spain)

- *An Application of Linear Mixed Model with Spatio-Temporal Correlated Random Effects for Small Area Estimation in the Italian Labour Force Survey* - Michele D'Alò, Stefano Falorsi, Monica Russo, Fabrizio Solari (Italy)

- *Small Area Estimation Under Auxiliary Sample Information* - Albert Satorra, Eva Ventura, Alex Costa (Spain)

1.5.1   On standard errors of model-based small-area estimators

Nick Longford; SNTL, Leicester, United Kingdom

Abstract. A key assumption made in model-based small-area estimation is that the areas form a random sample from a (super-) population of areas. Although very constructive, enabling borrowing strength across areas (exploiting their similarity), this assumption is incorrect, because hypothetical replications of the sampling scheme would be based on the same population.

In estimating the mean squared error, this 'incorrectness' is interpreted as a form of averaging over the distribution of the area-level deviations. The consequences of this are evaluated.

I derive an estimator of the mean squared error (MSE) of the empirical Bayes and composite estimator of the local-area mean in the standard small-area setting.

The MSE estimator is a composition of the established estimator based on the conditional expectation of the random deviation associated with the area and a naive estimator of the design-based MSE. Its performance is assessed by simulations. Variants of this MSE estimator are explored and some extensions outlined.

1.5.2   Diagnostics for the use of random effects models in small area estimation

Li-Chun Zhang; Statistics Norway, Norway

Abstract. Small area models are primarily used to make prediction. While it may be difficult to justify the use of a model that does not fit the data well, a better fitting model is not necessarily superior when it comes to prediction. A general issue here is the choice between a fixed effects model and its mixed effects model extension.

That is, given the auxiliary variables of the model, should we allow random effects in addition? The basic idea of our approach to this problem is as follows. When the finite population of interest is fixed, all model-based methods are biased. On the whole, while the estimators derived under a mixed effects model are less biased than those derived under a model that only contains the corresponding fixed effects, the former have larger sampling variances due to heavier estimation burdens. There arises, thus, a trade-off between bias and variance in general, which we attempt to assess by means of the proposed diagnostics.

1.5.3   Linear Mixed Models with Restrictions in Small Area Applications

A.F. Militino; University of Navarra, Spain

M.D. Ugarte; University of Navarra, Spain

Tomas Goicoa; University of Navarra, Spain

Abstract.   Small area techniques are becoming popular because in local and central Governments there is an increasing demand for obtaining precise estimates in domains that were not planned when the study was designed, or their inclusion would produce an increase in the cost of the study. These domains are called small areas and they usually correspond to small geographic areas, such as counties, municipalities or administrative divisions. The term small area can also refer to a speci.c small subpopulation, such as a type of crop, a particular economic activity or a subgroup of people from the same sex, race or other characteristic, within a large geographic domain.Traditionally, the sample sizes are chosen to provide reliable estimates for large regions or aggregates of small areas, but the statistical methods used for large domains can rarely be applied to small ones because the classical direct estimators yield very large standard errors or even they can not be calculated because of the lack of sample in some unplanned domains. When the sample in a given area is very small, a solution to the estimation problem is to use indirect estimators, in particular, model-based estimators, that "borrow information"from related areas, past occasions or both. Rao (2003) give an account of di.erent model-based methods to accomplish small area estimation tasks. One of the most popular methods has been the use of linear mixed models as a basis to obtain empirical best linear unbiased predictors (EBLUP) of population means and totals (see, for example, Battese et al., 1988). A drawback ofthese methods is that the estimates obtained for eachsmallarea do notaddupto the estimate provided by statistical o.ces at some level of aggregation with appropriate estimators for larger areas, and then, calibration weights are needed. To avoid the calibration weights, we propose to impose some restrictions in the estimation process of the model parameters to achieve the desired consistency in the results. The obtained predictors are robust in the sense that they perform well on the average if the model does not hold (Pfe.ermann

and Barnard, 1991). The mean squarederrors ofthe predictors are calculated byusinga bootstrap procedure.The results are illustrated with real data from the Basque Country Business Survey. In this example, the estimates at county level (small areas) are obtained from a linear mixed model with restrictions and the estimates at province level (the small areas aggregation) are obtained from a synthetic estimator.

1.5.4 An Application of Linear Mixed Model with Spatio-Temporal Correlated Random Effects for Small Area Estimation in the Italian Labour Force Survey

Michele D'Alò; ISTAT, Italy

Stefano Falorsi; ISTAT, Italy

Monica Russo; ISTAT, Italy

Fabrizio Solari; ISTAT, Italy

Abstract.   Usually, large scale surveys conducted by the Italian National Institute of Statistics (Istat) produce estimates not only for the target population but also for some subpopulations, defining unplanned domains usually called "small areas". Therefore, often, large scale surveys are repeated over the time.

In the Labour Force Survey, for instance, estimates are produced for provinces, NUTS3, and for Local Labour Market Areas (LLMAs), which are defined as an aggregation of municipality (NUTS5) that can be considered as NUTS4. In particular, in each LLMA, estimation of the parameters of interest are computed by means of composite estimator.

Recently Istat has carried out an empirical studies to evaluate the performances of standard model assisted and model based estimators. Special efforts have been addressed to the assessment of model based methods that make use of spatial autocorrelation structures. This estimator, based on a linear mixed model with spatially correlated random effects, has shown good results in terms of bias and mean squared error.

To improve the quality of the estimates in planned and unplanned domains it is necessary to take into account time effects in the estimation process. More precisely we introduced time random effects in the linear mixed model.

Since estimation of parameters in linear mixed models is highly computer intensive, the parameters estimates formulas have been reformulated by means of suitable algorithms.

1.5.5 Small Area Estimation Under Auxiliary Sample Information

Albert Satorra; University Pompeu Fabra, Spain

Eva Ventura; University Pompeu Fabra, Spain

Alex Costa; Catalan Institute of Statistics (Idescat), Spain

Abstract.   We consider small area estimation when the sample information about the target variable is supplemented by another survey of a variable correlated with the target variable. An application to the EPA (Spanish Labour Force Survey) is presented in which data from EPA is complemented by a survey conducted by CIS (Spanish Centre for Social Research). Inferences are desired about the variable recorded by EPA (ILO unemployment), and CIS records a variable with a very similar definition. We compare several composite small area estimators that combine the information from the two sources. Monte Carlo simulation based on an artificial population shows that the MSE is reduced substantially by integrating the auxiliary data in the estimation. The reduction is for some areas equivalent to doubling the sub-sample size.

1.6   Comparative research on religion

Session chair:
   Heiner Meulemann; University of Köln, Germany

This session takes place on:
   Wednesday, July 20; 15:00 to 17:00

Presentations:

- Invited presentation: *Altruism, Religiosity, and Involvement in Different Kinds of Organizations in Europe* - Heiner Meulemann (Germany)

Contributed presentations:

- *Religious Divide in Europe: measurements and opportunities for analysis in the ESS round 1* - Jaak Billiet (Belgium)

- *Religion in Europe: Secularization versus Rational Choice* - Loek Halman, Veerle Draulans (Netherlands)

- *Religion and Life-Satisfaction in a Comparative Perspective* - Wolfgang Jagodzinski (Germany)

1.6.1 Altruism, Religiosity, and Involvement in Different Kinds of Organizations in Europe

Heiner Meulemann; University of Köln, Germany

Abstract. Organizations of the Civil Society will be classified into interest organizations which predominantly serve private interests (as e.g. sports clubs), issue organizations which predominantly serve public issues (as e.g. environmental organizations), and religious organizations which serve at the same time interests in the other world and issues in this world. Therefore, involvement in interest organizations should be more strongly determined by egoistic attitudes, and involvement in issue and religious organizations should be more strongly determined by altruistic attitudes of which religiosity is the most important one. This is examined with the data of the European Social Survey 2002.

On the level of persons, it is examined if involvement in all three kinds of organizations will depend equally on resources and on attitudes of persons which are important for involvement in all kinds of organizations (as e.g. the subjective norm of involvement), while involvement in interest organizations will be less strongly depend on altruism and, in particular, religiosity than in issue and in religious organizations. Empirically, interest and issue organizations are very similar as for their determinants while religious organizations differ from both. While resources are useful for involvement the two forms of secular organizations, they are irrelevant for involvement in religious organizations; and while denominational membership and religiosity have no or negative impacts on involvement in secular organizations, they have a very strong positive impact on involvement in religious organizations. As for their individual level determinants, then, organizations have to be differentiated –yet not according to interest and issue, but to secular and religious ones.

On the level of countries, it is examined if involvement in all three kinds of organizations depends more strongly on the social order or on the opportunity structure which both pertain to politics, economics, and culture. Intercept models of linear hierarchical regressions show that the Protestant tradition of a country increases involvement in every kind of organization and social differentiation increases involvement in secular (interest and issue) organizations.

1.6.2 Religious Divide in Europe: measurements and opportunities for analysis in the ESS round 1

Jaak Billiet; University of Leuven, Belgium

Abstract.   I start with an overview of two possible measurements in ESS, a categorical variable 'religious involvement' and a metric variant 'religious commitment' (latent variable with four indicators). The factorial invariance of the latter is tested in a multi-group model with Lisrel_8. The results are discussed. Can we compare the latent means of religious commitment over 22 countries and order these? In the second part, I explore what kind of analysis are possible with the ESS data. There are several possibilities depending of the choice of religious commitment/involvement as a dependent or an independent variable. The rational choice versus the secularisation theory are tested with in a hierarchical model with both individual level and country level variables.

1.6.3 Religion in Europe: Secularization versus Rational Choice

Loek Halman; University of Tilburg, Netherlands

Veerle Draulans; University of Tilburg, Netherlands

Abstract.   A large body of literature has developed yielding evidence that religion in general and churches and church leaders in particular have lost their once dominant and prominent position in contemporary Europe. Evidence is often found in the declining levels of church attendance. Whether Europe should also be qualified as secularized in terms of religious beliefs remains unanswered. We investigate to which degree European people are indeed secular and we focus not only on religious practices, but also on beliefs. We argue that trajectories of religious change occur all over Europe, however not at similar speed. We formulate some hypotheses on the differences in the degree to which individuals and societies are secularized. Data from the recent European Values Study surveys are used to empirically test a number of hypotheses on the varieties in religious beliefs and practices. The analyses provide evidence in favor of secularization theories and contradicting rational choice theories. In Europe, religious pluralism does not produce higher levels, but lower levels of religiosity. The analyses reveal also that religious denomination is an important factor and that cultural and socio-economic heritages should be considered to explain and understand the patchwork pattern in levels of religiosity and religious participation in contemporary Europe.

1.6.4   Religion and Life-Satisfaction in a Comparative Perspective

Wolfgang Jagodzinski; Central Archive, Germany

Abstract.   Even though the relationship between religiosity and life-satisfaction has been investigated in a number of papers, so far no clear picture has emerged. Theoretically, there are compelling arguments in favour of a positive relationship. Empirically, we find fairly happy and fairly unhappy religious societies in Western Europe, and happy and unhappy secular societies as well. While there seems to be no macro-level relationship in Western Europe, the pattern on the micro-level according to the available studies is more consistent The relationship in Asian countries is rarely examined. This paper attempts estimate a multi-level model for the impact of religiosity on life-satisfaction in Asian countries.

1.7   Measuring basic human values

Session chair:
    Shalom Schwartz; Hebrew University of Jeruzalem, Israel

This session takes place on:
    Thursday, July 21; 9:00 to 11:00
    Thursday, July 21; 11:30 to 13:30
    Thursday, July 21; 15:00 to 17:00

Presentations:

- Invited presentation: *The Structure And Implications Of Individuals' Value Systems* - Shalom Schwartz (Israel)

Contributed presentations:

- *Bringing Values Back In: A Multiple Group Comparison with 20 Countries Using the European Social Survey 2003* - Eldad Davidov, Peter Schmidt (Germany), Shalom Schwartz (Israel)

- *European Value Map: Generations And Countries* - Hans Bay (Denmark)

- *Structural Equivalence Of The Values Domain Across Cultures: Separating Sampling Fluctuations From Systematic, Meaningful Variation* - Johnny Fontaine (Belgium), Ype H. Poortinga (Netherlands), Luc Delbeke (Belgium), Shalom Schwartz (Israel)

- *The Measurement Of Value Preferences By Paired Comparisons* - Michaela Brocke, Wolfgang Bilsky (Germany)

- *A Two Dimensional Model For Presenting Values Measured With Schwartz' 21 Item Portrait Values Questionnaire* - Markku Verkasalo, Jari Lipsanen (Finland)

- *Value Orientations in Europe. One Kind of European Identity?* - Florian Pichler (Austria)

- *Organizing Diverse Sets Of Data With The Schwartz' Value Circle* - Micha Strack (Germany)

- *Social values: salience and consensus in 10 European countries* - Alice Ramos, Jorge Vala, Henrique Duarte, Diniz Lopes (Portugal)

- *Attitudes and values of the Europeans: a gender perspective* - Anália Torres, Rui Brites, Rita Mendes, Tiago Lapa (Portugal)

- *Explorative And Confirmatory Factor Analysis In The Study Of Values* - Andu Rämmer (Estonia)

1.7.1   The Structure And Implications Of Individuals' Value Systems

Shalom Schwartz; Hebrew University of Jeruzalem, Israel

Abstract.   After clarifying the nature of basic human values, I will examine the bases for the structure of individuals' value systems. I will ground the value system in the everyday experience of choice among action alternatives. I will then present dimensions which organize the theorized and near-universally observed structure of the value system. These include the extent to which a value is a response to anxiety or is anxiety-free, whether its goal is to prevent loss or promote gain, whether it is concerned with protecting the self against threat or self-expansion and growth, and whether it regulates how we express our personal interests and characteristics or how we relate socially to others and affect them. The structure may be found universally because it expresses four innate drives in the form of goals. Some evidence suggests that the motivational structure of the value system can serve as a general theory of human motivation that applies to needs, personality, and behavior. I will briefly present the two main methods for measuring individual value priorities and the rationales that underlay their development. Finally, I will note a sampling of the significant behaviors that individual differences in value priorities have predicted.

1.7.2   Bringing Values Back In: A Multiple Group Comparison with 20 Countries Using the European Social Survey 2003

Eldad Davidov; University of Giessen, Germany

Peter Schmidt; University of Giessen, Germany

Shalom Schwartz; Hebrew University of Jeruzalem, Israel

Abstract.   Values are important because they can function as a black box and intermediate between socio demographic characteristics and attitudes and behaviour. In this paper we analyse the newly developed instrument to measure values in population surveys (Portrait Value questionnaire (PVQ)). For this purpose we have used a multi-sample confirmatory factor analysis to test the measurement properties in 20 European countries. The data we use are from the European Social Survey 2003. We tested for configural invariance, measurement invariance and scalar invariance. Only seven of the ten values underlying the 21 items as postulated by Schwartz were verified by the configural invariance model of multi-group confirmatory factor analysis. We accepted metric invariance. Factor variances, factor covariances and measurement errors vary considerably over the countries, but groups of countries with higher similarities in the value means can be identified. Future research will address the causal structure of these values on attitudes and behaviour and of socio demographic characteristics on these values.

1.7.3   European Value Map: Generations And Countries

Hans Bay; Danish National Institute of Social Research (SFI), Denmark

Abstract. The European Social Survey (the ESS) is a new, academic-driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. One central aim of the ESS is to develop and conduct a systematic study of changing values, attitudes, attributes and behaviour patterns within European policies. Academically driven but designed to feed into key European policy debates, the ESS will try to measure and explain how people's social values, cultural norms and behaviour patterns are distributed, the way in which they differ within and between nations, and the direction and speed at which they are changing. A two-dimensional map (European Value Map) is based on 21 questions. The questions (named 21-item Basic Human Values Scale, developed by S. Schwartz) have been tested internationally (Schwartz, 1992). The two dimensions in the map are: Individualism and Social Conformity.

These two dimensions can be found in all the countries, when the countries are analysed separately. It is proven that the two dimensions are identical. The two dimensions are therefore based on the total sampling. The total sampling covers 20 countries and 35,608 respondents.

1.7.4   Structural Equivalence Of The Values Domain Across Cultures: Separating Sampling Fluctuations From Systematic, Meaningful Variation

Johnny Fontaine; University of Gent, Belgium

Ype H. Poortinga; University of Tilburg, Netherlands

Luc Delbeke; University of Leuven, Belgium

Shalom Schwartz; Hebrew University of Jeruzalem, Israel

Abstract.   We examine the value domain, as measured by the Schwartz Value Survey, assessing equivalence of its internal structure. Data come from 38 countries, each represented by a student and a teacher sample. We seek to disentangle problems with the fit of the theorized value model from a lack of equivalence between cultural groups We then disentangle the impact of random sampling fluctuations from genuine structural deviations. We found that (1) the Schwartz value theory represents the consensus value structure across samples very well, (2) sampling fluctuation causes deviations from this average structure in specific samples, (3) the deviations from the average structure also reflect genuine, systematic structural variation, (4) the most striking genuine variation is the stronger opposition between protection-oriented values and growth-oriented values in countries at a higher level of development.

1.7.5   The Measurement Of Value Preferences By Paired Comparisons

Michaela Brocke; Westfaelische Wilhelms-Universitaet, Germany

Wolfgang Bilsky; Westfaelische Wilhelms-Universitaet, Germany

Abstract.   The importance of values as guiding principles in one's life has usually been investigated by asking subjects to rank order value items or to use Likert items for assessing value preferences. Our study used an alternative approach based on a paired comparison technique. The main purposes of this study are (1) to validate this approach by comparing results from paired comparisons with those obtained from a Likert-type value survey, and (2) to compare value structures resulting from an MDS of paired comparisons with those known from cross-cultural value research. All instruments are administered online. Subjects receive a total of 45 graded paired comparison tasks. On each trial, two of the ten values proposed by Schwartz (1992) were presented. The ten values were operationalized by using the value definitions from the Schwartz Value Survey (SVS). Subjects were asked to indicate the degree to which one value is more important for them than the other. To compare results from paired comparisons with those of a Likert scale approach, subjects were also asked to complete an online version of Schwartz' Portrait Values Questionnaire (PVQ). To control for order effects, the sequence of the paired comparisons and the PVQ was counterbalanced between subjects. Results from a pilot study (N=60) are reported and methodological differences between the paired comparison method and Likert scale ratings are discussed.

1.7.6   A Two Dımensıonal Model For Presentıng Values Measured Wıth Schwartz' 21 Item Portraıt Values Questıonnaıre

Markku Verkasalo; University of Helsinki, Finland

Jari Lipsanen; University of Helsinki, Finland

Abstract.   We will demonstrate a method for presenting basic individual values on the two main dimensions of Schwartz' theory of individual values; Conservation (versus Openness to Change) and Self-Transcendence (versus Self-Enhancement). The application of this method will be illustrated by examining the relations of value priorities to background variables. The data are from representative national samples in 19 countries from the European Social Survey in 2002-2003. 34,073 respondents completed the 21-item version of Schwartz' Portrait Values Questionnaire (PVQ). First, we split the data into two samples. We then generate an orthogonal three factor model from the 21 PVQ value items, separately in each sample. Second, we rotate the factor model graphically, so that the second factor represents the Conservation dimension and the third factor the Self-Transcendence dimension. We evaluate the equivalence of the two factor models. After verifying that the models are sufficiently similar and represent Schwartz' two dimensional model, we combine the data from the two samples. We then compute factor

scores representing all three factors. The factor score obtained with the factor coefficients from the first factor correlates .97 with the mean response to the 21 PVQ items. Thus, the first factor represents response style. We will present relations of scores on the two substantive factors, representing the basic dimensions of values, with age, gender, and schooling, as well as differences between the 19 countries.

1.7.7   Value Orientations in Europe. One Kind of European Identity?

Florian Pichler; Institute for Advanced Studies, Austria

Abstract. It is argued that dominant values and their relation to each other partly shape a society and thus determine its identity to a respectable degree. In my attempt to capture European iden-tity, I analyse value orientations designed by Shalom H. Schwartz and take data from the ESS 2003 (Jowell, NSD). The analysis of values leads towards numerous other questions:

1. Are values contested within and across European societies, i.e. can meaningful variances be observed concerning value orientations?

2. To what degree do people differentiate commitment to particular values (indi-vidually), or: How many answer categories are truly necessary to decide whether peo-ple identify/do not identify with a particular value or not?

3. What do suggested (Schwartz) corrections of data, e.g. centralization, mean as regards content and what do those imply for statistical analysis and interpretation of value orientations?

In the paper I address these and other methodological aspects of measuring Basic Human Values by comparing statistical results with various computed data. I stick to raw data from the ESS, apply the suggested centralizations by Schwartz, dichotomise value orientations in alternative ways and apply a wide set of statistical methods in national and comparative in-ternational perspectives. Methods applied are: preliminary basic univariate and bi-variate de-scription of value orientations (correlations), cluster analysis and exploratory and confirma-tory factor analysis.

The main conclusions are: survey designers shall rethink the necessity of multi-answer cate-gories and thus emphasize/support the development/application of statistical methods for the analysis of categorical data as some questions are easier/more meaningful to answer with a small number of categories. Social research shall disasso-ciate from the unquestioned use of "metric"variables. Data corrections and manipulations should only be applied if interpreta-tion allows for it.

1.7.8   Organizing Diverse Sets Of Data With The Schwartz' Value Circle

Micha Strack; University of Goettingen, Germany

Abstract.   This talk demonstrates the integrative power of the 1992 Schwartz' Value Circle (SVC). Convergent models (e.g., Parsons' functions of social systems or the Competing Values Model for organizational culture and leadership roles; Quinn & Cameron 1999, Quinn 1988) show the SVC validity for the semantics of values. Even Inglehardt's WVS surface seems to fit well. Empirical research reveals that the SVC can be identified in a variety of value scales given that items are positive and entail sufficient representative content diversity (even a six-item Allport-Vernon scale can be turned into an SVC). I reanalyze data from German surveys (ALLBUS 1998, 2002; Shell Youth Study 2002; GIP 1999) which included different value scales. Ordinary factor analysis of ipsative data (eliminating acquiescence) positively identifies the SVC. The resulting axes of the circle allow correlating third constructs such as age, status, personality, attitudes or behavior, as well as locating individuals or group means (e.g., voting preferences of Ss from Western and Eastern Germany); both statistics are visualized in illustrative plots. Joint inspection of these pictures can be used to establish consistent links between regions of the SVC and social milieus.

1.7.9   Social values: salience and consensus in 10 European countries

Alice Ramos; University of Lisboa - ICS, Portugal

Jorge Vala; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Henrique Duarte; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Diniz Lopes; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Abstract.   In this presentation we analyse social values salience and consensus in Europe and whether they are predicted by individual and/or contextual variables. The analysis is based on data collected in the European Social Survey 2002. We analyse the 10 basic values structure proposed by Schwartz, which can be summarized with two orthogonal dimensions: self-enhancement vs. self-transcendence (opposing power and achievement values to universalism and benevolence values) and openness to change vs. conservatism (opposing self-direction and stimulation values to security, conformity and tradition). The basic values structure proposed by Schwartz was found in several studies (e.g. Schwartz, 1992, 1994, 2003; Schwartz & Sagiv, 1995) and its measurement equivalence validated in 21 countries (e.g. Spini, 2003). The questions addressed are: a) Are there any differences in the salience of the 10 values across Europe?; b) What is the degree of value homogeneity across countries; c) What are the individual and contextual predictors of those values?

Multilevel analysis will be performed in order to analyse the predictors of basic values salience and consensus. GDPpp will be used as contextual predictor.

1.7.10 Attitudes and values of the Europeans: a gender perspective

Anália Torres; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rui Brites; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Rita Mendes; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Tiago Lapa; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Abstract.   This paper is based on the data provided by the "European Social Survey", a survey that aims to measure and interpret changes occurring, along the time, in the attitudes, perceptions and social behaviour of the Europeans, and their interaction with the changes in the economic, social and political spheres. The data is from 2002, the first wave of enquiries (Round 1), and is available for 21 countries .

Although it is known that the differences within sexes are much more significant than the differences between sexes, the tendency is to overstress the later ones. In this paper we analyse, in a transversal perspective centred on gender, all the answers of the questionnaire, addressing some of the main issues of the survey like the exposition to media, politics, the subjective perception of well-being and security and human values and trying to evaluate differences and similarities between men and women. What will prevail? Differences or similarities?

1.7.11 Explorative And Confirmatory Factor Analysis In The Study Of Values

Andu Rämmer; University of Tartu, Estonia

Abstract.   Empirical analysis of beliefs underwent both important conceptual and methodological shifts in the past century. Researchers' focus has been on the concept of values in the last decades. One of the key questions in relevant empirical studies concerns to the relation between the measured items. Different data reduction techniques have been applied to the detection of latent structures. In the presentation, two different approaches, explorative and confirmatory factor analyses will be discussed. While the aim of principal axis factoring is to detect latent structures then CFA allows to test supposed factor structure.

On the basis of two examples the structure of Estonian school-leavers' work orientations will be discussed. Three work orientations that were detected in previous Estonian youth studies, namely self-expression, social recognition and career, emerged also among post-Soviet youth. In total, the data of 8133 individuals were analysed.

Keywords: Opinions, Value Orientations, Principal Axis Factoring, Varimax Rotation, Confirmatory Factor Analysis.

1.8   The added value of worldwide surveys: methodological aspects in the cross cultural study of values

Session chair:
   Ronald Inglehart; University of Michigan, United States

This session takes place on:
   Thursday, July 21; 17:30 to 19:30

Presentations:

- Invited presentation: *Introduction* - Ronald Inglehart (United States)

Contributed presentations:

- *Explanatory variables in international comparative research: A comparison of background socio-demographic variables in four major international surveys (WVS-EVS, ESS, ISSP, CSES)* - Juan Diez-Nicolas (Spain)

- *Misinterpretations of the ecological fallacy: Evidence from the World Values Survey* - Chris Welzel (Germany)

- *Different surveys, similar results? A comparative analysis of cross-national data on religious involvement, life satisfaction, protest behavior, and social trust* - Thorleif Pettersson (Sweden)

- *Measuring values in Islamic publics* - Yilmaz Esmer (Turkey)

1.8.1   The added value of worldwide surveys

Ronald Inglehart; University of Michigan, United States

Abstract.   The emergence of large-N cross-national survey projects such as the Values Surveys, the International Social Survey Program, the European Social Survey, the Comparative Study of Electoral Systems and the various Barometers make it possible to test the impact of individual-level beliefs and values on societal-level phenomena such as democratic institutions.

Survey research has always assumed that individual-level beliefs matter, and in the case of electoral research the linkage is relatively obvious: the party supported by the largest number of individuals usually wins the election. But with other important variables, such as democratic institutions, the assumed linkage has been based on an act of faith: the ground-breaking Civic Culture study measured individual-level beliefs in five countries and found differences between the orientations of the publics of two stable democracies (Britain and the U.S.) and those of two recently Fascist regimes (West Germany and Italy) plus Mexico, and concluded that these attitudinal differences explained why Britain and the U.S. were stable democracies, while the other three countries were not (as of 1959). Numerous critics found this conclusion unconvincing. For example, it was perfectly possible that the characteristics shared by the two English-speaking countries simply reflected their joint Anglo-Saxon heritage and had no causal connection with democracy. With only five cases, it was impossible to control for this and various other possible factors or to get statistically significant results.

The emergence of large-N studies makes it possible to carry out much more conclusive analyses of the linkages between individual-level attitudes and societal-level phenomena-and the results indicate that mass attitudes have a major impact on a wide range of important societal variables, ranging from democratic institutions to good governance versus corruption, human fertility rates, and the percentage of women in parliament or other indicators of societal-level gender equality such as the UN Gender Empowerment Measure.

1.8.2   Explanatory variables in international comparative research: A comparison of background socio-demographic variables in four major international surveys (WVS-EVS, ESS, ISSP, CSES)

Juan Diez-Nicolas; Universidad Complutense de Madrid, Spain

Abstract.   The paper compares some of the more common socio-demographic variables that are used to explain attitudes and/or behaviors in European countries (defined as the 25 members of the EU) in four major international surveys: WVS-EVS, ESS, ISSP and CSES.

The socio-demographic variables included in this analysis are: sex, age, education, occupation and income.

For each country a comparison is made across the four international surveys on how each of the five variables has been measured (what instrument was used, i.e., question and response categories), and how similar or different results were.

Results derived from surveys are also compared with national statistics.

The analysis focuses on those cases were significant discrepancies in results are found, both when comparing results from a survey with those of another survey for the same country and with national statistics. But a comparison is also made between the measurement instrument used in different surveys within the same country, and between the measurement instrument used in different countries within the same survey.

In the discussion of the findings some proposals for improving comparison of results from different surveys are presented.

1.8.3   Misinterpretations of the ecological fallacy: Evidence from the World Values Survey

Chris Welzel; Bremen International University, Germany

Abstract.   Scholars have often referred to the ecological fallacy, arguing that aggregate-level relations that are not reflected at the individual-level within each aggregate unit are spurious. Based on evidence from the WVS, the presentation gives examples demonstrating that the above argument is a misunderstanding of the ecological fallacy problem. Since this misunderstanding is rather widespread, it leads to confusion in the notion of the relationship between political culture and democracy. Also, the presentation shows that survey researchers are well adviced to be more sensitive to the individualistic fallacy, not only the ecological one.

1.8.4 Different surveys, similar results? A comparative analysis of cross-national data on religious involvement, life satisfaction, protest behavior, and social trust

Thorleif Pettersson; University of Uppsala, Sweden

Abstract.   The paper reports results from both descriptive and analytical comparative analyses of ESS and EVS/WVS data for 19 European countries. Based on theories from the sociology of religion, four social orientations have been selected for analysis. These are religious involvement, life satisfaction, protest behavior, and horizontal social trust. Given the differences between the two surveys with regard to sampling procedures, questionnaire design, field work, etc, the paper concludes that the results from the two surveys are surprisingly similar.

1.8.5 Measuring values in Islamic publics

Yilmaz Esmer; Bogazici University, Turkey

Abstract.   Repeated surveys have consistently demonstrated that Islamic societies exhibit the highest levels of religiosity in the world both with respect to faith and practice. To give but one example, the 'importance of God' scale that has been asked since 1981 as part of both the European and World Values surveys attains its highest means within Islamic
populations.

On the other hand, certain important differences between Islamic and non-Islamic populations can result in misleading interpretations of the data as well as serious validity problems. At the most basic level, the questionnaires commonly translate 'church' as 'mosque' for Muslims. Yet, the church and the mosque refer to fundamentally different concepts. And data on church or mosque attendance, as the case might be, are not strictly comparable.

This paper examines some of the basic issues involved in conducting value surveys in Islamic societies using examples from past studies and suggests ways of dealing with these problems.

1.9   Cross-Cultural Comparability of Background Variables

Session chair:
    Christof Wolf; Centre for Survey Research and Methodology (ZUMA), Germany

This session takes place on:
    Friday, July 22; 9:00 to 11:00

Presentations:

- Invited presentation: *Methodological Considerations of the Measurement of Religiosity in Cross-Cultural Surveys* - Christof Wolf (Germany)

Contributed presentations:

- *Harmonisation of Survey Data in the International Social Survey Programme (ISSP)* - Evi Scholz (Germany)

- *Methodological Discussion of the Income Measure in the European Social Survey* - Uwe Warner (Luxembourg)

- *How to Measure Education in Cross-National Comparison* - Juergen H.P. Hoffmeyer-Zlotnik (Germany)

- *Evaluation of Behavioural Indicators after Data Collection: Measuring Political Participation of Young People in Europe* - Vlasta Zucha, Brigitte Salfinger (Austria)

1.9.1   Methodological Considerations of the Measurement of Religiosity in Cross-Cultural Surveys

Christof Wolf; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   Although religious membership and, to a lesser extend, religiosity traditionally are thought to belong to the central core of background variables the problem of their measurement in cross-national surveys has not attracted much attention. The European Social Survey (ESS) offers a new opportunity to study this problem.

This survey contains a number of indicators related to religion. In addition to current or former religious affiliation the ESS data contain three measures of religiosity each representing a distinct area of religiousness: public religious behavior, i.e. church attendance; private religious behavior, i.e. prayer; and intensity of belief, i.e. subjective religiosity.

Using these data the paper addresses two sets of questions: Firstly, is the relationship between the three indicators stable across European countries? If there are differences, can they be explained by the religious composition of the countries or related country differences? Secondly, how do these indicators vary across denominations? And is the denomination-specific religious profile constant across European countries

1.9.2   Harmonisation of Survey Data in the International Social Survey Programme (ISSP)

Evi Scholz; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   For cross-cultural surveys, the comparability of information collected in different cultural contexts is the crucial point. To reach comparable information in cross-cultural surveys, different harmonisation strategies can be followed either by input harmonisation, ex-post harmonisation or, alternatively, using elements from both. The first strategy, input harmonisation, starts with discussion and consideration of questions, question wording, and coding in advance. In this approach, cross-cultural surveys usually create a source questionnaire in a well-known language and then ask for a translation into country-specific linguistic equivalent field questionnaires.

The other route to harmonisation is based on a given concept; variables are measured by different indicators asked country specifically. Ex-post harmonisation by regrouping, recoding or reconstructing the collected data aims to result in conceptual equivalence.

This paper first discusses the pros and cons of harmonisation approaches and then presents details of harmonisation strategies in the International Social Survey Programme (ISSP). The ISSP applies different harmonisation strategies: input harmonisation for the so-called substantive questions and a mixed mode harmonisation for background variables. Finally the paper demonstrates how one of the ISSP background variables, educational degree, is harmonised in selected countries.

1.9.3   Methodological Discussion of the Income Measure in the European Social Survey

Uwe Warner; CEPS/INSTEAD, Luxembourg

Abstract.   During the last decade, the number of cross-national and cross-cultural empirical research has increased; at the same time the need for comparative survey data also enlarged.

Four strategies of making data comparable are common:

1. the output harmonization - after the data collection by using different survey instruments in different nations, the comparative data are created;

2. the target harmonization - after the data collection nation by nation, a comparative set of indicators based on a common set of rules and definitions is established;

3. the input harmonization - by using common fieldwork instruments and survey regulations comparative data are obtained;

4. the ad hoc harmonization –by the data users according to the needs of their individual comparative research.

Looking on the question the total net household income, we discus the advantage and weakness of an input harmonized social survey. We demonstrate the impact of the national social, economic and legal particularities on the answering behavior of the surveyed respondent by comparing across countries the interview outcomes from the European Social Survey (ESS) and the European Community Household Panel (ECHP). ESS used a measurement of the total net household income interviewing only one randomly selected household member. ECHP surveyed all persons living in a sampled household and asked all income details and components of the respondents and the household. In this paper we use ECHP as a reference showing the most accurate method to measure income, and compare this with the interview results of ESS.

For comparative social surveys we propose a set of questions on income that takes into account the national circumstances of the respondent. We get comparable data across countries reflecting the national tax systems, the particular practices in the earning structures and the national habits in summing up the different income components. We expect that such a new fieldwork instrument integrated into cross-national surveys may increase the analytical power of the comparative socio-demographic variable "total net household income".

1.9.4   How to Measure Education in Cross-National Comparison

Juergen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   Education is the basic variable constituting social inequality. Therefore education is one of the most important socio-economic variables in national as well as cross-national survey research.

Comparing different national educational systems one has first to analyze these educational systems to find out structural similarities and/or equivalences in different national educational systems . The presentation will demonstrate the process from first, concept to second, national structure and final to harmonized categorical system. Different instruments as CASMIN Educational Classification and Hoffmeyer-Zlotnik's Index of Education, as well as ISCED97 (International Standard Classification of Education) will be discussed. ISCED97 is offered by the European Social Survey to users of the international data set.

The quality of these three indices and the index "years of schooling"will be analyzed using the European Social Survey data set. Therefore national categories of general and vocational education, as far as these variables are necessary for constructing the above mentioned indices, are combined. By transforming national measurements to harmonized indices rules for index construction will be documented. In a next step a comparison of the different harmonized instruments will be done. The goal is to find out which index is most adequate with respect to index construction and in predicting social status

1.9.5   Evaluation of Behavioural Indicators after Data Collection: Measuring Political Participation of Young People in Europe

Vlasta Zucha; Institute for Social Research and Analysis - SORA, Austria

Brigitte Salfinger; Institute for Social Research and Analysis - SORA, Austria

Abstract.   Only few comparative studies have analysed the political situation of young people in Europe so far. The general objective of EUYOUPART (Political Participation of Young People in Europe), which is a project within the 5th framework programme of the European Commission, is to develop an instrument for measuring political participation of young people (aged 15-25 years) in Europe. The project started in February 2003 and covers a wide range of countries –Austria, Estonia, Finland, France, Germany, Italy, Slovakia and the UK. Its aim is to contribute to the identification of relevant and valid indicators for the involvement of young people in politics in different European countries.

The methodological challenge of EUYOUPART was to try to establish equivalence between eight European countries and to test comparability in different phases of the project. Especially in the last phase of the project, the analyses focussed on testing the cross-national comparability of the main indicators and on integrating explanations for the lack of comparability within country-specific contexts.

Many forms of equivalence referring to comparability of the measurement instrument are described in literature. Within the framework of statistical analysis of the recently finished main survey, the focus lies on functional equivalence as basic form of comparability (e.g. van Deth 1998). This means "that concepts should be related to other concepts in other settings more or less in the same way". The present paper shows how comparability of behavioural indicators was tested after data collection, the results of this evaluation and the attempt to find explanations for the lack of cross-national comparability.

There is a variety of methods for testing equivalence –different statistical methods which rely on the structure of variables and dimensions can be applied. The most commonly applied methods are Exploratory Factor Analysis or Multidimensional Scaling. In the context of EUYOUPART, the indicators measuring political participation consist mainly of behavioural variables, which dimensionality and structure was tested. This paper focuses on one important item battery of behavioural variables covering different conventional and unconventional forms of political participation.

In general, the research question aims at functional equivalence in cross-national research. It examines whether the structure of forms of political participation is comparable for the analysed countries. The method applied to the behavioural variables to identify similar groups of participation was Hierarchical Cluster Analysis, because it seems the most adequate method to deal with behavioural questions.

The testing of comparability of behavioural indicators with Hierarchical Cluster Analysis seems to be the most appropriate way to deal with this type of variables and to identify functional equivalence in the final phase of the project. The objective of this process is to (1) evaluate the measurement instrument, (2) identify possible reasons for the lack of comparability and (3) make recommendations on the usage of the final measurement instrument. Although testing behavioural variables after data collection is difficult, a Hierarchical Cluster Analysis might support the detection of non-comparable indicators. In general, it seems helpful to test this type of variables also after data collection.

1.10 Attitudes towards migration in Europe

Session chair:
Ian Preston; University College London, United Kingdom

This session takes place on:
Friday, July 22; 11:30 to 13:30

Presentations:

- Invited presentation: *Immigration, Economy and Culture: Analysis of Attitudinal Responses* - David Card (United States), Christian Dustmann, Ian Preston (United Kingdom)

Contributed presentations:

- *Resistance to immigrants and asylum seekers in the European Union* - Marcel Coenders, Marcel Lubbers, Peer Scheepers (Netherlands)

- *Measuring Attitudes towards Immigration across Countries with the ESS: Problems of Equivalence and Reasons for Inequivalence* - Nina Rother (Germany)

- *Format effects in measurements of political tolerance* - Kristen Ringdal, Anders Todal Jenssen, Albert Simkus (Norway)

- *Attitudes Towards Migration In Europe: A Cross-Cultural And Contextual Approach* - Bart Meuleman, Jaak Billiet (Belgium)

1.10.1   Immigration, Economy and Culture: Analysis of Attitudinal Responses

David Card; University of California, Berkeley, United States

Christian Dustmann; University College London, United Kingdom

Ian Preston; University College London, United Kingdom

Abstract.   In this paper we study attitudinal responses of host country residents towards immigration. We consider opinions on a variety of relevant impacts and provide empirical analysis that is based on data from the first European Social Survey specifically suited to pick up the many channels of interest through which benefits and costs of immigration may be felt. Results establish strong associations between individual characteristics and a wide range of responses to questions relating to perceived impact of immigrants on economic and social outcomes and point towards the relative importance of relevant considerations.

1.10.2   Resistance to immigrants and asylum seekers in the European Union

Marcel Coenders; University of Nijmegen, Netherlands

Marcel Lubbers; University of Nijmegen, Netherlands

Peer Scheepers; University of Nijmegen, Netherlands

Abstract. In this contribution we focus on the present resistance to immigrants and asylum seekers in European Union countries. We take advantage of recently collected cross-national high quality data providing means to rigorously test hypotheses from previously accumulated theoretical propositions with advanced statistical methods. We find large differences between countries which, however, could not be explained well by economic and demographic characteristics of these countries. We found that resistance to immigrants and asylum seekers strongly prevails among under-privileged people but also among self-employed people. Moreover, we found that distrustful people, people overestimating the presence of out-groups in their country, and people perceiving migrants to pose a threat to resist strongly to immigrants and asylum seekers.

1.10.3   Measuring Attitudes towards Immigration across Countries with the ESS: Problems of Equivalence and Reasons for Inequivalence

Nina Rother; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract. To know about attitudes towards immigration is essential not only for researchers but especially for politicians. Immigration and integration plays a more and more important role in European societies nowadays. In order to provide a knowledge base on attitudes towards immigration in Europe, a special module on attitudes towards immigration was included in the ESS round 1. However, attitudes towards immigration are not easy to measure and especially to compare across cultures. Different migration histories and policies in the different European countries make it hard to think of a shared understanding of what immigration and immigrants are.

A cross-cultural comparison of attitudes towards immigration can only be done when functional equivalence is given, i.e. that the same construct is measured in all cultures. Hence, this has to be tested before analysing the data on the substantive level. Functional equivalence of data can be affected by three sources of biases: construct bias, method bias and item bias. In order to analyse the functional equivalence of ESS immigration attitudes measures two distinct analytical strategies are employed. The first is made up by a secondary analysis of the ESS 2002/2003 data. Data-analytical tools employed among others factor analyses with target rotation. However, in many cases, evidence from secondary analysis is not conclusive and additional studies are needed, such as cognitive interviews and split-ballot experiments. For this purpose, the results of 96 cognitive interviews conducted in 2003/2004 in 6 European countries are presented, e.g. by demonstrating reasons for the presence of bias in the ESS.

The results show that some of the ESS measures might not be regarded as functionally equivalent. Some possible solutions are outlined

1.10.4   Format effects in measurements of political tolerance

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

Anders Todal Jenssen; Norwegian University of Science and Technology (NTNU), Norway

Albert Simkus; Norwegian University of Science and Technology (NTNU), Norway

Abstract. Tolerance for (extremist) groups one dislikes or even fears is important for a well functioning democracy. Tolerance for such groups may be measured in several ways in a survey. The traditional Stouffer approach is used in the American General Social Survey (GSS). The respondents are asked whether each group should be forbidden a specific activity. In the Sullivan, Pierson and Marcus format the respondents are asked to pick the groups that should not be allowed a specific activity from a list. In addition the respondents should pick the group they dislike the most. The purpose of this paper is to test whether the two ways of measuring political tolerance will yield different estimates of the level of political tolerance, and whether the formats are related differently to explanatory variables.

Data: A survey in Poland based on personal interviews conducted by CEBOS in a representative sample of N=4000. Half of the respondents in a random split will be exposed to each form.

We measure whether a set of nine groups should be allowed to arrange public meetings, and participate in radio and television programs. This yields 18 item in the GSS format, and the pick groups format yields two questions plus three questions on the most disliked group to get at intensity of feelings and fear for that group.

The survey will be conducted in April 2005. For this reason no results are yet available.

1.10.5   Attitudes Towards Migration In Europe: A Cross-Cultural And Contextual Approach

Bart Meuleman; University of Leuven, Belgium

Jaak Billiet; University of Leuven, Belgium

Abstract.   Given the tendency towards harmonization of the migration policies of the EU-member countries, international comparisons of attitudes towards migration are highly relevant. In this paper, attitudes towards migration are compared across 22 countries that participated in round 1 of the European Social Survey. The ESS contains an extensive module of migration-related items. Out of this module, 4 scales are made up, namely (1) the willingness to allow immigrants into the country, (2) conditions for allowing immigrants, (3) the perceived threat of immigration and (4) the evaluation of the consequences of migration.

In order to make valid comparisons across countries possible, the concepts have to be measured in a sufficiently equivalent way. For comparisons of the county-means, at least scalar equivalence is needed. For each of the four scales, this level of equivalence is tested using the multi-group structural equation modelling (MGSEM) approach.

Once scalar equivalence is established, the 22 countries involved are compared with respect to the four constructed latent variables. Apart from a mere comparison, a statistical model is constructed to explain the differences between countries. The influence of a number of relevant individual level and country level variables is scrutinized using multi-level modelling. The results and shortcomings of this analysis are discussed.

1.11   Cross cultural comparable scales in the ESS

Session chair:
    Jaak Billiet; University of Leuven, Belgium

This session takes place on:
    Friday, July 22; 15:00 to 17:00

Presentations:

- Invited presentation: *Are Differences In Meaning Detected By Tests Of Factorial Invariance? Evidence From ESS Round 1* - Jaak Billiet, Bart Meuleman (Belgium)

Contributed presentations:

- *Functionality of rating scales in survey research* - Peter Hagell (Sweden), Katarína Vasilová (Slovakia)

- *Cross-cultural patterns in attitudes towards biotechnology* - Sally Stares (United Kingdom)

- *Criteria for equivalence of measurement instruments in cross cultural research* - Willem Saris (Spain)

1.11.1   Are Differences in Meaning Detected By Tests Of Factorial Invariance? Evidence From ESS Round 1

Jaak Billiet; University of Leuven, Belgium

Bart Meuleman; University of Leuven, Belgium

Abstract. One of the prime objectives of cross-national survey research is to compare concepts across countries or cultures. It is therefore important that these concepts are measured adequately in all of the countries involved in the survey. Moreover; in order that country-scores on items or scales can be compared in a valid way, concepts have to be measured in a sufficiently equivalent way. In the past we assessed the equivalence of a set of indicators dealing with the acceptance of immigrants. It was fount that different methods like multi-group SEM and the proportional odds model (POM) lead to somewhat different conclusions. It is more difficult to conclude to a complete invariance with POM since more aspects of the response distribution of each particular indicator are considered.

In this paper, two other sets of items are considered in which we fount at least a translation problem in one item that caused considerable differences in the observed response distributions in comparable countries. The first translation bias was found in an item on asylum seekers (D41) in France leading to considerable differences in response distribution comparing with Luxemburg, Switzerland, and Wallonia (French speaking part of Belgium). The second translation bias was detected in the context effect of two items on 'serious' and 'any' crime (D23-D24) in Denmark in contrast with Norway and Sweden. The problems were detected in the course of substantial analysis or by the media because of implausible response distributions. Is it possible to detect this kind of bias during tests of factorial invariance by SEM or by POM. The two approaches are compared on their ability to detect this kind of bias in translation.

1.11.2   Functionality of rating scales in survey research

Peter Hagell; University of Lund, Sweden

Katarína Vasilová; Slovak Academy of Sciences, Slovakia

Abstract. European Social Survey is a survey designed to obtain answers that reflect the actual attitudes, beliefs, values and behaviors of respondents across Europe. As attitudes, values and beliefs are abstract constructs and one can only assume their relationship with physical events (like behaviors or the tick in the questionnaire), these variables are known as "latent variables". This survey research, similarly to other surveys used the questionnaire with different response scales. Among the most commonly used are rating scales typically consisting of 3-11 defined or non-defined numerical steps. A larger number of steps gives the impression of improved separation between responders and can improve correlation and reliability coefficients. However, outcome interpretation is dependent on the functionality of the rating scale. Thus, as people move from less to more along the underlying latent variable the successive response categories should express more and more of the latent variable, and each category should be modal (i.e., the most probable) outcome at some point along the latent continuum.

The objective of the study was to explore the functionality of 11- and 6-pont rating scales and whether collapsing categories may improve rating scale functionality and affect reliability. Data from the European Social Survey (ESS) 2002/2003 were used. Data from the UK (n=2052) on social trust (3 items with 11-point rating scales), and human values (21 items with a 6-point rating scale) were analyzed by means of the Rasch measurement model.

The results indicate that parts of the 11-point social trust scales displayed disordering and/or non-modality along the underlying latent variable. For the 6-point human values scale, category 3 displayed borderline modality. After collapsing response categories steps into 5-point scales all categories in both scales appeared in a distinct successive modal manner without any disordering along the underlying latent variable. Long response scales failed to display expected functionality, but improved following combination of adjacent response categories. Reliability was marginally affected. These observations may be due to difficulties in making fine tuned distinctions among levels of endorsement. These pilot observations support the use of 5-point scales.

1.11.3   Cross-cultural patterns in attitudes towards biotechnology

Sally Stares; London School of Economics, United Kingdom

Abstract. This paper presents analyses of cross-national survey data asking for citizens' opinions of biotechnology. Such data are of significant interest not only to social scientists working in the broad field of Public Understanding of Science, but also to other 'actors' in biotechnology: those in politics, academic and industrial science, media and activist groups regularly make reference to surveys describing levels of support for various applications of the technology, such as genetically modified foods, cloning, and genetic testing. Cross-cultural comparisons of opinions are often crucially informative to those engaging in debates surrounding the future of these applications. For example, such comparisons have been invoked in discussions of the World Trade Organisation trade suit filed by the United States against the European Union for its moratorium on genetically modified crops.

Empirical research to date such as Gaskell et al. (2003) and Hviid Nielsen et al. (2002) demonstrates that underlying simple 'yes or no' verdicts on biotechnology are a range of patterns of opinion. For example, citizens tend to judge medical and agricultural applications using different criteria of evaluation: medical applications tend to be judged as more risk-worthy, for example, than agri-food applications. However, such judgements vary not only according to type of application, but also according to respondent characteristics. Some important characteristics are: feelings of optimism or scepticism about technology generally, levels of engagement with and knowledge about science, and values such as religious beliefs. Interacting with these characteristics are socio-demographic factors such as age, gender, education and nationality.

This paper presents a set of analyses using data from the Eurobarometer 2002 module on biotechnology. It focuses on a battery of questions with Likert-type response options that ask for judgements of utility, riskiness, moral acceptability and overall support for different applications of biotechnology. Using latent class analyses, patterns of support and opposition are identified, and these are compared between country samples and between types of biotechnology application. A key feature of these data is the high rates of 'don't know' responses –more than 40% on some items. Therefore an essential part of the analysis is an interpretation of such responses. Since the latent class models are suitable for nominal variables, one approach to this is simply to include 'don't know' responses in the class models, to determine with which types of response profiles they are associated. Another approach is to use latent trait models to ascertain where 'don't know' appears on each underlying response continuum. These trait analyses also give an indication as to the ordinality or otherwise of the response scales, and to the scalability of the item sets.

The analyses presented, therefore, give some insights into substantive patterns of opinion on biotechnology across European countries in 2002. They also offer an evaluation of differential item functioning across samples, which provides valuable diagnostic information regarding the extent to which we can be confident in drawing comparisons between country samples using these survey questions.

Keywords: attitudes, biotechnology, latent class models, latent trait models, 'don't know' responses

1.11.4 Criteria for equivalence of measurement instruments in cross cultural research

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. Normally equivalence of measurement instruments in cross cultural research is evaluated on the basis of equality of factor loadings. But the size of the factor loadings is a product of quality of the observed measure and the internal consistence of the items with respect to the general factor. If the internal consistency coefficients are equal but the quality of the measures is not across countries then the factor loadings will differ from country to country. On the other hand the factor loading may be the same but the internal consistency coefficients may be different then the instruments are not comparable across countries.

On the basis of this argument we suggest that the criterion for equivalence of measurement instruments in cross cultural research is not equality of factor loadings but equivalence of internal consistence coefficients i.e. de factor loading after correction for measurement error.

2   Data collection (Aula 004)

2.1 Quality control of surveys

Session chair:
    Giovanna Brancato; ISTAT, Italy

This session takes place on:
    Tuesday, July 19; 9:00 to 11:00

Presentations:

- Invited presentation: *Enhancing the harmonisation of the statistical production within ESS: Recommended Practices for Questionnaire Development and Testing methods in CATI and CAPI surveys* - Giovanna Brancato (Italy)

Contributed presentations:

- *Who is concerned about their privacy? A socio-demographic background and consequences of the concern about privacy on the cooperation with surveys* - Vicky Storms, Geert Loosveldt (Belgium)

- *Coverage and quality of the UK Annual Survey of Hours and Earnings (ASHE)* - Pete Brodie (United Kingdom)

- *Quality assurance framework for the 4th European Working Conditions Survey* - George Petrakos (Greece), Enrique Fernández Macías (Ireland)

- *On some non sampling errors in the Norwegian Election Surveys 1997-2003* - Øyvin Kleven, Ib Thomsen, Li-Chun Zhang (Norway)

2.1.1 Enhancing the harmonisation of the statistical production within ESS: Recommended Practices for Questionnaire Development and Testing methods in CATI and CAPI surveys

Giovanna Brancato; ISTAT, Italy

Abstract. The first systematic work on quality in the European Statistical System, conducted in the framework of the Leadership Group (LEG) on Quality, has highlighted that much efforts have been placed in the harmonisation on the side of the outputs, whereas still there is the need of harmonising the inputs, and in particular the methodology, and of sharing the existing good practices developed within the European Organizations (Grünewald et al., Lyberg et al.). Many standardising tools have been evaluated and it has been concluded that the most appropriate instrument to be implemented at European level is represented by the Recommended Practices (RPs). The RPs are a collection of proven good methods for performing different statistical operations (Berghdal et al.).

It was identified that one of the most relevant areas to start to develop European RPs was the questionnaire development and testing methods. To this aim a European project, financed in the framework of Eurostat grants for 2004, Theme 10, "Quality Management and evaluation", was launched. The project is coordinated by the Italian National Statistical Institute with the participation of representatives from Germany and Portugal . The work is focused on Computer Assisted Telephone and Personal Interviewing (CATI and CAPI) techniques. The first step was to gather information both on the most recent methodology and on the tools and practices in use in the NSIs by means of literature search and the submission of a state-of-the-art questionnaire. On the basis of this review the RPs will be developed, tested, and reviewed with the support of experts from some network countries (Statistics Netherlands, Statistics Norway, Statistics Finland and Statistics Austria).

2.1.2 Who is concerned about their privacy? A socio-demographic background and consequences of the concern about privacy on the cooperation with surveys

Vicky Storms; University of Leuven, Belgium

Geert Loosveldt; University of Leuven, Belgium

Abstract. Concerns about privacy are said to be one of the causes of non response. Starting from the identification of the problem of privacy in surveys (1979, American Statistical Association) up till now, numerous research was performed in order to investigate the problem further (for an overview –see Mayer, 2002). To monitor this concern about privacy and some other attitudes toward surveys, the Flemish Government's Administration of Planning and Statistics (APS) decided to introduce some new questions into their annual survey about socio-cultural topics. The results show that a majority of the Flemish people were concerned about their privacy at some point. However, this concern did not seem to be linked to surveys since only 7% of the respondents worried about privacy during surveys. Concerns about privacy are linked to the gender and age of the respondent, but not to the educational level. Trust in government and its institutions is also a good predictor of the respondent's general concern about privacy, but not of the concern about privacy in surveys. Concerns about privacy are important for survey researchers since they seem to have consequences for the cooperation in surveys. One of the consequences of the concern about privacy in surveys is more item non response on a question about the family income. Other consequences of a concern about privacy are a smaller inclination to cooperate again in similar surveys in the future and a smaller cooperation rate to a self-administered supplemental survey at the end of the interview. This effect is even more evident when considering the reactions of the sample persons at the doorstep (responders as well as refusers). Only 12% of the sample persons who make a remark about the survey as an invasion of their privacy can be convinced to cooperate in the survey.

2.1.3 Coverage and quality of the UK Annual Survey of Hours and Earnings (ASHE)

Pete Brodie; Office of National Statistics, United Kingdom

Abstract. This paper describes analysis carried out using data collected for the 2004 Annual Survey of Hours and Earnings (ASHE). From 2004 ASHE replaced the New Earnings Survey (NES) which had been running since the 1970s. Details of the results and methodology for ASHE can be found on the National Statistics Website (www.statistics.gov.uk under Labour Market, Earnings). The main part of the survey consists of a one in a hundred simple random sample of all employees currently in a Pay As You Earn (PAYE) scheme while the supplementary portions of the survey aim to capture those employees that have moved jobs (changed employer) between selection

and questionnaire despatch as well as those employees that are outside the PAYE scheme (mostly people who work for smaller employers). The supplementary portions of the sample were included as pilot studies (not to be included in the calculation of the published estimates) for 2004 and 2005 with the decision of whether or not to continue after 2005 to be based on the added value that is produced by each supplement. The overall sample design is a combination of single stage element and single stage cluster design. Estimates of earnings, hours, numbers of low paid etc. are then produced by weighting the dataset. Weighting is carried out by calibration to population estimates produced by the main Labour Force Survey (LFS - details also available at www.statistics.gov.uk) subdivided by age, sex, location and broad occupational category. The analysis examines the coverage, bias and variance of the estimates produced and comments on the value added by the supplementary portions of the survey.

2.1.4 Quality assurance framework for the 4th European Working Conditions Survey

George Petrakos; Agilis SA, Greece

Enrique Fernández Macías; European Foundation for the Improvement of Living and Working Conditions, Ireland

Abstract. The European Foundation of Dublin has been carrying out its European Working Conditions Survey since 1991, with a new wave every five years. In line with the recent increased concern about quality in surveys, the Foundation has devised a quality assurance framework that will be implemented for the first time in the 2005 wave of the survey. This paper presents the general approach to quality assurance of this framework, discussing the complexities and problems involved in the implementation of a homogeneous survey throughout all European countries, both from a methodological point of view and in a purely organizational sense.

2.1.5 On some non sampling errors in the Norwegian Election Surveys 1997-2003

Øyvin Kleven; Statistics Norway, Norway

Ib Thomsen; Statistics Norway, Norway

Li-Chun Zhang; Statistics Norway, Norway

Abstract. During the last 40 years Statistics Norway has performed several election surveys. For some key variables, like party distribution and electoral turnout, we have observed an increase in the discrepancies between the survey results and the actual vote, even after having weighted the sample using demographic information. During the same period, we have observed an increase in non reponse rates and other types of errors, which might explain the negative development.

Several authors have argued that surveys on political topics tend to get higher response rate among the more political interested, and that people who are less interested in politics tend not to participate. In this paper we explore the influence of a variety of factors affecting response rates and measurement errors and their impacts on the final results from the surveys. Several alternative non-response and measurement error models will be tested. The key variables are electoral turnout, party identification, interest in politics, refusals, converted refusals and other non-respondents.

The data used are the Election surveys 1997, 2001 and 2003. The design of these surveys is a rolling panel. In addition to information collected in the surveys, we also have information from the municipal election committee on whether each person in the samples voted or not

2.2   Survey research of special populations

Session chair:
   Keming Yang; University of Reading, United Kingdom

This session takes place on:
   Tuesday, July 19; 11:30 to 13:30

Presentations:

- Invited presentation: *Overcoming Constraints on Conducting a Social Survey in a British Town* - Keming Yang, Deborah Støer (United Kingdom)

Contributed presentations:

- *A Pilot Survey among Turks in Denmark* - Mette Deding, Torben Fridberg, Vibeke Jakobsen (Denmark)

- *Mode effects on data quality: Benefits and drawbacks of mixed-mode business surveys* - Anna Stangl (Germany)

- *Development of a questionnaire for standardized study of undocumented immigrants* - Mila Paspalanova (Belgium)

- *The effect of data collection mode and ethnicity of the interviewer on response rates and self-reported alcohol use among Turks and Moroccans in the Netherlands: and experimental study* - Regina van den Eijnden, Aafje Dotinga, Willem Bosveld, Henk Garretsen (Netherlands)

2.2.1　Overcoming Constraints on Conducting a Social Survey in a British Town

Keming Yang; University of Reading, United Kingdom

Deborah Støer; University of Reading, United Kingdom

Abstract.　It has been widely acknowledged that conducting a social survey is an art of balancing scientific standards and practical constraints. This is especially true when resources at the disposal of the survey researcher are limited. Nevertheless, reports and advices on how best to overcome the constraints with strategies that are at least acceptable, if not admirable, are still rare. It seems that researchers have only two opposite options –either they should conduct a social survey only when the available resources are enough for satisfying the demands of scientific principles, or they should give up the idea of administering a first-hand survey to analyse data collected by those with enough resources but a different research objective in mind. In this paper we report our experience of dealing with such a dilemma –being requested by Thames Valley Energy to conduct a survey on local residents' attitude toward green energy in the town of Reading with very limited financial and personnel support, and no access to information essential for drawing a random sample. Our strategy was to compile a sampling frame based on a local map and to adopt the drop-off-and-collect method in administering the fieldwork. We shall discuss both the benefits and the disadvantages of following such a strategy.

2.2.2　A Pilot Survey among Turks in Denmark

Mette Deding; Danish National Institute of Social Research (SFI), Denmark

Torben Fridberg; Danish National Institute of Social Research (SFI), Denmark

Vibeke Jakobsen; Danish National Institute of Social Research (SFI), Denmark

Abstract. Conducting surveys among immigrants is typically more difficult than surveying the majority population. One problem is that immigrants are difficult to get in contact with due to wrong addresses, cash card mobile phone numbers, and long periods away from Denmark. Furthermore, language barriers and the accumulation of socio-economic factors connected to higher non-response rates are more prevalent among immigrant groups. The consequence is that both total and partial non-response is often relatively high in surveys among immigrants.

In preparation to a larger survey among immigrants in Denmark, a pilot survey among 1st and 2nd generation immigrants from Turkey is carried out. The purpose of the pilot survey is to evaluate the data collecting method as well as analysing total and partial non-response qualitatively. Concerning the data-collecting method, we focus on the introduction letter and the situation of the interview. We study the effect of language barriers and gender issues, e.g. translation problems and different interviewer-interviewee gender combinations. In the pilot-questionnaire, we focus on non-partial response by including clarifying questions about reasons for non-response. Although there are large differences with different immigrant groups, the pilot survey only includes Turks. This allows us to exclude the effects of cultural differences from our results. The findings from the pilot survey will be used to tailor the larger survey and thus increase the quality of the data.

2.2.3　Mode effects on data quality: Benefits and drawbacks of mixed-mode business surveys

Anna Stangl; Ifo Institute for Economic Research, Germany

Abstract.　The study is based on a panel of independent economic experts from over 90 countries who were surveyed quarterly in 2002 and 2003 in the context of the World Economic Survey (WES). The results of the longitudinal analysis demonstrate that response rates to Internet-based questionnaires, given the choice of survey mode, are approaching the response rates to traditional paper-and-pencil techniques and Internet based questionnaires can be regarded as almost equally popular among a panel of economic professionals. Furthermore, on-line participants tend to be more strongly committed to the survey. The reason is seen in the stronger and more frequent e-mail communication between on-line participants and the surveyor that helps to foster social bonds and respondents' commitment.

Particular attention is drawn to the precision / reliability of the Internet-based questionnaire as a data collection method. The frequency of error scores / inconsistent responses is in Internet questionnaires significantly higher than in questionnaires received with the traditional paper-and-pencil technique. An interesting finding is that in the

on-line mode the occurrence of inconsistent answers coincides with a lower item non-response. This combination provides reasons for concern that in the case of on-line surveys a low item non-response is not a quality criterion, but may derive from the tendency of on-line participants to be less thorough in answering to the questions. Some practical issues of conducting an on-line and fax survey simultaneously are discussed.

2.2.4 Development of a questionnaire for standardized study of undocumented immigrants

Mila Paspalanova; University of Leuven, Belgium

Abstract.   The presence of undocumented immigrants in the industrialized countries is a largely studied and discussed phenomenon. The research in the field, though concentrated around several key questions (who are the undocumented immigrants, how did they arrive and what are their future plans) is rarely uniform and the lack of systematic approach towards answering these basic questions does not allow for comparisons on national, cross-cultural and even regional level. The abundance of empirical results is seldom explained or interpreted in relation to the method of data collection. In the field of empirical migration research, data is collected via non-structured or the semi-structured interviews, designed especially for the particular research purposes, and virtually never reported by the researcher.

The current paper is a response to the lack of standardized instruments for studying undocumented immigrants. The questionnaire is developed to systematically register the primary questions of interest when an under-studied population of undocumented immigrants is explored; and would allow for comparison of undocumented populations. An exhaustive pool of open and closed ended questions allows for recording, among others, of undocumented immigrants': 1. Educational status and language skills, including pre and post migration educational history and the relation between immigrants' educational qualifications and employment status; 2. Living conditions and tendencies for upward social mobility over time; 3. Social life, prevailing type of social relationships, and subjective perception of integration; 4. Migration history, motivations for choosing the destination city, information sources about living and working possibilities in the destination city; 5. Employment history and occupational activities before and after migration; 6. Health status and work accidents; 7. Remittances and other resource flows between the immigrant and the home country; 8. Stimulating further migration; 9. Legal aspects (reasons for undocumented residence status; regularization plans and motivations pro and against regularization) and contact with law enforcement authorities; 10. Minor children residing with their parents in and out of the city of immigration.

Based on the questionnaire administration among 80 undocumented Bulgarian and 30 Polish immigrants in Brussels, the paper will offer a discussion over the choice between open and closed ended questions in relation to the studied topic; justification of the selection of the particular research questions; an analysis of the non-working questions and response categories; time for questionnaire administration and coding of the responses

2.2.5 The effect of data collection mode and ethnicity of the interviewer on response rates and self-reported alcohol use among Turks and Moroccans in the Netherlands: and experimental study

Regina van den Eijnden; IVO (Addiction Research Institute), Rotterdam, Netherlands

Aafje Dotinga; RIVM, Bilthoven, Netherlands

Willem Bosveld; Dienst Onderzoek en Statistiek, Amsterdam, Netherlands

Henk Garretsen; University of Tilburg, Netherlands

Abstract. Alcohol research among migrants with an Islamic background faces several methodological problems. Because of their Islamic religious orientation, which prescribes alcohol abstinence, it is generally assumed that these migrant groups underreport their alcohol use. The present study investigates self-reported alcohol use among individuals with a Turkish and Moroccan background living in the Netherlands.

The study had an experimental design. The effect of data collection mode (mail survey mode versus face-to-face interview mode) and ethnicity of the interviewer (ethnically matched interviewer versus Dutch interviewer) was tested on response rates and self-reported alcohol use. The study was conducted among 728 Turks and 745 Moroccans, which were randomly allocated to the experimental conditions. The study was accompanied by non-response analyses.

The results showed that non-response was higher in the mail survey mode than in the interview mode. Respondents to the interview mode more often reported using alcohol than respondents to the mail survey mode. However, respondents to the mail survey reported higher alcohol use and more excessive drinking than respondents to the interviews. Within the interview mode, Dutch interviewers generated higher alcohol reports than ethnically matched interviewers.

Based on the assumption that higher alcohol reports reflect more reliable data ('the more is better' rule), mail surveys seem more appropriate to measure degree of alcohol use and excessive drinking among Turks and Moroccans. However, face-to-face interviews held by Dutch interviewers seem more reliable to measure the prevalence of drinking among these ethnic groups. On basis of these finding, a multi-method approach combining these research methods seems recommendable.

2.3   Sampling and variance estimation

Session chair:
  Sabine Häder; Centre for Survey Research and Methodology (ZUMA), Germany

This session takes place on:
  Tuesday, July 19; 15:00 to 17:00
  Tuesday, July 19; 17:30 to 19:30

Presentations:

- Invited presentation: *Sampling for the European Social Survey* - Sabine Häder (Germany)

Contributed presentations:

- *A mixed design for studying policies of childcare institutions in Denmark* - Michael Davidsen, Inge Lissau (Denmark)

- *Responserate and Non-response Using the Gabler-Häder Sampling Technique* - Wolfgang Donsbach, Olaf Jandura (Germany)

- *The Probability Space of Statistical Surveys* - George Petrakos (Greece)

- *Variance Estimation in Complex Surveys from a Cross National View* - Ralf Münnich (Germany)

- *Software packages for variance estimation - Results from the Dacseis project* - Seppo Laaksonen (Finland)

- *A new measure of interviewer variability for a complex sampling design* - Siegfried Gabler (Germany), Partha Lahiri (United States)

2.3.1 Sampling for the European Social Survey

Sabine Häder; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   In the meantime the second round of the European Social Survey (ESS) has been finished. The survey –conducted as face-to-face interviews –covered again more than 20 nations. It employed the most rigorous methodologies, including sampling strategy. Strict random sampling was a basic requirement. Additionally, three rules for the implementation were set up: full coverage of the defined target population, a minimum target response rate of 70% and the same minimum effective sample size in all participating countries.

In the first two rounds the sampling resources, especially the available frames, varied considerably between countries. This lead to very different designs from simple random sampling to clustered four-stage designs. It was the task of an expert panel to give advice and support in finding workable and equivalent sampling strategies in the different countries. Besides, design effects due to clustering and due to differing inclusion probabilities could be predicted based on the results of round 1. These efforts increased the quality of sampling and improved cross-national comparability. In the presentation one member of the sampling expert panel will report on the experience with this project.

2.3.2 A mixed design for studying policies of childcare institutions in Denmark

Michael Davidsen; National Institute of Public Health, Denmark

Inge Lissau; National Institute of Public Health, Denmark

Abstract. Background: In these years there is focus on obesity, food habits and physical activity among children. The nutrion and physical activity policies are important for the food and as well as for activity in childcare institutions. Therefore nutrition and physical activity policies of schools, after school care and kindergarten/pre-school are of interest. In 2004 a study of Danish childcare institutions was conducted.

Aim: To describe the complex design involving mixing of well-known design methods and a stopping rule and its statistical analysis.

Design: The following demands were set up prior to construction of the design: 20% of all childcare institutions should be included, the design should ensure geographical spread, special attention should be paid to cities and municipalities included in the study should have a usable feed-back. To fulfil these criteria the country was subdivided into 4 regions such that each of the four biggest cities in Denmark was in separate regions. In each region the following strategy was applied: the city was chosen and 20% of each of its childcare institutions (kindergarten, after school care institutions and schools) were chosen at random. Among the remaining institutions all institutions in a number of randomly selected municipalities was chosen. The municipalities were selected consecutively stopping the selection when 20% of the institutions in the region (city excluded) had been chosen. The selection probability of each municipality was proportional to the number of childcare institutions in the municipality. All selected institutions received an institution-specific questionnaire. The design thus includes stratification, simple random sampling and cluster sampling in combination with 'certainty units' in two different levels.

Statistical analysis: Weighting was necessary to obtain country-specific figures. Institutions in a city received a weight of 5. All institutions in a municipality received a weight calculated as the number of institutions in the region divided by the number of municipalities chosen ('steps') times the number of institutions in the municipality. The SUDAAN-programme includes facilities to compute reliable variance estimates in this mixed setting.

Results: 8758 institutions were identified. Of these 4229 (48.3%) were kindergarten, 2450 (28.0%) after school care institutions and 2079 (23.7%) were schools while 1662 (19.0%) were situated in the four cities. In all 1875 institutions (21.4% of all identified childcare institutions) were selected. These came from 38 municipalities including the four cities –this corresponds to 14.6% of all Danish municipalities. The distribution of institution types in the sample was similar to that in the entire country. In all 1323 childcare institutions (70.6% of all selected institutions) returned the questionnaire, the distribution across institution types was quite similar; A few results from the study will be presented.

Conclusion: Seemingly simple criteria may lead to quite complicated designs. The design worked and produced reliable results.

2.3.3 Responserate and Non-response Using the Gabler-Häder Sampling Technique

Wolfgang Donsbach; University of Dresden, Germany

Olaf Jandura; University of Dresden, Germany

Abstract. For years, in Germany samples for telephone samples that claimed to be representative were drawn using the numbers published in telephone directories. This was possible because of an obligation to have one's number listed there. This, however, changed in 1993 when customers were allowed to decide for themselves whether or not their number should be listed. Since then, the number of people not represented in Germany's phone directories has been ever in-creasing, making it impossible for researchers to ensure that all members of a given popula-tion had the same chance of being included in a survey when the selection was based on printed or electronic telephone directories.

Unsurprisingly, when studies were conducted in as early as the mid-1990s when the abolition of the listing obligation dated only a few years back they produced hard evidence that there were systematic differences between listed and unlisted households. Hence, in Germany sam-ples could no longer be drawn from phone books or CD-Roms. Of course, other countries had experienced similar problems much earlier and thus other methods of drawing samples for telephone surveys were developed. In the USA, for instance, phone numbers were generated by chance, either by substituting the last number of a given phone number by a random digit (LNDD) or by creating the whole number by chance (RDD). The latter, especially, cannot be done in Germany however as the numbers here differ in length and thus too many invalid numbers would be produced. In 2002, Gabler and Häder developed a numbering system that is based on existing phone numbers and thus diminishes the production of nonexistent num-bers. The chance of inclusion is the same for all households with a telephone no matter whether the number is listed or not.

We have a dataset of 18.013 phone numbers that were contacted during phone surveys con-ducted at the Communication Research department of the University of Technology, Dresden. Those numbers were drawn using the Gabler-Häder design. The numbers were provided by the German research institute ZUMA which also checks and discloses to the researchers whether these numbers are listed or not. For every single try to contact a number we kept re-cord of the date(s), time of day, number of try and whether the try was successful or not. In all, more than 30.000 tries to contact those 18.013 numbers lead to 3.061 interviews. As we used an identity code for the numbers contacted and the interviews conducted we could match the contacting data with the data of the respective respondents without breaking the rules set out by data protection regulations.

Using this unique dataset we want to answer two research questions:

(1) Do the respondents which have their numbers listed differ from those who refrain from publishing their number regarding (a) their willingness to take part in surveys and (b) socio-demographic variables such as age, sex, education, income, or employment?

(2) Are there differences between those respondents that can be reached by one or two tries from those that you can only reach after four or more calls?

Our data suggests that (1a) people who have their number listed are more likely to refuse to take part in a survey than those who withhold their number and that (1b) there are significant differences between the two groups in terms of all sociodemographic variables mentioned. On the other hand we can show that consisting with previous findings people that can be reached easily vary from those that have to be contacted more often until they are reached (2). How-ever, even if we took those together who shared the same sociodemographic characteristics (i.e. were of the same age, sex and education) we found that when splitting them into two groups based on how hard to was to contact them we found that their attitudes varied.

Hence we can show that the Gabler-Häder design is a very useful method to draw samples for telephone surveys in Germany as it ensures that both listed and unlisted numbers have an equal chance of inclusion. This is important because –as we demonstrate –people who make use of their right to withhold their number vary in more ways from their listed counterparts than just their preference of number publication. At that, we can add to the evidence that it is vital to undertake several attempts to contact numbers as people who are easy to get differ substantially from those who are hard to reach.

2.3.4 The Probability Space of Statistical Surveys

George Petrakos; Agilis SA, Greece

Abstract. Key Words: Statistical Survey, Probability Space, Probability Measure, Sampling.

A statistical survey consists, among others, of a random experiment where a sample is selected from a target population with a well-defined procedure. The outcomes of the experiment, the selection of the sample and the construction of probability measures associated with them, form a probability space $(\Omega, B, \mu\iota)$. Each element of the triplex is defined and examined for properties, like completeness, additivity continuity and more, with emphasis given to the construction of different probability measures $\mu\iota$ associated with respective sampling schemes. Effort has been made so that the proposed probability system can accommodate the individualities and the complexity of modern statistical surveys within official statistical organizations.

2.3.5 Variance Estimation in Complex Surveys from a Cross National View

Ralf Münnich; University of Tübingen, Germany

Abstract. When comparing statistics across countries, one generally has certain ideas of rankings. In order to adequately compare estimates gained from different countries and sources, one should investigate confidence intervals or at least the variances of the statistics. Therefore, adequate variance estimates have to be gained from the samples in addition to the point estimates.

The aim of this presentation is to elaborate a wider spectrum of variance estimation methods and their efficiency for household and individual surveys. These include direct and approximate methods as well as several resampling based methods. The methodology is enriched by an appropriate consideration of non-sampling errors, focusing on non-response problems.

To elaborate the impact of practical factors on the point estimators, such as non-sampling errors and the complexity of the design, a large simulation study was conducted. The results aim at building an important source for best practice recommendations. The entire study was conducted within the IST-2000-26057-DACSEIS project which was financially supported by Eurostat within the 5th Framework Programme of the European Commission.
Source:
Data Quality in Complex Surveys within the New European Information Society
http://www.dacseis.de

2.3.6 Software packages for variance estimation - Results from the Dacseis project

Seppo Laaksonen; Statistics Finland and University of Helsinki, Finland

Abstract. The EU research project DACSEIS (www.dacseis.de) provided a thorough overview on the methodology and practice of variance estimation covering various special cases and problems in the field of survey sampling. The project included also a part for evaluation of statistical software for variance estimation (standard errors, confidence intervals) in surveys. This paper presents a summary of these results. The software evaluated were Bascula, Clan, Poulpe, SAS, SPSS, Stata, Sudaan and WesVar but we looked the characteristics of some other software too including R. The evaluation was based on the pre-developed criteria and then some real survey data files were exploited. A general conclusion is not easy to present but we can say however that no software package is so superior that all important requirements of a demanding user would be satisfied without additional programming. On the other hand, all packages together are quite exhaustive in such cases when the data do not include missing values.

2.3.7 A new measure of interviewer variability for a complex sampling design

Siegfried Gabler; Centre for Survey Research and Methodology (ZUMA), Germany

Partha Lahiri; University of Maryland, United States

Abstract. In the European Social Survey (ESS) one of the important tools for comparing the different sampling designs of the different participating countries is the use of design effects which involve the effect due to unequal weighting and the effect of clustering. But this measure does not take into consideration the interviewer variability which is known to be one of the major sources of variability in surveys. In this presentation, we provide a model-based justification of a new measure of interviewer variability for a complex sampling design. This general formula for the overall design effects incorporates effects due to unequal weighting and homogeneity effects arising from spatial clustering and the interviewers. For self-weighting designs with no spatial clustering and equal interviewer workload the effect due to interviewer is the usual formula ieff = 1 + (n - 1)rho.

2.4 Data collection problems in business surveys

Session chair:

Jacqui Jones; Office of National Statistics, United Kingdom

This session takes place on:

Wednesday, July 20; 9:00 to 11:00

Wednesday, July 20; 11:30 to 13:30

Presentations:

- Invited presentation: *Introduction* - Jacqui Jones (United Kingdom)

Contributed presentations:

- *Data collection problems is business research* - Antoni Dorse Lopez, Irmtraud Gallhofer, Willem Saris (Spain)

- *Frame Error Impact on Structural Business Statistics Surveys* - Salvatore Filiberti, Caterina Viviano (Italy)

- *Mixed mode data collection and nonresponse bias reduction in a panel survey among municipalities in The Netherlands* - Robert Voogt (Netherlands)

- *Is the mixed mode design the most efficient design for carrying out business surveys? Evidence from an employer survey in the United Kingdom* - Emanuela Sala, Peter Lynn (United Kingdom)

- *The Challenges of Implementing Web Data Collection for Mandatory Business surveys - the respondents' perspective* - Zoë Dowling (United Kingdom)

- *Measuring Perceived and Actual Response Burden in Business Surveys* - Trine Dale, Gustav Haraldsen (Norway), Jacqui Jones (United Kingdom), Dan Hedlin (Sweden)

- *Reducing Response Burden: Redesigning the Dutch Annual Business Inquiry* - Ger Snijkers, Deirdre Giesen, Robert Göttgens, Hank Hermans, Evrim Onat, Jo Tonglet, Myra Wieling (Netherlands)

- *Data Collection Trough Questionnaires With Value Added Responses* - Alberto Gimeno (Spain)

2.4.1 Introduction

Jacqui Jones; Office of National Statistics, United Kingdom

Abstract. Business surveys incorporate a wide range of different subject matters ranging from pay and personnel information to detailed financial information such as pension funds. Similarly businesses are diverse covering different industries, size and geographical locations. Until recently survey methodologists had not awakened to the facts that business surveys also needed to utilise the methodological developments e.g. question testing that had been seen in social surveys; and that collecting data from businesses involved more response steps than in social surveys.

This paper will provide an overview of the characteristics of business surveys, focusing on how data has been traditionally collected; the issue of response burden; and the respondent(s) response process. Throughout the paper potential data collection problems will be identified and discussed.

2.4.2 Data collection problems is business research

Antoni Dorse Lopez; ESADE, Universitat Ramon Llull, Spain

Irmtraud Gallhofer; ESADE, Universitat Ramon Llull, Spain

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. In a study with respect to the Transfer Pricing Policies of Multinationals data have been collected from the financial managers of subsidiaries of Multinationals in Catalonia. The sample was drawn from a list of firms obtained from the Chamber of Commerce. The companies were contacted by phone to make an appointment for a face to face interview. In the paper we present the sampling design, the way the data collection was organized and the efforts it took to get the cooperation of 134 companies out of the 300 selected for the study. Attention will be given to the use of interviewers, the position of the secretary, the effect of company size and other factors. We will provide some recommendations for future research.

2.4.3 Frame Error Impact on Structural Business Statistics Surveys

Salvatore Filiberti; ISTAT, Italy

Caterina Viviano; ISTAT, Italy

Abstract. Keywords: Frame Errors, Administrative Data, Structural Business Statistics, Business Register.

Frame errors represent a considerable non sampling error source within the structural business statistics (SBS) surveys. In fact, main information that are necessary for the data collection about enterprises are not perfectly updated at the time of the building of the frame of the population of interest and launch of the sampling surveys. Business registers may have technical difficulties in getting all the relevant information about economic units from the supplying data sources in good time. This may cause bias and it also may reduce accuracy of the final SBS estimates. For these reasons it is important to assess the effects that such error sources may cause on final SBS estimates. In this paper a methodology based on combination of auxiliary administrative data and survey data is proposed. Also, first results of the application are shown.

2.4.4 Mixed mode data collection and nonresponse bias reduction in a panel survey among municipalities in The Netherlands

Robert Voogt; Ministry Of Social Affairs and Employment, Netherlands

Abstract. This paper desribes a four wave panel survey among the 500 municipalities in The Netherlands. The goal of the survey was to monitor the implementation of a new social law by the Dutch municipalities. The results of the survey were used to trace possible problems in the implementation process and to inform the Secretary of State of Social Affairs and Employment and the Parliament about the progress the municipalities made. This meant it was very important that the results correctly reflected the state of affairs concerning the implementation. In a period of a year –the municipalities had one year to implement the new law - the municipalities were asked on four different moments to answer a short, factual questionnaire on the internet.

After getting a response rate of only 25% at the first panel wave, several adjustments were done to secure a higher response rate at the second panel wave: the results of the first survey were published in a newsletter that was send to all municipalities, the possibility was given to send in a paper version of the questionnaire, after a week all municipalities that hadn't answered were send an e-mail reminder and after two weeks a subsample of nonresponding municipalities was asked by phone to fill in the questionnaire. This lead to a substantial higher response of 76%. In the third and fourth panel wave, when these adjustments were further improved, an even higher response rate was obtained –topping at 91% in the fourth panel wave. This proved that, by having a good design, it is possible to reach a very high response rate in survey among municipalities.

By applying a hot deck imputation method that used both the results of the subsample of nonrespondents (for this group a response rate of more than 95% was reached) and the results of previous and later panel waves, it was possible to correct for nonresponse bias and to use the complete sample in the analyses that the data were used for.

We also compared the results obtained for the group of municipalities that only participated after having received a reminder by phone and the group that participated after having received the e-mail reminder. It turned out that leaving out the phone group did not significantly changed the results. As a reminder by phone is much more expensive than an e-mail reminder, this means that a substantial financial saving can be obtained without having to compromise on the quality of the data.

The approach, which combined mixed mode data collection and subsampling a group of nonrespondents, is currently repeated for a much more burdensome two wave panel survey in 2005. Probably some first results of the first panel wave can be presented on the conference.

2.4.5   Is the mixed mode design the most efficient design for carrying out business surveys? Evidence from an employer survey in the United Kingdom

Emanuela Sala; University of Essex, United Kingdom

Peter Lynn; University of Essex, United Kingdom

Abstract.   Establishment surveys have to deal with methodological issues that are quite different from the ones researchers normally face while carrying out other kinds of surveys. The process of making contact and obtaining co-operation, for example, is usually longer and more complicated, while the dynamics of the response process are specific to this kind of survey. This paper addresses issues related to the efficiency of alternative designs for business surveys. Specifically, we aim to evaluate the efficiency of different mode designs with respect to two main issues, namely sample bias and costs. The objective is to assess the relative cost-effectiveness of two alternative designs at minimising non-response bias.

We use data from an employer survey that has an unusual and advantageous characteristic for our purposes. It was carried out as a part of a methodological project aimed at assessing the quality of a longitudinal dataset and consequently a set of auxiliary information collected previously is available for the whole sample (including non respondents). These data will therefore allow us to evaluate non-response bias with the survey designs that we adopted.

The paper is divided in 6 sections. After briefly discussing the methodological issues related to establishment surveys (section one), we describe the context in which the employer survey was carried out and we discuss the survey design that we have adopted for this study, a postal stage with a telephone follow-up (section two). In section three we briefly examine the response rates obtained at the different stages of the survey and we describe reasons for refusals. We then compare the structure of the sample obtained by using a one mode (postal) design with the one obtained adopting a mixed mode (postal plus telephone) design in order to check if a) the sample obtained with the first design is biased, b) the sample obtained with the multimode design reduces such bias. We will assess the sample bias using three sets of variables: variables related to the firm (number of people employed, Standard Industrial Classification 92, location of the firm), variables related to the employees' occupation (Standard Occupational Classification, having managerial duties, being an employee or a self-employed, hours of wok, pay), demographic variables of the employees (sex, age, race). Section five gives some information on the marginal costs of data collection accruing to the different stages of the survey. Section six concludes giving an overall evaluation of the two stage (postal and telephone) design adopted, compared with a simple one stage postal survey.

2.4.6 The Challenges of Implementing Web Data Collection for Mandatory Business surveys - the respondents' perspective

Zoë Dowling; University of Surrey, United Kingdom

Abstract. Keywords: Business surveys, Web data collection, respondent perspective, new mode data collection, challenges

Government statistical agencies that require businesses to provide various details about their activities face many challenges in the data collection process. The statistical agency works within the demanding parameters of minimising response burden to the business whilst maintaining acceptable levels of data quality, timeliness and curtailing survey costs. Thus statistical agencies are perpetually investigating new means to improve their data collection processes. The World Wide Web, as a result of ever-growing internet connectivity within businesses, now presents a new mode for data collection and one that appears to offer many improvements to current methods.

Whilst Web reporting may be advantageous to the statistical agencies, it is necessary to ask how respondents view this new option. What expectations would they have of the Web as a mode of response, what design and functionality features would make it worthwhile and finally would the respondent, given the choice to complete a Web questionnaire, over a paper or telephone version, chose the Web option?

During 2004, 30 face-to-face interviews were undertaken with current respondents to UK Office for National Statistics mandatory business surveys. These interviews focused upon the respondent's experience of the mandatory survey(s), their use of the internet, and importantly, design and functionality features of Web questionnaires. The interviews included the use of visual prompts –a paper version of a mandatory survey and a Web version of a section of the questionnaire.

The data gathered demonstrates that, on the whole, respondents are receptive to the idea of returning their data via the Web. Some concerns around security and workflow issues were raised. Expectations of the Web questionnaire varied owing to a number of factors ranging from computer competency levels to perception of the task. In general, ease of use was the most commonly cited expectation.

This paper arises from on-going PhD research at the University of Surrey. It funded by the UK Economic and Social Research Council with the UK Office for National Statistics as collaborating partners. The views expressed in this paper are entirely my own and do not necessarily represent those of my funders.

2.4.7 Measuring Perceived and Actual Response Burden in Business Surveys

Trine Dale; Statistics Norway, Norway

Gustav Haraldsen; Statistics Norway, Norway

Jacqui Jones; Office of National Statistics, United Kingdom

Dan Hedlin; Statistics Sweden, Sweden

Abstract. Keywords: perceived response burden, business survey burden, response burden guidelines

Statistics Norway, The UK Office for National Statistics, and Statistics Sweden have been working together on a Eurostat Leadership in Quality project: Developing methods for assessing perceived response burden. Based on qualitative research in the earlier phases of the project, question sets to measure both actual response burden and response burden as perceived by business survey respondents were developed and tested both qualitatively and quantitatively in all three countries. The questions measure factors such as what aspects of the survey respondents perceive as burdensome, the perceived usefulness of the survey, the actual time used on collecting information and filling in the questionnaire as well as the number of people involved in completing a survey questionnaire. The tests were carried out on both paper and pencil and on web surveys and show that the questions are well suited to measuring response burden. Measurements of response burden by the use of these questions can also be used to assess the quality of the data collected.

In this paper the question sets and the results from the field tests will be presented and discussed. We will also present guidelines on how to measure response burden that were developed based on the research findings in the project.

2.4.8   Reducing Response Burden: Redesigning the Dutch Annual Business Inquiry

Ger Snijkers; Statistics Netherlands (CBS), Netherlands

Deirdre Giesen; Statistics Netherlands (CBS), Netherlands

Robert Göttgens; Statistics Netherlands (CBS), Netherlands

Hank Hermans; Statistics Netherlands (CBS), Netherlands

Evrim Onat; Statistics Netherlands (CBS), Netherlands

Jo Tonglet; Statistics Netherlands (CBS), Netherlands

Myra Wieling; Statistics Netherlands (CBS), Netherlands

Abstract.   Statistics Netherlands is facing a major challenge with regard to the data collection of business surveys. The major reason behind this challenge is the reduction of response burden. Both the policy of the Dutch government as well as individual businesses demand this. In order to get a high enough response rate, that is non-selective, and valid and reliable data in the future, Statistics Netherlands has to adapt to these demands. A second reason is that Statistical Netherlands by law is entitled to use registers for its production of statistics. But apart form these reasons, Statistics Netherlands feels that the data collection should be more respondent friendly, as well as efficient.

As for the Dutch Annual Business Inquiry, a number of projects have been carried out to achieve these goals.

1. A first project was aimed at reducing the length of the questionnaire by investigating the output demands.

2. In a second project was investigated how the sample could be reduced by use of register data.

3. In yet another project the remaining paper questionnaire was redesigned in order to make it more respondent friendly.

4. A last project is aimed at developing an internet questionnaire.

With these projects this important survey is adapted to the response process for businesses.
In the presentation these projects and their results will be presented.

2.4.9   Data Collection Trough Questionnaires With Value Added Responses

Alberto Gimeno; ESADE, Universitat Ramon Llull, Spain

Abstract. This presentation shows the data collection strategy developed by ESADE's research group in family business. Through an alliance with the service provider Family Business Knowledge (FBK), the researchers have access to a continuous increasing database of family business companies.

Based on the Structural Risk Management Model, developed by Alberto Gimeno from ESADE and Gemma Baulenas (collaborator in ESADE's programs), FBK has developed an on-line system that generates a diagnosis of the family-business situation and, according to the theoretical model, makes suggestions of possible management practices to be developed by the respondent.

The diagnosis requires the fulfilment of an anonymous quite extensive questionnaire. The quality of the response determines the quality of the feedback received, which enhances the reliability of the responses.

The combination of valuable feedback and anonymity, make the participation in the survey research valuable to the participant. In this way large amounts of data about business firms are obtained without any costs for data collection.

Compared with the tedious way data are collected normally in business research, this approach seems to work very well. In the presentation more details of the approach and it limitations will be discussed.

2.5   Data collection, non-response and design in Health surveys

Session chair:
    Cornelia Bormann; Gesundheitsforschung, BMBF, Germany

This session takes place on:
    Wednesday, July 20; 15:00 to 17:00

Presentations:

- Invited presentation: *Overview on Health surveys in Europe* - Cornelia Bormann (Germany)

Contributed presentations:

- *National Health Surveys in Germany* - Thomas Ziese (Germany)

- *The Danish Health Interview Survey Programme* - Mette Kjøller (Denmark)

- *Effect on trend estimates of the difference between survey respondents and non-respondents: results from 27 populations in the WHO MONICA Project* - Hanna Tolonen (Finland), Annette Dobson (Australia), Sangita Kulathinal (Finland)

- *Cognitive problems with health questions: determinants and verbal indicators* - Stasja Draisma (Netherlands)

2.5.1   Overview on Health surveys in Europe

Cornelia Bormann; Gesundheitsforschung, BMBF, Germany

Abstract.   The health status of populations is measured in most of the European countries by population based health surveys. These surveys differ by design aspects as f. ex.
  ●Cross-sectional vs. longitudinal approach
  ●Postal vs. interview vs. examination survey
  ●"subjective"or "objective aspects of health
  ●different target groups (children, adults, elderly)
  ●different health indicators and instruments
Athough the survey approach is mostly recognized and the quality of data from population surveys in Europe has greatly improved over the last 10 –15 years, there exist still some difficulties, which concern the comparability of these data especially. The establishment of agreed standards in terms of recommended common methods and measurement instruments is one topic, which is tackled by the WHO-Euro and by the EU (EUROHIS project).

In the speech an overview of the health surveys in the European countries will be given and the different approaches and activities concerning the comparability of the data will be presented.

2.5.2   National Health Surveys in Germany

Thomas Ziese; Robert Koch Institute, Germany

Abstract.   Health surveys are an important data source for information on health status and health behaviour. The combination of Health Interview Surveys and Health Examination Surveys provide comprehensive and comparable information to show changes in health and supplies central indicators for the development of health policies.

The Robert Koch Institute has carried out health surveys on national level at irregular intervals since the 1980s. The first three national health surveys (1984 –86, 1987 –89 and 1990 –91) were originally designed to serve as 'control group' to measure the impact of certain health intervention. The last German Health Examination Survey was carried out in 1998. In this survey, about 7000 subjects of a representative sample of the population aged between 18 and 79 years were interviewed and examined. The response rate was 61.4%. These surveys cover a wide range of health relevant topics: e.g. perceived health, chronic conditions and morbidity, quality of life, risk factors, living and working conditions, health care utilisation, sociodemographic factors. This information is obtained by interview or questionnaire. Furthermore a blood and urine specimen was collected, as well as measurements of blood pressure, height and weight.

On one hand examination surveys provide a broader range of information on health and offer the opportunity of validation by comparing laboratory data with questionnaire data for example. One the other hand, examination surveys are more complex to organize and much more expensive than Health Interview Surveys. Therefore Interview Health surveys supplement the examination surveys to provide continuous health data on a annual basis.

The Department of Health Reporting at the Robert Koch Institute conducts nationwide telephone health interview surveys since 2002, financed by the Federal Ministry of Health and Social Security. The Health telephone surveys will also serve as a module of a health-monitoring system in Germany, which is currently in discussion.

From September 2002 to March 2003, 8,313 German speaking people randomly selected out of the resident population aged 18 years and over were asked on topics like chronic deceases, health behaviour and the extent to which they make use of health care services and others. In addition 7,341 telephone interviews were conducted in a follow-up survey from September 2003 to March 2004 dealing with further aspects of health and illness or health related behaviour. Currently a third survey is on the run. In order to be able to identify changes in the health status and in health-related behaviour at the level of individuals, for the first time people are recalled, who took part in the first survey and consented to participate in further surveys. A well trained team of interviewers assures high quality of data. The sampling frame guarantees random selection, including non-listed numbers in public phone directories.

For the first time the Robert Koch Institute is currently conducting a national health examination survey for children and adolescents. About 18 000 persons in the age from zero to 18 take part in the survey. This survey covers a broad area of health relevant aspects and living conditions which will be described in details in the presentation.

2.5.3   The Danish Health Interview Survey Programme

Mette Kjøller; National Institute of Public Health, Denmark

Abstract.   Introduction

The Danish National Institute of Public Health regularly conducts national representative health interview surveys. The programme changes between general and broad HIS, which were conducted in 1987, 1994, 2000 and HIS with specific subjects (1991 and 2003). Data collection for the 2005 health interview survey is ongoing.

Purpose

The purpose of the programme is to describe the status and trends in health and morbidity in the adult population and in the factors that influence health status, including health behaviour and health habits, lifestyles, environmental and occupational health risks and health resources. The results are used in national, regional and municipal planning and monitoring as well as in research.

Design

The 1987 and 1994 surveys were each based on a random sample of 6000 Danish citizens aged 16 years or older. In the 2000 survey the sample was increased to 22500 persons and organized in such a way that we had 1000 completed interviews in each of the 15 counties. Further all invited persons from the 1994-survey were included, thus combining a cross sectional and a longitudinal design. The sample design for the 2005-survey is similar to the 2000-survey taking account of a forthcoming new regional structure in Denmark.

The programme is based on a set of core elements that are used in every HIS. The core elements are structured in a traditional epidemiological model describing risk factors that might lead to ill health and morbidity, as well as the consequences of ill health. Different additional topics are included in each survey due to current discussion on health policy or requests from researchers.

Data are collected through personal interview in the respondent's home, followed by a self-administered questionnaire, which the respondent is asked to fill in and return after the interviewer has left.

Non-response

The total non-response increased from 20.0% in 1987 to 22.0% in 1994 and 25.8% in 2000, with the percentage of refusals increasing from 12.1% to 16.4% and 22.4% respectively. Non-response is associated with gender, age, marital status and geographical area. Non-response is highest among older women and among persons living in the metropolitan area but lowest among widowed and married men.

Detailed studies, where the whole sample is followed in the Danish Hospital Register, have documented, that the hospital admission rate is higher among non-responders than responders during the data collection –but identical before and after. Anyhow, this difference is too small to bias the estimation of morbidity in the population based on information from the respondents in the survey.

Presentation of results

Examples will be given.

2.5.4   Effect on trend estimates of the difference between survey respondents and non-respondents: results from 27 populations in the WHO MONICA Project

Hanna Tolonen; National Public Health Institute (KTL), Finland

Annette Dobson; University of Queensland, Australia

Sangita Kulathinal; National Public Health Institute (KTL), Finland

Abstract.   Introduction. Sample surveys are commonly used to obtain data on the epidemiological characteristics of the population. There is evidence that survey respondents and non-respondents differ from each other in socio-economic and health profiles. In this study, our first purpose was to examine differences in socio-economic and health profiles among respondents and non-respondents in 27 populations. Our second purpose was to investigate the effect of non-response on estimates of trends.

Methods. In the WHO MONICA Project, which involved a series of cross-sectional health examination surveys, an effort was made to obtain basic data on non-respondents. Non-respondents were contacted either by mail, phone call or in few cases by home visit and asked to fill in a short questionnaire, which included few key questions about the socio-economic status and health profile. These data from the final survey were compared between respondents and non-respondents. The potential effect of non-response on the estimates of trends between the initial survey and final survey approximately ten years later was then investigated.

Results. In most of the populations, non-respondents were more likely to be single, less well educated, and had poorer lifestyles and health profiles than respondents. The differences were smaller among women than men. Ten-year trends in prevalence of daily smokers were overestimated in most populations if they were based only on data from respondents. For example, in one population among men, the results based only on respondents indicated that the smoking prevalence had declined 6% in ten years. After adjusting for non-response, there still was decline in the smoking prevalence was only 3% in this period.

Conclusions. Our results support the earlier findings that the respondents and non-respondents differ in their socio-economic and health profiles. The difference between respondents and non-respondents is fairly consistent across the populations. Estimators of population trends based only on respondent data might also be biased. Declining response rates therefore pose a threat to the accuracy of estimates of risk factor trends in many countries. It would be important to try to collect some information about non-respondents, which then can be used to evaluate the potential bias caused by non-response.

2.5.5 Cognitive problems with health questions: determinants and verbal indicators

Stasja Draisma; Vrije Universiteit Amsterdam, Netherlands

Abstract. Questions that pose problems to respondents can result in verbal and paralinguistic doubt expressions (e.g. filled pauses, speech rate et cetera). Such doubt expressions indicate inaccuracy and uncertainty, shown earlier in a validity study of Draisma & Dijkstra (2004). For the present paper, verbal interactions in CATI health interviews are analyzed, to identify underlying cognitive problems in answering health questions. Not only doubt indicators may reveal problems with questions, specific types of qualified answers, with the use of certain arguments signify the kinds of troubles that occur during question answering. For example, an older respondent may respond with a conditional qualification to the item 'I am healthy' with the expression: 'considering the fact that I have lung problems for already fifteen years, I feel relatively healthy, so I agree'. In the health survey used for this study, a sample of 610 respondents was confronted with four different question types: opinion questions, simple behaviors/facts, subjective perceptions of own health and frequency estimates of (un)healthy behaviors. These four categories are found to be related to different kinds of verbal qualifications, also known as projective reports (e.g. Moore, 2004).

In the present study the relationship between tangible variables like age and reported health history on the one hand and qualified answer types on the other hand are also investigated, as it is assumed that such variables are determinants of problems encountered during question answering. Problems regarding specific items and response categories will become apparent from the arguments used in projective reports. The analysis of projective reports is a relatively new method to identify cognitive and communication difficulties, which will enhance insight into issues regarding validity and specificity of health survey instruments.

Keywords: interaction analysis, qualified answers, validity, cognitive problems and question types, pretesting questionnaires, health questionnaire.

2.6   Nonresponse effects

Session chair:
  Uwe Engel; University of Bremen, Germany

This session takes place on:
  Thursday, July 21; 9:00 to 11:00

Presentations:

- Invited presentation: *Nonresponse effects in sample surveys* - Uwe Engel (Germany)

Contributed presentations:

- *Controlling for patterns of non-response bias in small scale surveys in Australia* - Steven McEachern (Australia)

- *Respondents and Non-Respondents in European Social Survey: Similar or not? A Case of Poland* - Pawel Sztabinski, Franciszek Sztabinski, Dariusz Przybysz (Poland)

- *Non-response bias in Switzerland: lessons from the ESS 2002 and 2004* - Nanette Wälti, Nicole Schoebi, Dominique Joye (Switzerland)

- *Consequences of Panel Mortality in the German Socio Economic Panel for the Estimation of Life Expectancy* - Mark Trappmann, Rainer Schnell (Germany)

2.6.1   Nonresponse effects in sample surveys

Uwe Engel; University of Bremen, Germany

Abstract.   The paper deals with determinants and consequences of nonresponse in sample surveys. The main focus is on unit nonresponse. Starting points are two meta-analyses of experimental research findings on reported response rates in sample surveys. A multilevel analysis is carried out on incentive effects studied under a total of 177 conditions, while another analysis on the effectiveness of contact related fieldwork procedures is based on an enlarged set of 278 conditions. Findings are discussed with respect to explanations offered in the relevant literature on survey participation, in particular referring to the prominent social exchange, social integration and rational choice hypotheses. In doing so, special attention is paid to the question when prepaid incentives work or do not work. In a certain sense, this is tantamount to ask whether a person rejects the survey request for rational choice reasons or is cooperative for reasons of social or economic exchange. And regarding the exchange respectively tit for tat principle, the question is whether its effectiveness is due to getting paid for the time spent on the interview or an inherently moral commitment. To explore this social vs. economic exchange topic in terms of possible expectations held, findings from a sample of students as well as a sample drawn from an access panel population that were asked for relevant information are presented.

Nonresponse is a potential threat to the accuracy of sample estimates. It is somewhat misleading practice to assess the quality of samples by reference to the response rate criterion, since by and large this criterion neglects the bias component. To assess the latter, evidence on nonrespondents is necessary but usually difficult to obtain. Facilitation might be provided by tailored survey designs affording the flexible adoption of variably long questionnaires and flexible scheduling of interviews in conjunction with response propensity modelling and weighting that exploits the inherent accessibility information.

Finally, the topic of nonresponse effects in panel surveys is addressed by example of a multi-wave panel analysis of GSOEP data carried out to test status inconsistency hypotheses.

2.6.2   Controlling for patterns of non-response bias in small scale surveys in Australia

Steven McEachern; University of Ballarat, Australia

Abstract.   Understanding and compensating for response bias remains a high priority for researchers interested in maximising the value of social survey data in social science research. Dillman (2000) identifies four basic areas of error in social surveys: coverage, sampling, non-response and measurement. Correcting or controlling for each of these forms of error enables the researcher to improve the overall quality of their data and their research.

Recent research in the area of survey non-response has demonstrated the impact of geographic (Fraboni et al, 2002), demographic (Dillman et al, 2001), interviewer (Houtkoop-Steenstra, 2000, Hox et al, 1991) and mode (Abraham et al, 2002) effects on non-response in social surveys. There is also strong evidence that the involvement of particular institutions, either as sponsors or data collectors (Groves and Couper, 1998) has a significant effect in improving response rates. However, much of the recent research in survey response has been based on government-sponsored studies, using national sampling frames and institutional support. Less well understood is the impact of such institutional effects on small-scale surveys conducted for academic or commercial purposes. This paper therefore seeks to address this gap in existing knowledge.

The paper presents an overview of the recent work of the computer-assisted survey research facility, part of an Australian Research Council funded data collection facility established as a pilot to facilitate computer assisted survey research in Australian universities. It then provides an analysis of the survey outcomes and patterns of non-response of a selection of recent CATI survey projects conducted by the facility in both academic and contract research contexts. A multilevel modelling approach is adopted to analyse the impacts of study topic, sponsor, interviewer, demographics and geographic location on non-response.

The paper then continues on to discuss a methodological experiment incorporated into the CATI survey introduction to address the problem of understanding demographic patterns of non-response. An approach to controlling for variations in survey non-response in presented, using interviewer estimates of age and gender among both respondents and non-respondents, along with actual age and gender information provided by respondents. This approach provides a novel means of re-weighting for non-response error in situations where non-respondent characteristics are unknown

2.6.3 Respondents and Non-Respondents in European Social Survey: Similar or not? A Case of Poland

Pawel Sztabinski; Polish Academy of Sciences, Poland

Franciszek Sztabinski; Polish Academy of Sciences, Poland

Dariusz Przybysz; Polish Academy of Sciences, Poland

Abstract.   Over the past few years many countries, including Poland, have seen a clear decline in response rates in face-to-face interviews. While fieldwork standards introduced in the ESS project helped Poland to raise the response rate, it still oscillated around 70% in ESS round 1 and round 2. Hence, a question remains to be answered about the differences between individuals who participated in the survey and those who did not. As a matter of fact, this is a question about the legitimacy of extrapolations based on the effective sample.

We tackle this problem drawing on data from ESS Round 2 pilot study and the main ESS survey (Round 2) in Poland. An mail questionnaire has been circulated to individuals who did not participate in these surveys. It contained selected questions from the ESS questionnaire, the same questions in both cases (after the pilot study and after the main study).

First, the respondents (i.e. these who participated) and non-respondents (individuals who did not participate in the surveys but responded to the mail questionnaire) are compared with regard to their socio-demographic characteristics. Then we check if ESS response (opinion) distributions differ between those two groups in a statistically significant way. We also check if the control of socio-demographic variables modifies those relationship in any way, i.e. to what extent the differences between the groups can be explained through differences in their respective structures.

Secondly, we ask if the associations between variables (e.g. opinions and socio-demographics) differ between the two groups. Log-linear models are used for data analysis to attempt the answers.

We have also drawn on the available data to verify the hypothesis stating that a long fieldwork period (here: in main study ESS as opposed to the pilot study) generates greater differences between respondents and non-respondents. A long fieldwork period reduces the percentage of non-contacts and ineligible individuals ('light' non-respondents). As a result, 'heavy' non-respondents may be expected to differ more substantially from the respondents.

It is important to note that the scope of conclusions from our analysis is limited since the data (responses to mail questionnaire) were obtained only from some non-respondents (the respective response rates: 52.4% in the pilot study and 35.3% in the main study). The second limitation is that we have compared face-to-face interview data against mail questionnaire data

2.6.4 Non-response bias in Switzerland: lessons from the ESS 2002 and 2004

Nanette Wälti; SIDOS, Switzerland

Nicole Schoebi; SIDOS, Switzerland

Dominique Joye; SIDOS, Switzerland

Abstract. Different researches have already shown that the people who do not answer to surveys are probably not identical to the people who do participate (i.e. non-respondents). Furthermore, it seems that the non-respondents do not constitute a homogenous group and that we have at least to distinguish between people that cannot be reached and people that can be contacted but refuse to participate. If this is correct, this means that the quality of responses is also de-termined by the strategies used in order to obtain the collaboration of potential respondents.

In Switzerland, this question is particularly important because the response rate for face to face surveys is relatively low. In the 2004 edition of the European Social Survey (ESS) great efforts were made in trying to reduce the number of refusals with the result that the response rate was notably higher than in the 2002 edition. This means that the characteristics of the respondents taken as a whole may have changed. In this contribution, we will try to identify in both surveys the main features of respondents that where difficult to reach but finally partici-pated and respondents that had to be converted to participate. We will also use this informa-tion to estimate the possible non-response bias in the ESS in Switzerland

2.6.5 Consequences of Panel Mortality in the German Socio Economic Panel for the Estimation of Life Expectancy

Mark Trappmann; University of Konstanz, Germany

Rainer Schnell; University of Konstanz, Germany

Abstract. One major source of bias in panel studies results from panel mortality:

People who once participated drop out of the study due to refusal or unreachability. Problems occur when this dropout is informative, i.e. when dropout is connected to an outcome under study. These problems get worse if other variables in the study do not suffice to predict dropout. This is most likely the case with respect to health variables. Not only do we expect that people who are sick are more likely to refuse, but moreover do we assume that especially those people might refuse whose state of health deteriorated since the last panel wave. If this assumption is correct, we are not able to correct for the bias either by weighting or by imputation.

In this presentation we address this problem concentrating on the estimation of life expectancy from panel data from the German Socio Economic Panel (GSOEP). As Heller and Schnell (2000, 128) showed, death one to three years after a wave is a reasonable predictor for refusal in this wave even in the presence of all other predictors from the GSOEPs propensity weighting procedure used to correct the bias caused by panel mortality. Because the information about future death is not known at the moment of refusal, this result suggests informative censoring.

Using models of survival analysis to estimate life expectancy from panel data, dropouts have to be coded as censored because there is no information about their survival thereafter. Our hypothesis is –as stated above

- that contrary to the assumptions of such models refusers are on average more sick than people who are still in the panel and thus are at greater hazard of dying in the years following their refusals than people who are still in the panel. We also assume that propensity weighting used in the GSOEP is not sufficient to correct this bias, because important predictors are unobserved .

We can test our hypothesis with data from the GSOEP respondent tracking in

2001: A request at registration offices could confirm the year of death or the place of living (and thus the survival) of most of the dropouts.

Comparing estimates for life expectancy from the GSOEP before and after the respondent tracking we can quantify the amount of bias assuming that the data are error free afterwards. We can validate these calculations by comparing both results to data from official statistics.

2.7   How to cope with decreasing response rates

Session chair:
    Leo van Doorn; Marketresponse, Netherlands

This session takes place on:
    Thursday, July 21; 11:30 to 13:30

Presentations:

- Invited presentation: *Non response in marketing research* - Leo van Doorn (Netherlands)

Contributed presentations:

- *The long-term effectiveness of refusal conversion procedures on longitudinal surveys* - Jonathan Burton, Peter Lynn, Heather Laurie (United Kingdom)

- *Reducing Nonresponse by SMS Reminders in Three Sample Surveys* - Vesa Virtanen, Timo Sirkiä, Juha Nurmela (Finland)

- *Enhancing Response Rate in ESS Round 2: Case of Slovakia* - Michal Kentoš, Jozef Výrost, Anna Takáčová (Slovakia)

- *How to decrease response rates and reduce bias* - Ineke Stoop (Netherlands)

2.7.1 Non response in marketing research

Leo van Doorn; Marketresponse, Netherlands

Abstract. As other forms of social science research it's commercial cousin marketing research has been affected by rising refusal rates and greater difficulties in reaching people in general. Cost of research is of great importance in the marketing research arena but time constraints are even more important and very often determine whether or not a project will be done at all not. This means that not all the options of reducing (the effects) of non response can be employed in those situations. In my paper I will go describe the strategies that are employed in commercial research to combat nonresponse. My evaluation will show some of the bad, even some of the downright ugly but also some of the good ways to provide users of marketing research with reliable and representative samples.

2.7.2 The long-term effectiveness of refusal conversion procedures on longitudinal surveys

Jonathan Burton; University of Essex, United Kingdom

Peter Lynn; University of Essex, United Kingdom

Heather Laurie; University of Essex, United Kingdom

Abstract. The maintenance of high response rates is an important objective for most surveys. Increasingly, survey organisations are having to make greater efforts in order to achieve respectable response rates. This paper examines a method used by some survey organisations: the attempted "conversion"of sample members who refuse to take part in a survey.

Persuasive techniques are used in an effort to get the refusers to change their mind and agree to an interview. This is done in order to improve response rate and, possibly, to reduce non-response bias.

However, refusal conversion attempts are expensive and must be justified. Previous studies of the effects of refusal conversion attempts are few and have been restricted to cross-sectional surveys.

The criteria for "success"of a refusal conversion attempt are different in the case of a longitudinal survey, where for many purposes the researcher requires complete data over multiple waves. This paper uses data from the British Household Panel Survey (BHPS) from 1994 to 2002 to examine the characteristics of those most likely to refuse, and those likely to be converted and to assess the long term effectiveness of refusal conversion procedures in terms of sample sizes, sample composition and non-response bias.

2.7.3 Reducing Nonresponse by SMS Reminders in Three Sample Surveys

Vesa Virtanen; Statistics Finland, Finland

Timo Sirkiä; Statistics Finland, Finland

Juha Nurmela; Statistics Finland, Finland

Abstract. In this article we will describe our experimental use of the short message service (SMS) for reducing nonresponse in three mail surveys (n1=532, n2=5,000 and n3=11,300). SMS as a new technical solution for survey research can be seen as a supplementary tool for the data collection process. We exclude pure SMS questionnaires from this review. The mobile phone has virtually penetrated the whole population of Finland. New users are being recruited even from older age groups. Has text reminding any effect on nonresponse? Is there any evidence of nonresponse bias or this so-called 'digital divide' bias in using SMS in questionnaires? These practical questions of survey implementation are in the focus in this article. We performed an analysis with the co-operation rate as the dependent variable. We tested whether there was any difference between the experimental (SMS text message) and control (standard postcard reminder/advanced letter or no SMS message) groups in response rate. We found statistically significant differences between these respondent groups. The results were encouraging. The SMS text message can be seen as a supplementary tool for self-administered surveys. These recently collected tests were very practical ones, but they can also be considered from the background of the Leverage-Saliency Theory of Survey Participation (Groves et al.) or as a new toolkit for the Total Design Method (Dillman).

2.7.4 Enhancing Response Rate in ESS Round 2: Case of Slovakia

Michal Kentoš; Slovak Academy of Sciences, Slovakia

Jozef Výrost; Slovak Academy of Sciences, Slovakia

Anna Takáčová; Slovak Academy of Sciences, Slovakia

Abstract. European Social Survey perceives enhancing response rate as essential element of striving to achieve high standards. For these purposes certain fieldwork procedures in ESS were prepared by CCT. In our paper we would like to describe on empirical data, how this procedures worked in our cultural settings. Most of ESS participating countries in order to achieve high response rate prepared some additional, country-specific procedures. It was the case in Slovakia, too, and in the second part of our paper we would like to discuss the experiences gained during the fieldwork .

2.7.5 How to decrease response rates and reduce bias

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Abstract.   Whereas most researchers focus on enhancing response rates, the present paper will present an overview of strategies that are bound to decrease response rates. Some of these are well-known, others may require some explanation. A few examples of successful strategies (mainly in face-to-face surveys) are:
- ask people for their cooperation on the telephone first (although you might end up with a high response rate if you ignore initial nonresponse);
- set a low upper limit to the number of calls;
- make face-to-face calls during office hours only (interviewers have a family life too);
- follow-up appointments only when you need more respondents;
- don't bother with an advance letter and a leaflet describing the survey: people do not read this stuff anyway;
- don't try to convert initial refusals (a refusal is a refusal);
- stick to a single interview mode;
- try to economize on interviewer remuneration;
- don't monitor interviewers closely (as they prefer organizing their own schedules);
- don't spend to much money on interviewer' training and instruction as their turnover rate is high anyway;
- use a questionnaire efficiently and include as many questions as you can (better safe than sorry);
- don't hand out an incentive (a survey is expensive enough as it is).

The effectiveness of these strategies can be deduced from The Hunt for the Last Respondent (Stoop, 2005) in which an overview is given of strategies to enhance response rates. The presentation will show how very high response rates can be achieved.

Aiming for lower response rates does not reduce bias. The sad truth, however, is that lower nonresponse rates neither guarantee a reduction of nonresponse bias. The relationship between response rates, fieldwork efforts and nonresponse bias is difficult to unravel, as in many cases it will not be clear how much and which type of efforts were required to establish contact and obtain the cooperation of increasingly difficult participants, and if and how respondents differ from nonrespondents. An empirical study will be presented in which this information was collected. From this study it appears that –beyond a certain level –efforts aimed at enhancing the response rate may be less effective than directly aiming for bias reduction, or at least for correcting for nonresponse, by collecting data on the fieldwork process and the final nonrespondents.

Stoop, Ineke A.L (2005) The Hunt for the Last Respondent. Nonresponse in Sample Surveys. The Hague, SCP (www.scp.nl)

2.8   Weighting for nonresponse in European comparative research

Session chair:
> Vasja Vehovar; University of Ljublana, Slovenia

This session takes place on:
> Thursday, July 21; 15:00 to 17:00

Presentations:

- Invited presentation: *Population weights in European Social Survey (ESS)* - Vasja Vehovar, Maja Mrzel, Katja Lozar Manfreda (Slovenia)

Contributed presentations:

- *How to correct for nonresponse bias in election research* - Robert Voogt (Netherlands), Willem Saris (Spain)

- *Correction for nonresponse in cross-national studies: The effectiveness of weighting for background variables* - Bart Meuleman, Jaak Billiet (Belgium)

- *Modelling attrition in panel data: the effectiveness of weighting* - Annelies Debels, Leen Vandecasteele (Belgium)

- *How to correct for survey nonresponse in a single step* - Barry Schouten (Netherlands)

2.8.1   Population weights in European Social Survey (ESS)

Vasja Vehovar; University of Ljublana, Slovenia

Maja Mrzel; University of Ljublana, Slovenia

Katja Lozar Manfreda; University of Ljublana, Slovenia

Abstract.   In this paper, the European Social Survey, first round with 22 countries, was closely observed for the effect of the basic population weighting. The sampling strategy was based on probability samples of the eligible residential populations aged 15+ years (face-to-face interview). The paper first summarizes results of post-stratification weighting by gender and age done by J. Billet and B. Meuleman (2004). Next, it continues with a systematic post-stratification for gender, age and education. The topics (substantial survey questions) that create more (unweighted) biases are systematically detected. The relative bias is also compared to the sampling error and significance levels are observed. For Slovenia, the cluster weights and designs effects are also calculated. The findings are not surprising: in a survey with decent response rates (in general much above 50%) the effect of these weights are negligible for large majority of variables. However, there are some topics and there are some countries, which have larger share of variables where simple age-gender-education weights do produce significantly different results. What is even more important is the fact that the overall weights in ESS do produce considerable increase in sampling variance. The confidence intervals are thus much broader than they seem.
   KEY WORDS: design effect / bias/ weighting / ESS /sample

2.8.2   How to correct for nonresponse bias in election research

Robert Voogt; Ministry Of Social Affairs and Employment, Netherlands

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract.   A wide variety of methods have been developed to correct for nonresponse bias in survey research. In this paper, a number of easily applicable correction methods is evaluated, using data gathered from the 1998 Dutch national elections. Weighting, extrapolation, hot deck imputation and multiple imputation are used to correct for bias in election research. The performance of these correction methods is compared for different response levels and for the inclusion of only background variables or of both background variables and two so-called central question variables in the estimation process. The results show that correction methods that use the central question variables perform much better than the methods that only use the background variables. All correction methods perform better when the response level is higher, which indicates that maximizing response does lead to substantially better estimates. The estimates obtained by the different correction methods only differ slightly when the same information is used in the calculation of the corrected estimates. Although correcting for nonresponse bias does lead to much better estimates of the true values of variables, there is barely any improvement in the estimates of relations between variables.

2.8.3   Correction for nonresponse in cross-national studies: The effectiveness of weighting for background variables

Bart Meuleman; University of Leuven, Belgium

Jaak Billiet; University of Leuven, Belgium

Abstract.   Since the size of non-response bias varies across countries, this source of bias poses an important threat to the validity of cross-national comparisons. Different approaches have been proposed to tackle this problem. In this paper, a small simulation study is performed to assess the effectiveness of one method, namely weighting with respect to background variables. European Social Survey data (round 1) for five countries –Austria, Belgium, Switzerland, Finland and Italy –is used.
   Two sets of weighting coefficients are calculated: with respect to gender and age, and with respect to gender, age and education. The latter is obtained through an iterative procedure, namely iterative proportional fitting. Frequency distributions and means of items that measure attitudes towards migration are compared before and after weighting. Weighting for gender and age hardly produces any effect. Weighting for gender, age and education seems somewhat more effective. However, sample estimates of the voter turn-outs of the last national elections

suggest that non-response bias remains present to a large extent, even after weighting for gender, age and education. It is concluded that weighting for background variables is not a proper method to eliminate non-response bias in cross-national surveys.

2.8.4 Modelling attrition in panel data: the effectiveness of weighting

Annelies Debels; University of Leuven, Belgium

Leen Vandecasteele; University of Leuven, Belgium

Abstract. Although panel studies are rich sources for studying social life, they often suffer from substantial attrition among participants. This attrition can lead to biased results if it does not happen at random. Weighting can overcome these problems, on condition that data are missing at random within the categories of the variables used for weighting. This paper argues that the choice of variables for the construction of weights is therefore crucial. This argument is supported by an analysis of the effectiveness of weighting in the European Community Household Panel. The focus is on three variables of interest, namely social class, main activity status and education, but the methodology can be applied to all possible variables of interest.

After elaborating on some fundamental determinants and mechanisms of dropout and the major variables used in weighting procedures, we focus on the effect of social class, main activity status and education on attrition. These effects are investigated by means of discrete-time logistic hazard models. We find that social class, main activity status as well as education have significant effects on dropout, even though the direction of these effects differs across countries. Next, this paper tests whether dropout is random within the categories of a number of relevant variables used in longitudinal weighting procedures. This is accomplished by examining whether the effects of social class, main activity status or education on dropout continue to exist under the control of weighting covariates used in the ECHP base weight, again using discrete-time logistic hazard models. The effects of the three variables remain significant in most countries. We conclude that weighting is not effective in these countries in the sense that dropout stays non-ignorable, even when the most commonly used weighting variables are included. Hence it is recommended to panel researchers in general to conduct a similar effectiveness analysis in order to highlight possible problems related to non-random dropout.

2.8.5 How to correct for survey nonresponse in a single step

Barry Schouten; Statistics Netherlands (CBS), Netherlands

Abstract.   Unit nonresponse to surveys may distort estimators if on average the answers by respondents and nonrespondents are different. Statistical agencies often resort to nonresponse adjustment methods to remove possible bias in their estimates for population parameters.

Nonresponse adjustment methods make use of covariates that are available for both respondents and nonrespondents. Incorporation of all available covariates, however, usually leads to estimators with large variances, so that a selection is needed. The adjustment is most effective in case we select covariates that relate both to the key survey questions and to the response behaviour. Therefore, often the process of selection is performed in two steps. In the first step a set of covariates is selected that explains response behavior. In the second step this set is reduced by accounting for the relation with the important survey questions.

We present a classification tree method that allows for the construction of strata that simultaneously account for the relation between response behaviour, survey questions and covariates. The method creates a tree in which the leaves can be used as strata in the adjustment for nonresponse.

The classification tree uses a splitting rule that is based on the maximal absolute bias under the worst case scenario, i.e. we make no explicit assumptions about the missing data mechanism. However, we assume that correlations are not affected by the nonresponse.

We apply the classification tree method to survey data from Statistics Netherlands

2.9   Surveys using new ICTs in substantive research

Session chair:
    Katja Lozar Manfreda; University of Ljublana, Slovenia

This session takes place on:
    Friday, July 22; 9:00 to 11:00
    Friday, July 22; 11:30 to 13:30

Presentations:

- Invited presentation: *Surveys using new ICTs in substantive research* - Katja Lozar Manfreda (Slovenia)

Contributed presentations:

- *Web survey design for predicting performance using network questions* - Luis Coromina (Spain)

- *How Internet Surveys Are Changing Data Collection Practices: The Case of University Student Surveys and Related Populations in the United States* - Don A. Dillman, Thom Allen (United States)

- *Online student feedback surveys. Methodological issues in comparison to the traditional classroom survey mode* - Marek Fuchs (Germany)

- *Monitoring quality of life in small and medium sized cities –results of online-survey research* - Reinhard Aehnelt, Manfred Kühn, Inga Schütte (Germany)

- *Sampling Problems in Web Surveys* - Jan B. Steffensen (Denmark)

- *Web surveys: inference using weighting and imputation in the survey on graduates* - Silvia Biffignandi, Enrico Fabrizi, Monica Pratesi, Nicola Salvati (Italy)

- *Expert Appraisals of Business Web Survey Applications* - Gustav Haraldsen (Norway)

- *Mobile phones and their impacts on survey data* - Sara Pasadas del Amo, Manuel Trujillo Carmona (Spain)

- *Nonresponse segments in Internet and mobile phone surveys* - Vasja Vehovar, Eva Belak, Darja Lavtar (Slovenia)

2.9.1 Surveys using new ICTs in substantive research

Katja Lozar Manfreda; University of Ljublana, Slovenia

Abstract. New information-communication technologies (ICTs) bring several challenges in survey research methodology. Especially, Internet and lately also mobile phone surveys have considerably changed survey industry. New types of survey designs, functions and uses have evolved. In addition, methodological issues, such as sampling, coverage, non-response, and also measurement gain new aspects. In the introductory talk we are going to present the uses of new ICTs for survey data collection. In addition, we are going to highlight main methodological problems and solutions that are used in practice (such as access panels, weighting adjustments, methods for non-response conversion, principles of questionnaire design, incentives, etc.).

2.9.2 Web survey design for predicting performance using network questions

Luis Coromina; University of Girona, Spain

Abstract. This paper describes the design and evaluation of a web survey done for predicting performance for PhD students and their supervisors at university. The attractiveness of the example is that the web survey includes many kinds of questions, such as social network questions, socio-demographic questions, questions about publications, environment in the research group and attitudes towards work. Some of these questions were sensitive and complex to answer. We give a general overview about web survey design, advantages and inconveniencies over classics surveys (personal, telephone and mail) and afterwards we show our survey procedure. Some tips will also be suggested in order to improve response rates. Finally, the results for our web survey show that respondents are satisfied with the length, layout, difficulty, susceptibility and distribution of the questions. Moreover, we can differentiate between PhD students and supervisors in order to know which web survey characteristics we should take more in consideration when we are asking to junior or senior researchers.

2.9.3 How Internet Surveys Are Changing Data Collection Practices: The Case of University Student Surveys and Related Populations in the United States

Don A. Dillman; Washington State University, United States

Thom Allen; Washington State University, United States

Abstract. Each of the traditional modes of data collection, whether face-to-face, telephone, or mail—has a rhythm and character that sets it apart from the others. The selection of mode has a significant effect on how samples are drawn, how individual questions get worded, and how items are ordered to form questionnaires. These differences often influence how people provide answers to individual questions, and even whether they respond at all to a particular survey.

It would be surprising if the introduction of surveys by the Internet, did not also bring with it many new, often unintended, data collection practices that set it apart from other modes of surveying and which also influence data quality. Our purpose in this presentation is to identify a number of significant changes in data collection practices that are consequences of surveys on the Internet that are based upon periodic surveys of random samples of students at Washington State University conducted by the authors. These surveys are designed to provide useful longitudinal information to University leadership for tracking how students are responding to the education being provided by our University.

In this paper we discuss a wide variety of coverage, measurement and non-response issues that require us to approach the web surveying process differently than when we relied only on telephone and/or mail systems. The changes include coverage issues, e.g. not being able to get a single list of postal, telephone, or web addresses, because of changes in what address information students are required to provide to the university and the change from mandatory updating to making it voluntary as to which address is available for university use. Another change we discuss is the benefit of using contacts by an alternative mode (mail) to deliver meaningful incentives in order to achieve high web survey response rates. Based upon extensive experimentation which shows that differences in the visual layout of questionnaires influence how people respond to web survey questions we discuss the use of procedures such as Cascading style sheet construction and other measures in order to control the visual aspect of questions seen by respondents in an effort to provide the same stimulus to each respondent. Another critical change has to do with the development of monitoring systems to quickly analyze and respond to respondent

problems and concerns that are critical in a web context for converting likely refusals to completes. These are examples of the topics to be discussed.

The learnings from these periodic surveys have provided a basis for developing a web survey system now used by the Social and Economic Sciences Research Center that we now employ in a general way for doing web surveys of many other populations. In the proposed paper we will describe in detail both the problems and solutions associated with maintaining high quality survey data collection processes, making specific reference to how the web has required us to change long-standing data collection practices that previously supported our telephone and mail data collection systems.

2.9.4   Online student feedback surveys. Methodological issues in comparison to the traditional classroom survey mode

Marek Fuchs; University of Kassel, Germany

Abstract.   Students are asked on a regular basis to evaluate the lectures and seminars that they are attending. Until recently this has been done by paper and pencil surveys in a classroom setting. In recent years some systems became available to approach student using web surveys. Since this methodology has various advantages (in terms of cost and time, data protection), some methodological concerns have been raised:

(1) Over-coverage and under-coverage: In order to be included into the study, all students have to be registered with a valid e-mail address. Since almost every student has an e-mail address nowadays this no longer seems to be an issue. However, there is a small population who abandons the Internet or are not willing to provide their e-mail address. These populations cannot be reached using web surveys. By contrast, using web survey methodology we can address also students who abandoned the class or course in the course of the semester. As a result, the two survey modes address overlapping, however, slightly different populations which has consequences for the results.

(2) Non-response: Experiences from several waves of online student feedback web surveys suggests that non-response is of great concern. Compared to a class room setting where social pressure leads to almost 100 percent cooperation, students feel less obliged to participate in web surveys. Several strategies in reducing non-response for the web survey condition were tested.

(3) Measurement: Completing a paper form in class in the presence of the teacher implies a completely different social setting compared to responding to a web survey at home, in the dorm or at some other place. Thus, concerns were raised, that measurement bias related to the mode of administration should be of great importance. However, experimental research (split half design) suggests, that only minor differences of the data occur.

The paper summarizes this line of methodological research for student feedback and course evaluation web surveys based on a series of four studies conducted over the past several years.

2.9.5   Monitoring quality of life in small and medium sized cities –results of online-survey research

Reinhard Aehnelt; Advis, Germany

Manfred Kühn; Institute for Regional Development and Structural Planning, Germany

Inga Schütte; Institute for Regional Development and Structural Planning, Germany

Abstract.   In the context of the research project 'The Quality of Life in Small and Medium-sized Cities' sponsored by the German Ministry of Education and Research (2002-2005) the network of seven Cities in Brandenburg (Germany) has developed an monitoring system regarding the quality of life. An essential aim was the assessment of objective living conditions and subjective well-being. An objective data-set expanded by a survey about subjective well-being of the citizens. Because the monitoring system should be operated by the cities themselves after the funding for the project has run out, an online questionnaire should deliver the level of satisfaction enjoyed by the city population.

A main target was to test the suitability of this instrument especially in small and medium-sized cities. Therefore the number of completed questionnaires, the extent of selection of the answering group as well as the quality of the answers were tested by a parallel conducted telephone survey. In the two years the participation was very stable (2003: 1.324; 2004: 1291), although levels of participation in one city varied from year to year.

The opinions of residents were very similar concerning the results of the telephone- and the online-based survey. The differences lay in more critical replies corresponding to the social structure of the participants. The

online users tend to be younger, better educated and employed. Furthermore, more men took part in the online-based questionnaire.

Fundamentally, online-based questioning is a suitable method for collecting subjective data especially in medium-sized cities. In small towns there is a problem of inadequate participation. The instrument is useful for showing trends, but not for conducting a representative study.

2.9.6   Sampling Problems in Web Surveys

Jan B. Steffensen; Biblioteksskolen, Denmark

Abstract. The fast-spreading use of web surveys in social surveying accentuates the methodological cornerstones related to sampling and user identification. While the obvious advantages of web survey methodology –cost reduction, automation of data collection, accuracy, validation, timeliness –in first hand are of a practical nature, the problems are of greater consequence to more basic issues of survey research, such as sampling, representative samples and respondent identification. The non-probabilistic convenience sampling of many web surveys traditionally poses serious problems to the power of generalization and prediction of the survey, but do these fundamental problems also indicate actual problems in concrete library user surveys.

This paper presents the results of a study comparing various traditional sampling methods to web sampling. The study was carried out at The Regional Library of Northern Jutland in Aalborg, Denmark, during spring 2005. Three samples of library users were created as a web sample drawn by self-selection from the library web site, a convenience sample drawn among visiting users over two weeks, and a random sample drawn from the library registration of users. Respondents from all three samples were presented to the same questionnaire, and the composition of the samples as well as the results and predictions were compared in order to determine and analyze the effect of sampling methods.

During the web survey an automated collection of data on respondent behaviour was conducted in addition to the traditional data collection. The analysis of these data gives an insight into how respondents navigate through the questionnaire as well as detailed information on unit and item non-response. The data are also used to identity repetitious replies or "ballot stuffing"through identification of respondents. The paper presents an analysis of the problems connected to respondent identification and repetitious replies

2.9.7   Web surveys: inference using weighting and imputation in the survey on graduates

Silvia Biffignandi; University of Bergamo, Italy

Enrico Fabrizi; University of Bergamo, Italy

Monica Pratesi; University of Pisa, Italy

Nicola Salvati; University of Pisa, Italy

Abstract. If a Web survey allows respondents to self select into the survey, it is necessary to adjust for selection bias. In our paper we handle the statistical problem working on empirical data coming from a Web survey on graduates of an University; the task of the survey is their enrolment in the labour market.

We investigated whether it is possible to adjust for selection bias using propensity scores. The target population is represented by a statistical register maintained by the university administration: for each graduate there is an administrative record where each event in the student career - from the university enrolment to the degree or to the abandon - is registered. This register is used to estimate the propensities, in other words to find out the variables that capture the difference between the Web respondents and the general population. In the contribution a propensity based estimator of the mean and an estimator of its variance are proposed. For few study variables we have tried also to impute the response for not web respondents using the donor pool individuated by the propensity matching. In the contribution we examine the sensitivity of the results to alternative specifications of the method and of the model for the imputation, we refer to the choice of variables for the propensity score model and the choice of matching method (number of digits in the matching, radius or nearest neighbour matching).

2.9.8   Expert Appraisals of Business Web Survey Applications

Gustav Haraldsen; Statistics Norway, Norway

Abstract.   Expert appraisals are a well known technique used to evaluate comprehension and task problems in questionnaires. Used for this purpose a classification system based on the cognitive steps that a respondent normally runs through is applied. The classification system names well known problems with the wording, tasks or response categories in questions. It is recommended that at least three censors first make their individual evaluations and then come together to work out joint conclusions. We have constructed a different classification system, but have nevertheless used a similar technique to evaluate two business web survey applications.

Our expert appraisals are a part of a project guided by a conceptual model that links together the instrument and data collection procedure developed and presented to the respondent, the characteristics of the respondent and his environment, and the outcome of the response process. At present we test various methods of measuring the different elements of this model separately, but eventually the plan is of course also to analyze the relationships between different parts of the model. The expert appraisal procedure is used as one way of evaluating the quality of the first element of the model, namely the data collection instrument in web surveys.

In Norway two main portals are presently used to collect information from companies. First it is the System for Information and Data Exchange with Businesses (Idun), which is a portal developed for statistical questionnaires by Statistics Norway. Secondly it is the Altinn portal that has been developed by the Tax Authorities, The Brønnøysund Registers and Statistics Norway, with the ambition to offer one common web portal to all official institutions that collect data from companies or private persons. There are both similarities and differences in the structure and design of these two applications. For about a year ago a current best practice publication named "Advise for development and design of web questionnaires"was written. The content was partly based on tests run during the development of Idun and Altinn, but first of all based on international research. We have used the recommendations given in this publication to classify possible problems in web survey applications. Then a few expert appraisals of Idun and Altinn questionnaires were carried out with the help of this classification system. We tried to run the evaluations according to the procedure recommended in classical expert appraisals.

In our presentation we first want to introduce the classification system itself, because it mirrors what we think is the current best way of designing business web survey applications. Secondly we would like to show some the results from the appraisals that we think are interesting. And finally we want to discuss if it is this or other methods that should be used to evaluate the design element in our conceptual model.

2.9.9   Mobile phones and their impacts on survey data

Sara Pasadas del Amo; Institute for Social Studies of Andalusia, Spain

Manuel Trujillo Carmona; Spanish Council for Scientific Research (IESA/CSIC), Spain

Abstract.   The produced technological changes and the rapid adoption of certain emergent technologies on the part of the population, such as the mobile telephone, have strongly impacted communicative behaviour and uses towards this end. As such, in some Spanish households, cell phones have replaced land lines as the principal long distance communication device. These households present a specific profile that differentiates them from those that have both technologies at their disposal, and in greater measure from those that only have a land line. The most important consequence of this process for research based on telephone surveys is that, although telephone coverage has not ceased to enlarge until practically becoming universal, more are the homes each day that are excluded from the sampling frame of the surveys for not having a land line. With this paper we propose to analyse the importance of this phenomenon and to evaluate the impact on different kinds of survey data.

2.9.10   Nonresponse segments in Internet and mobile phone surveys

Vasja Vehovar; University of Ljublana, Slovenia

Eva Belak; Statistics Slovenia, Slovenia

Darja Lavtar; University of Ljublana, Slovenia

Abstract.   After the rise of the Internet, rapid expansion of mobile phone usage creates another challenge to the survey industry. As for now, the potential impact still remains more or less latent. However, with further changes in prices and in the share of mobile-only persons, we can expect considerable changes in telephone survey practice.

Paper addresses the issue of noncoverage and nonresponse patterns of these surveys. The aspect, which is strongly connected to coverage and noncoverage, is the actual availability. Persons may have internet access or not, but with mobile device it can be most of the time turned off. When defining the noncoverage and nonresponse in internet and mobile phone surveys, these issues have to be carefully considered. Further specifics are arising from the patterns of usage as well as the corresponding attitudes and life-styles. The main purpose of the paper is thus to identify the segments that have specific noncoverage properties and nonresponse behavior.

The empirical data are based on Slovenian Labor Force Survey 2002-2004, where the fixed and mobile phone usage is systematically monitored. There, the apparent stagnation in the share of mobile-only households that appeared in 2004 (after growth in 2003), is analyzed. Next, the experience with follow-up telephone surveys for later waves of LFS is studied: the interviews and persons that revealed the mobile phone and compared with the ones that revealed the fixed telephone.

Another specialized face-to-face survey with high response rate (80%, n=1,500) was also carried out in April 2005, which studied the mobile phone usage. The propensity of various segments to participate in a mobile phone survey was empirically elaborated. The same propensity for participation in Web survey was also analyzed.

2.10   Computer assisted data collection

Session chair:
> Jelke Bethlehem; Statistics Netherlands (CBS), Netherlands

This session takes place on:
> Friday, July 22; 15:00 to 17:00

Presentations:

- Invited presentation: *Computer Assisted Interviewing, History and Development* - Jelke Bethlehem (Netherlands)

Contributed presentations:

- *A strategy for guaranteeing high level of data quality in CATI surveys* - Manuela Murgia, Stefania Macchia (Italy)

- *Non-observation errors in Computer Assisted Telephone Interviewing* - Fannie Cobben (Netherlands)

- *Mode effects and maintenance of data quality in a CAPI/CATI survey* - Tim Burrell (United Kingdom)

- *A comparison of CAPI and PAPI in a nationally representative Danish Health Survey* - Ola Ekholm, Ulrik Hesse, Jeanette Nørlev, Michael Davidsen (Denmark)

2.10.1   Computer Assisted Interviewing, History and Development

Jelke Bethlehem; Statistics Netherlands (CBS), Netherlands

Abstract. Survey data collection has undergone substantial changes in the last 30 years or so. The rapid advent of the computer has made it possible to replace paper questionnaire forms by electronic ones. The change to computer assisted interviewing (CAI) started in 1970's with the emergence of Computer Assisted Telephone Interviewing (CATI) in the United States. In the 1980's, we witnessed the rise of Computer Assisted Personal Interviewing (CAPI) in Europe. Also in the 1980's, the first applications of Computer Assisted Self Interviewing (CASI) were developed.

Although CAI has some clear advantages (higher quality data, and improved timeliness of survey results), its introduction took a while. First, careful designed experiments had to show the new technology worked. Currently, CAI is a well-established mode of data collection. Powerful standard software assists survey researchers in developing effective electronic data collection instruments. However, there are also new challenges. Testing and documenting large electronic questionnaires (with sometimes thousands of questions) with complex skip patterns becomes ever more difficult.

Survey researchers are always confronted with nonresponse in their surveys. This may result in biased estimates. Mixed-mode surveys may help to reduce nonresponse. CAI systems can be used to implement mixed-mode surveys. However, to avoid mode effects as much as possible, much attention has to be paid to proper implementation of these systems.

The last couple of years a new means of computer assisted interviewing is growing in populairity, and that is computer assisted web interviewing (CAWI). It can be seen as a form of CASI. Currently, web surveys suffer from severe methodological problems, survey estimates are subject to substantial non-sampling errors. Nevertheless, there seems to be a bright future for this type of data collection, not only in single surveys, but also in the form of web panels.

2.10.2   A strategy for guaranteeing high level of data quality in CATI surveys

Manuela Murgia; ISTAT, Italy

Stefania Macchia; ISTAT, Italy

Abstract.   CATI technique is quite used in Istat. For the majority of surveys adopting this data collection mode, the usual strategy has been to design the questionnaire in Istat and to charge private companies with the development of software procedures for the management of both data capturing and monitoring phase.

This organisation has always succeeded in concluding the data capturing phase, but often presented problems, in terms of data quality, monitoring and timely delivery of data. These problems were generally due to the fact that private companies are often very experienced in telemarketing or opinion polls, but:

●had never faced in advance the development of electronic questionnaires so complicated in terms of skipping and consistency rules between variables;

●had rarely put in practice strategies to prevent and reduce non response errors;

●had not at their disposal a robust set of indicators or ad hoc procedures to monitor the interviewing phase.

That is why a new strategy, called 'in-house strategy for CATI surveys', was tested for a very important survey (the Birth Sample Survey) in 2002 and, after its success, repeated for other five surveys hold in 2004 or planned in 2005. It consists in relying on a private company only for the fieldwork phase (call centre plus interviewing staff) and in developing in Istat all the software procedures to manage:

●the data capturing phase: the electronic questionnaire, the calls scheduler and delivery of data;

●to monitoring phase: traditional reports to monitor the trend of telephone contacts and innovative tools for peculiar aspects of CATI surveys.

The procedure integrates different software packages, but the core is developed with the Blaise system (produced by Statistics Netherlands and already used by a lot of National Statistics Administrations for data capturing carried out with different techniques).

Following this strategy, all the software system is planned having the quality as the main purpose to fulfil. According to this, the 'in-house strategy' defines quality standards for the construction of the electronic questionnaire and, at the same time, offers generalised software procedures to take into account the peculiarities of each CATI survey.

The quality standards, defined for the data capturing phase, are aimed at preventing from some of the non-sampling errors, like partial non response, total non response and non-consistent response. They cover different aspects of the construction of the electronic questionnaire and of the other functions to be managed in the data capturing phase. They focus on different purposes, related to:

●the layout of the electronic questionnaire, to reduce or eliminate the 'segmentation effect';

●the customisation of questions' wording, to make the interview more friendly and questions easy to be answered;

●the management of errors, to prevent from all the possible type of errors without increasing the respondent burden and making the interviewer' job easier;

●the control of data with information from previous surveys or administrative archives, to improve the quality of the collected data;

●the assisted coding of textual answers, to improve the coding results and to speed up the coding process;

●the scheduling of contacts, to enhance the interviewers' productivity and to avoid distortion on the probability of respondents to be contacted.

The monitoring phase of the 'in-house strategy' is based on both traditional and innovative tools. The first ones are excel reports representing n-ways contingency tables and the latter are based on control charts that are a methodological tool for statistical quality control, traditionally applied in manufacturing process control. The two types of monitoring tools are used to monitor different aspects of the interviewers' job:

●excel reports monitor the telephone contacts variable, in order to keep under control the interviewers' productivity and the presence of odd behaviours in assigning contact results (e.g. an excessive use of appointments in order to reduce the refusal rate, if, for instance, a too high refusal rate causes some penalties for the interviewer);

●control charts monitor peculiar aspects for the survey, like the assisted coding of textual variable (if used). They are, therefore, dedicated to monitor complex variables, treated in the questionnaire, that require the analysis of a set of various indicators. Till now, in the 'in-house strategy', control charts have been used to monitor the assisted coding of Occupation, observing two aspects of the coding process: the quality of coding activity and the time necessary to assign a code.

The paper will describe in details the already mentioned standards and generalised tools and will report the main CATI results obtained in the surveys which used this new strategy, comparing them with those currently obtained in Istat for the other CATI surveys.

This analysis lead us to state that the quality standards and monitoring procedures of the 'in-house strategy' guarantee the efficiency of data capturing process as a whole and therefore a high level of data quality

2.10.3   Non-observation errors in Computer Assisted Telephone Interviewing

Fannie Cobben; Statistics Netherlands (CBS), Netherlands

Abstract. Individuals that are selected in a sample but that do not participate are referred to as nonrespondents. That nonresponse can cause statistics to be biased is a well known fact. In this paper we investigate an other non-observation error: the bias due to noncoverage. Noncoverage is the situation in which the sample is not representative of the population of interest because there are certain groups in this population that are not covered by the survey. If the noncovered groups respond systematically different to the questions in the survey, the statistics are biased due to the noncoverage. Noncoverage occurs even before the survey has been allocated to the interviewers and is therefore invisible in most datasets.

In this paper, we analyse the noncoverage bias in CATI-surveys. This bias is induced by the restriction of using the telephone as a communication medium. Obviously, only that part of the population that disposes of a listed, fixed-line telephone can be interviewed by means of CATI.. We use data from a CAPI-survey, the Integrated Survey of Living Conditions, to assess the noncoverage bias when restricting the sample to individuals that own a listed, fixed-line telephone. We compare individuals with and without a listed, fixed-line number. Two methods to adjust for the noncoverage bias are applied to the data from the Integrated Survey of Living Conditions: linear weighting and propensity score stratification. Finally, we analyse a strategy to simultaneously adjust for nonresponse and noncoverage bias.

2.10.4   Mode effects and maintenance of data quality in a CAPI/CATI survey

Tim Burrell; Office of National Statistics, United Kingdom

Abstract.   The Annual Population Survey (APS) is a ONS social survey launched in January 2004. This survey covers a sub-set of the topics in the ONS Labour Force Survey (LFS) and the 2 samples will be joined to provide improved local area data.

The LFS is a rotating panel survey containing 5 quarterly waves. The first wave is always CAPI, since the UK does not have an address register with phone numbers, follow-up waves are carried out via CATI. The majority of LFS interviews are via CATI and the structure of the interview is set up to best suit this method. The APS also has a panel element. The majority of first wave interviews are administered via CAPI. However, where a telephone number can be traced (about 25% of cases) first wave interviews are carried out via CATI. All second and subsequent interviews are carried out via CATI. The LFS caters for both CAPI and CATI within a single Blaise instrument designed at the lowest common denominator, most suitable for CATI. Other ONS social surveys have a different (ONS) standard single instrument structure which is more suited to CAPI. For the APS, we retained a single questionnaire instrument which is flexible enough to get the best out of both modes of interviewing and aims to maintain data quality.

This paper will describe the two different methods of interviewing, and steps that have been taken to assess the success of the new questionnaire. It also gives an opportunity to look at any differences arising in how the questionnaire structure and mode of interview impact on the length of interview (using audit trails) and data quality.

Tim Burrell is a Senior Social Survey Researcher in the Social and Vital Statistics Division (SVS) of the Office for National Statistics.

2.10.5   A comparison of CAPI and PAPI in a nationally representative Danish Health Survey

Ola Ekholm; National Institute of Public Health, Denmark

Ulrik Hesse; National Institute of Public Health, Denmark

Jeanette Nørlev; National Institute of Public Health, Denmark

Michael Davidsen; National Institute of Public Health, Denmark

Abstract.   Introduction: In the last decades computers have been introduced and used in all phases of data collection. One aspect of this is the conversion from Paper-And-Pencil Interview (PAPI) to Computer-Assisted Personal Interview (CAPI). An important reason for a conversion is the reduction in the time elapsed between the data collection and the availability of the data for analysis. Several studies have investigated the effect of CAPI on response rates, the interview durations and the responses to questions, but to our knowledge, no study have investigated the effect according to gender and age. Furthermore, few studies exist regarding health. Thus, the purpose of this study is to investigate the differences between the two modes with regard to response rates, interview durations and responses to questions concerning health, lifestyle and morbidity.

Methods: Data were derived from a Health Interview Survey in Denmark 2003. Four random samples of each 1,000 Danish citizens (age 18 or more) were drawn from the Danish population. One of the samples was interviewed using PAPI and the other three samples were interviewed using CAPI. The interviews were carried out in the respondents' home and were performed by a trained interviewer. All samples included identical questions on health, morbidity and lifestyle.

Results: The response rate in the CAPI group was 64,4% and in the PAPI group it was 67,5%. The differences in response rates were most obvious among women and elderly (67 years or older). The average interview duration was significantly ($p < 0.05$) shorter among CAPI cases than among PAPI cases (mean: 58.5 and 60.3 minutes, respectively). Differences in responses to selected questions were found in questions that could be apprehended as sensitive (questions concerning alcohol habits, weight and height).

Discussion: The results from this study indicate that women and elderly are less willing to participate when the data collection method is CAPI. These findings might be understood as an outcome of different habits and familiarity with computers. However, those who actually participated in the study seemed to apprehend the CAPI mode as more confident and anonymous than the PAPI mode, this making them more willing to report "socially improper behaviour".

3   Quality of Measurement (Aula 006)

3.1 Measurement of satisfaction

Session chair:
     Luigi Fabbris; University of Padua, Italy

This session takes place on:
     Tuesday, July 19; 9:00 to 11:00
     Tuesday, July 19; 11:30 to 13:30

Presentations:

- Invited presentation: *A Tournament System of Questions for the Collection of Customer Satisfaction Data through a CATI System* - Luigi Fabbris (Italy)

Contributed presentations:

- *Bias in life satisfaction/dissatisfaction ratings due to the response format* - Mehrdad Mazaheri (Iran), Peter Theuns, Jan L. Bernheim (Belgium)

- *The Pursuit of Happiness* - Timothy Bechtel (United Kingdom)

- *Wellbeing survey results in Algeria: Stability or change?* - Habib Tiliouine (Algeria)

- *Customers' loyalty as a measure of graduates satisfaction* - Maria Cristiana Martini (Italy)

- *Use Of Rasch Analysis To Classify Work-Related Low Back Pain* - D.F. Iezzi, F. Giannandrea (Italy)

- *Level and Inequality of Happiness in Nations* - Jan Ott (Netherlands)

- *Determinants of the Risk of Uneasiness in Paduan Families* - Sara Poffe, Luigi Fabbris (Italy)

- *Personal and institutional factors in the assessment of European public policies: The case of education* - Clara Riba, Anna Cuxart (Spain)

- *Relationship between Teaching Quality and Customer Satisfaction at University* - Marisa Civardi, Emma Zavarrone (Italy)

3.1.1 A Tournament System of Questions for the Collection of Customer Satisfaction Data through a CATI System

Luigi Fabbris; University of Padua, Italy

Abstract. The customer satisfaction measurement is often completed by the definition of the "importance"estimate of a set of aspects of the evaluated good or service, which is considered a measure of the relationship between the responses given and the value system of respondents. In this paper, we suggest a new method of data collection, suitable for CATI systems, of the relative importance of a set of items based on a se-ries of hierarchical comparisons between couplets, or triplets, of items.

The method of couplets comparison (of an even number of items) consists of:

- the definition of the first-round comparison within each couplet defined by the researcher and the identification of the "winner"of the first tournament;

- the second-round comparison within the couplets of items that passed the first-round of comparisons. The coupling of items is defined by the researcher according to a random rationale, unless a logical order may be defined,

- the within-couplet comparison is repeated until a single item wins the top tournament.

The method may be adapted to almost any number of items, provided an adequate design is defined by the researcher. Combinations of couplets and triplets may be planned at various levels of comparison.

Our method may be juxtaposed to the methods popular in customers' satisfaction re-search, i.e.

- the methods of "importance level"measurement, with questions administered item by item,

- the method of the "direct ranking"of a set of items, with the presentation of the full set of items and the request addressed to the respondent of ordering either all of them, or just the identification of the items on the tails of the set,

- the method of "paired comparison"based on the comparison of all possible couples of items.

The properties of our method are discussed both theoretically and empirically with reference to a set of experiments on customer satisfaction for services supplied to families and companies by a company in Padua. From several viewpoints our method shows to be valuable with respect to comparable methods.

3.1.2 Bias in life satisfaction/dissatisfaction ratings due to the response format

Mehrdad Mazaheri; University of Sistan & Baluchestan, Iran

Peter Theuns; Vrije Universiteit Brussel, Belgium

Jan L. Bernheim; Vrije Universiteit Brussel, Belgium

Abstract. Aim: The current study was designed to investigate the effects of variations in scale orientation and type of response format on global life-satisfaction and global life-dissatisfaction ratings.

Method: A sample of 1022 voluntary participants, mainly students (699 females and 323 males), randomly assigned to eight groups, were asked to rate their current overall life-satisfaction and life-dissatisfaction on one of eight different formats of a rating scale, differing in type of response format (bipolar versus unipolar), anchor points (-5 to +5 versus 0 to 10), and scale orientation (vertical versus horizontal).

Results: For life satisfaction ratings, a negative skew was found for all response formats, but, a higher percentage of respondents scored in the upper part (center to top) of the scale with anchor points (-5 to +5) than on the one with anchor points (0 to 10). For life dissatisfaction, a positive skew was found for unipolar response formats and a negative skew for bipolar response formats.

Univariate Analysis of Variance (ANCOVA) indicated significant main effects of (1) anchor points (-5 to +5 versus 0 to 10) on life satisfaction and (2) type of response format (Unipolar vs Bipolar) on life dissatisfaction. Moreover, (3) the type of response format $\times$ anchor points and scale orientation $\times$ anchor points interactions were found to be significant for life dissatisfaction (not satisfaction) ratings.

A significant correlation between life satisfaction and life dissatisfaction was found in all groups, but, remarkably, the direction of the correlation was negative for unipolar and positive for the bipolar response format.

Conclusion: Our results suggest that the type of response format can bias both life satisfaction and life dissatisfaction ratings, and more specifically, the relation between these two constructs.

3.1.3   The Pursuit of Happiness

Timothy Bechtel; University of Essex, United Kingdom

Abstract.   The present paper explores the determinants of happiness and poses the following research question: is happiness driven more by economic or psychological factors? This question has been investigated in the English and west German populations using the British Household Panel Survey (BHPS) and the German Socio-Economic Panel (GSOEP). The results of this study fall into two categories:

Methodological

●In explaining happiness it is important to follow respondents longitudinally so that (inevitably) unobserved personality traits can be included as individual effects in the panel regression model. The inclusion of these effects, which is impossible in cross-sectional regressions, protects the regression effects from being biased.

●When happiness is measured on numerical rating scales, its explanatory regression coefficients are almost identical with those produced when happiness is only ordinally scaled. This result, which justifies the cardinal measurement of a subjective dependent variable, is welcome news for social scientists, who have been numerically coding ratings for decades.

●The BHPS single item indicator of overall happiness almost perfectly traces the 9-item indicator of English happiness constructed from the present BHPS data. Thus, over the five years studied here, either indicator may be used to track national happiness. The validity of a single-item indicator is more good news for social scientists, who have been calibrating happiness in this way for decades, despite the expressed reservations of many.

Substantive

●In the economic, accommodation and medical spheres of experience subjective representations of physical variables (ratings), rather than the physical scales themselves (Euros and cents, square feet, etc.), are the operational determinants of life satisfaction. That is, subjective regressors are far superior to objective regressors in explaining happiness.

●English and German overall life satisfaction follow different time trends and have different compositions. English happiness is driven more by psychological than economic factors, whereas the major component of German happiness is the materialistic item standard of living. Thus, a preliminary answer to our research question about what drives happiness appears to be an ethnographic one, i.e., it depends on the specific culture studied. The English have a psychological orientation toward happiness, in contrast to the Germans who are more economically sensitive.

●The R-square = 0.60, achieved in the present study for explaining English happiness, appears to be the highest yet recorded in the quality-of-life literature.

3.1.4   Wellbeing survey results in Algeria: Stability or change?

Habib Tiliouine; University of Oran, Algeria

Abstract.   The purpose of the paper is to raise the issue of the usefulness of subjective indicators in developing countries and to draw some conclusions from the results of the two surveys using the International Wellbeing Index (IWI), besides other measures, that have been conducted in Algeria.

This country is a very interesting case of study as a third world and mostly Arab and Islamic country. It has witnessed since the end of the 80's a very harsh economic crisis, accompanied, since 1991, with a very bloody struggle between Islamist extremists and the army. The results of the first wellbeing survey of September 2003 (N= 1417), indicate that the population's wellbeing has been very low on all the dimensions of the measure, with some significant differences between the categories of the respondents.

In the present time security problems are less frequent, the political life is stable, and the major economic indicators, partly due to oil prices boost, have shown a real increase. Our second survey which has started in the 9th of March, 2005, for four weeks, aims to explore how does the Algerian population respond to such improvements? Do subjective indicators respond positively to improvements in economic indicators in the particular case of Algeria? Some other factors will be explored with relation to subjective wellbeing measurement.

3.1.5   Customers' loyalty as a measure of graduates satisfaction

Maria Cristiana Martini; University of Padua, Italy

Abstract.   Assessing a university study program consists in comparing its results with its purposes, but the definition of the aim of a cultural and formative service such as the university is, looks far from simple, since the expected products are mainly culture and job competences, in one word human capital. The assessment of university programs, then, will change according to which we consider as the main target of academic formation, that is, transmitting knowledge and culture, or supplying a tool to enter the job market more easily. Both of these approaches have their Achilles' heel, since the university as a place for mere knowledge transmission becomes estranged from the real needs of a society, while the ideal of the university as the breeding ground where high level labour forces are formed risks to give a too pragmatic and utilitarian point of view.

This contribution attempts to apply an alternative measure of evaluation of university programs, based on customers' satisfaction. This approach lies in a loyalty measure, that is on the idea that a study program is successful if its graduates would be willing to confirm their choice, could they go back to their enrolment. Data are drawn from a survey conducted on a sample of 2818 graduates of the university of Padua. The survey is longitudinal, and graduates are contacted for the first time at the completion of their study career with a self-administered questionnaire, and then for 6 waves at 6 months intervals by CATI interviews. The respondents were asked about their hypothetical will to repeat their study program in each wave. This study aims at investigating what are the components which influence and determine the graduates' loyalty in three temporal situations: (a) at the end of the university program, when the academic experience is still vivid; (b) six months after graduation, when they are experimenting the first contacts with the labour market; (c) eighteen months after graduation, when their working position is more stable.

In order to take into account the two-level hierarchical structure of these data (first-level units, graduates, are nested in second-level units, e.g. faculties) and the dichotomous nature of the dependent variable, the analyses are performed by applying multilevel logistic models.

The model fit and the parameter estimates for the three situations here considered show clear differences as the time reference changes:

- At the end of the university studies the fidelity level only depends on different aspects of the academic experience. The difference in the fidelity level between faculties is strong, and it is mainly due to the distinction between short-term and long-term study programs, to the existence of faculties with a mandatory post-lauream training, and to the differences between faculties in the mean final grade;

- Six months after graduation the fidelity level is still influenced by the same factors, but these become less important, and a group of predictors related to the working situation enter the model. Differences between faculties are still conspicuous, even if less than at the end of the university studies;

- Eighteen months after graduation differences between faculties are not statistically significant. The opinion on colleagues and facilities do not influence the loyalty level anymore, and the other factors related to the studying career have a weaker influence. Conversely, factors referring to the job situation become more and more important, and new aspects of the graduates' working life enter the model.

In general, this fidelity-based assessment measure lies between the judgment on the university as a cultural experience and a more pragmatic opinion on the graduation as a tool to penetrate the job market. In fact, right after graduation this evaluation largely coincides with the academic experience, while, as time goes on, considerations on the problems which are encountered in finding a job, as far as the kind and quality of the found job and its opportunities in terms of use of the acquired competencies and adherence to the studied topics, enter the determination of the assessment; this makes the proposed evaluation measure a sort of weighted average of different aspects, where the weights are decided each time by each graduate.

Moreover, this measure is highly affected by faculty differences only at the end of the studies, when individual expectations and ambitions are strongly dependent on the academic experience, but the first contacts with the labour market smooth these differences and make the faculties comparable with respect to the index of customers' satisfaction.

3.1.6   Use Of Rasch Analysis To Classify Work-Related Low Back Pain

D.F. Iezzi; University of Rome 'La Sapienza', Italy

F. Giannandrea; National Institute for Occupational Safety and Prevention (ISPESL), Italy

Abstract. Low back pain (LBP) represents one of the most frequent health problems in several working populations. Epidemiological studies have provided ample evidence that physical work factors like high physical load,

manual material handling and patient handling are associated with the occurrence of LBP. In several studies perceived disability associated with LBP has been collected by different questionnaires and a variety of scales. The most popular questionnaire is Oswestry Disability Questionnaire (ODQ).

The aim of this study is to use Rasch measurement to assess and modify the original classification categories of the ODQ, to examine the hypothesis that the items from the ODQ form a unidimensional construct and a hierarchical representation of LBP disability, and to compare ordinal Likert resultant scores to interval Rasch scaled scores with disability categories serving a framework.

In pre-processing step we translate and culturally adapt the Italian version of the ODQ, to validate its use for assessing low back disability perception in nursing personnel. Hospital workers, particularly ward nurses, are known to be at high risk for back pain, with patient-handling tasks being implicated in most cases.

This study included 132 healthcare workers (62 male and 70 female) who completed ODQ. These workers were classified into three clinical classes: 1) no symptoms (60 %), 2) LBP (28 %) 3) LBP with diagnosed lumbar disk hernia or spondylolisthesis (12 %).

A hierarchical representation of LBP disability was supported. A comparison of the disability categories based on Likert and Rasch scaling revealed them to be non equivalent. The new scaling changes the disability categories and highlights that the items more representative are lifting, low back pain intensity and walking.

A classification tool was developed to categorize different lumbar disability classes.

References:

Fairbank J, Couper J, Davies J, et al. (1980). The Oswestry low back pain questionnaire. Physiotherapy; 66:271-3

Fairbank J, Pynsent PB. (2000). The Oswestry Disability Index. Spine; 25:2940-2953

Giannandrea F., Iezzi D.F. (2004) "A method to classify Hospital Workers with lower back pain", Conference IFCS (International Federation Classification Society), Chicago, Illinois Institute of Technology, USA, July 14-18: 101.

Iezzi D.F., Giannandrea F., D'Onofrio A., Chirico F., Marini Bettolo P. (2004). "Un algoritmo per classificare la percezione del dolore lombare in un gruppo di lavoratori ospedalieri". G. Ital. Med. Lav. Erg. 26; 4: 261-262.

Lagerstrom M, Hansson T, Hagberg M.(1998). Work-related low-back problems in nursing. Scand J Work Environ Health; 24: 449-64.

Ozguler A, Guéguen A, et al. (2002). Using the Dallas Pain Questionnaire to Classify Individuals With Low Back Pain in a Working Population. Spine; 27:1783-1789.

3.1.7 Level and Inequality of Happiness in Nations

Jan Ott; Erasmus University, Netherlands

Abstract. Utilitarians and egalitarians have different priorities. Utilitarians prioritise the greatest level of happiness in society and are prepared to accept inequality, while egalitarians prioritise the smallest differences and are willing to accept a loss of happiness for this purpose. In theory these moral tenets conflict, but do they really clash in practice?

This question is answered in two steps. First we consider the relation between level and inequality of happiness in 80 nations; level of happiness is measured using average responses to a survey question on life-satisfaction and inequality is measured with the standard deviation. There appears to be a strong negative correlation; in nations where average happiness is high, the standard deviation tends to be low. This indicates harmony instead of tension. Secondly we consider the institutional factors that are likely to affect happiness. It appears that level and equality of happiness depend largely on the same institutional context, which is another indication for harmony. We may conclude that the discussion between utilitarians and egalitarians is rather academic and of little practical importance.

There is one issue on which the ideologies may diverge. This is the case of inequality in incomes. In poor countries such inequality correlates positively with the level of happiness, but it is unrelated to inequality of happiness; in rich countries it correlates positively with inequality in happiness, but is unrelated to the level of happiness. Utilitarians and egalitarians may continue their discussion on this matter for a while, but it is expected that economic growth and institutional developments in nations solve these remaining tensions.

3.1.8 Determinants of the Risk of Uneasiness in Paduan Families

Sara Poffe; University of Padua, Italy

Luigi Fabbris; University of Padua, Italy

Abstract. The family is a continually changing social body. It evolves according to the experiences and projects of its members.

Since it is an important place where people –heterogeneous with regard to sex, age, profession –interact, it may become a privileged observatory for the study of population's problems. These problems can be classified according to content: health, relational, economic and other social.

Within most Italian families, these obstacles are overcome as soon as they are evident. Some problems are unbearable by the family and require help from outer bodies. Normally, the family faces the problems which caused this situation mainly through its own strengths or with the help of the "enlarged"family, that is other relatives who do not live under the same roof.

Family uneasiness is a critical state involving each member. If the internal resources (time, money, and so on) are insufficient, the family gives up and asks for social or voluntary services help. When neither of these resources are able to solve the problem, the family can only outlive its tragic condition.

In this paper we aim at presenting a study about uneasiness in Paduan families, measured on a dichotomous scale:

1. no heavy problems or problems solvable with family's strength;
2. problems solvable with social or voluntary services help or not solvable heavy problems.

The criterion variable is the uneasiness risk. Our aim is to identify groups of families featured by high levels of risk of uneasiness, particularly dependent on interactions between structural and foreseeable characters of the family. The analysis is performed by means of a sample segmentation method with a software named LAID-OUT.

The data were collected in 2004 and 2005 by the Permanent Interistitutional Observatory on Paduan Families. The sample of families is segmented by a stepwise procedure in accordance with the largest difference between risks of competing categories identified by dichotomously splitting the sample according to a variable at a time.

The results highlight that families where invalids or other people affected by severe chronic-degenerative diseases, and families with low income or composed by a few members (typically, single member families, and single parent with children) are the most exposed to uneasiness.

3.1.9   Personal and institutional factors in the assessment of European public policies: The case of education

Clara Riba; University Pompeu Fabra, Spain

Anna Cuxart; University Pompeu Fabra, Spain

Abstract.   The paper analyzes the differences in the assessment of public policies among citizens of European countries focusing in the personal, social and institutional factors that drive to the individual evaluations. The interest of the paper is both, substantive and methodological. It is intended to explain the sources of the differences in the education evaluations by regions and countries and it is also intended to analyze the interaction between individual and political context with a modelization that takes into account the hierarchical structure of data. The effect of the degree of decentralization in the European countries is one of the relevant issues that the paper tries to deal with. Data come from European Social Survey wave 2002 and other international institutions as OECD, ONU. In particular data from the last PISA report are used.

3.1.10   Relationship between Teaching Quality and Customer Satisfaction at University

Marisa Civardi; Università di Milano-Bicocca, Italy

Emma Zavarrone; Università di Milano-Bicocca, Italy

Abstract.   This paper examines the relationship between teaching quality and customer satisfaction (student satisfaction) of some subjects taught during the first year of a Bachelor's degree in order to analyze their causality; i.e., does teaching quality precede student satisfaction in time or vice versa? The early purpose of the work was to check the assumption that, for each discipline, the workload (WL) necessary to prepare an exam (which is evaluated according to one's own studying commitment), the width of the programme and the difficulty of the subjects themselves, could be perceived on the basis of the didactic material actually used; this one is evaluated according to clarity, exhaustiveness, availability and adequacy.

Prior to building a structural model so as to discover the causality between teaching quality and student satisfaction, a preliminary descriptive analysis has been done in order to compare the average scores related to the two latent dimensions (teaching quality and student satisfaction) both in every discipline and in the single evaluated course. As these items are surveyed on an ordinal scale, it is not possible to calculate a total score and, traditionally, either the scale is dichotomized by considering the percentage of positive assessments for each dimension, or the median is calculated. Though the first solution entails a rather relevant loss of information, the synthesis by means of the median is even less informative because it adopts a limited 4-degree scale means, therefore reducing the possibility of discrimination both in the subjects and especially in the courses. A new index, starting from a previously exposed one (Civardi, 2002), has been calculated. It is based on the observed distribution of the different kind of answers and it gives a numerical score by synthesizing the aspect under examination. The index is the algebraic sum of two indexes: an index expressing the score obtained on the half-plane of the positive assessments and another corresponding to the half-plane of the negative ones. In brief, with reference to positive assessments, let us consider, for a generic teaching i, the whole distribution of the scores assigned to k items measuring a latent dimension. If x_i indicates the percentage of positive assessments and y_i the percentage of very positive assessments on the overall positive assessments, the couple (x_i, y_i) identifies, in the Cartesian plane, a point P_i which lies on a 100 square surface (the positive assessment area). By dividing the interval into ten parts both on the axis of abscissas and on the axis of ordinates, 100 squares may be obtained which can be ordered by the positive assessment decreasing intensity. The order is obtained by assigning a range r=1 to a square placed at the upper left vertex of the area till a range r=10 is assigned, proceeding vertically, to the tenth square of the outer column (the column of the disciplines which have obtained positive assessments ranging from 90 to 100 %). A range r=11 will be later assigned to the upper square of the following column (the column of the disciplines which have got positive assessments ranging from 80% to 90%) and so on. It is thus possible to assign a score V+ = 100-r+1, $0 \leq V+ \leq 100$ to the point P_i. A here-adopted better alternative considers even the distance of point P_i from the upper left vertex of the square where it is located (point of maximum relative positivity). The new score will therefore be calculated as V*+ = V+ $-\sqrt{2}d$ (d being the distance between P_i and the point of maximum relative positivity. The calculation of the V index (negative assessment) is totally similar and the algebraic sum results in the index V [-100,100].

The data, which have been collected from the surveys carried out by the NUCLEO DI VALUTAZIONE inside the University of Milan-Bicocca in the academic years 2000-2001, 2001-2002, 2002-2003, 2003-2004, concern three subjects taught in the first teaching year at the Faculty of Economics: Mathematics (MT, 807 questionnaires), Business Economics (BE, 1042 questionnaires), Statistics (ST, 717 questionnaires). The 6 items selected to check the above-mentioned assumption have all been pointed out on a four-degree ordinal scale (1=yes, very much, 2=yes, so and so, 3=no, little, 4=no, nothing at all). According to validity analysis (face and content) the three items relating to WL contribute to explain the latent dimension that is defined as Didactic Load Tenability (DLT), considered one of the dimensions of teaching quality, while the three items relating to DM define the Didactic Material Adequacy (DMA), one of the dimensions of student satisfaction.

First results highlight that student satisfaction precedes teaching quality as far as the evidence in first two years shows; as a matter of fact the CFA carried out on the three disciplines, considering the three latent variables, have resulted in values which are still very high (RMSEA equal to 0,01, 0,067, and 0,075 respectively for BE, MT and ST) and the goodness of fit index for each course is extremely adequate.

REFERENCES

Civardi, M. (2002). Le Opinioni degli studenti .Nucleo di Valutazione. Univ. Milano-Bicocca

3.2   Measurement of Social Indicators

Session chair:
  Guillaume Osier; European Communities Statistical Office (Eurostat), Luxembourg

This session takes place on:
  Tuesday, July 19; 15:00 to 17:00

Presentations:

- Invited presentation: *Computation of variance estimates for complex poverty indicators by the technique of linearization* - Guillaume Osier (Luxembourg)

Contributed presentations:

- *Structural model of coping with unemployment* - Denisa Fedáková, Miroslav Frankovský (Slovakia)

- *Income distribution and income mobility in Italy according to two household surveys* - Claudio Quintano, Rosalia Castellano, Andrea Regoli (Italy)

- *Spanish economic inequality: A gender approach* - Mercedes Prieto-Alaiz (Spain)

- *Out of European Labour Market: What changes after retirement?* - Elsa Fontainha (Portugal)

3.2.1　Computation of variance estimates for complex poverty indicators by the technique of linearization

Guillaume Osier; European Communities Statistical Office (Eurostat), Luxembourg

Abstract. Variance estimates for values calculated from a sample are key elements of survey quality, but they are in general hardly computed. The Eurostat project is to compute in-house variance estimates for the "Laeken indicators", a set of complex poverty indicators (at-risk-of-poverty rate, at-risk-of-poverty threshold, Gini coefficient, quantile share ratio...). These indicators are compiled from survey micro-data files sent to Eurostat by National Statistical Offices (NSO). The records are collected by NSO according to a national methodology which fulfils Eurostat requirements. In this paper, we suppose we do not call on the heavy computational "re-sampling"methods (Bootstrap, Jackknife, Balanced Half-Samples...). Our goal is to compute variance estimates by direct calculations on the sample values. The linearization technique, developed by Deville (1999), is the theoretical basis of the work. Its principle is to derive from the complex poverty estimator a linear estimator for which the variance is asymptotically equal. The variance of the linear estimator is then calculated using the software POULPE, a SAS/AF application developed by the INSEE, the French NSO. The paper will present the details of the computation of the variance and confidence interval estimates for the Laeken indicators, which thus take into account their non linearity and the complexity of the survey design implemented. After developing some theoretical aspects of the linearization, we present the structure and the underlying basic principles of the software POULPE. At the end, we provide numerical examples referring to some countries and draw some conclusions about the variance of such complex estimators.

3.2.2　Structural model of coping with unemployment

Denisa Fedáková; Slovak Academy of Sciences, Slovakia

Miroslav Frankovský; Slovak Academy of Sciences, Slovakia

Abstract. According to current results of the EUROSTAT (April 2005) unemployment is one of the most serious social problems in many European countries. It is confirmed by attention which is paid to this problem at the EU level (eg.: European employment strategy, European social dialogue)

Studying unemployment requires evaluation not only from European perspective but also from particular, first of all, cultural and situational aspects of each EU country.

In the general specification frame of the research on unemployment is the attention mainly focused at coping with the situation of unemployment.

Our study presents the proposal for structural model of coping with unemployment in specific conditions of Slovakia (transformation of social and economic system, high rate of unemployment, insufficient social experience with unemployment). The Carver's at al. (1989) multidimensional questionnaire COPE was adapted to specific situation of unemployment. The structural model of latent variables is based on exploratory factor analyses results of unemployed respondents' data (N=729).

Possibilities of global (cross-cultural) validity of presented model are discussed in comparison to structural models in general.

3.2.3　Income distribution and income mobility in Italy according to two household surveys

Claudio Quintano; University of Naples Parthenope, Italy

Rosalia Castellano; University of Naples Parthenope, Italy

Andrea Regoli; University of Naples Parthenope, Italy

Abstract. The starting point of this paper is the remark that the two most relevant sample sources on household income in Italy give a very different picture of this dimension of the economic well-being for the Italian households. The above mentioned surveys are the Survey of Household Income and Wealth (SHIW) conducted since the mid-1960s by the Bank of Italy and the Italian component of the European Community Household Panel (ECHP) whose experience started in 1994 and ended in 2001. Both surveys derive the annual disposable household income as the sum of the different income components received by the household members.

We are interested in investigating first the definition of income adopted by the two surveys as well as the reporting and the editing procedures for income data. For the same reference year we want to compare both levels

and inequality measures of income distribution. We focus on the lower and the upper tail of the distribution in order to stress who the 'poor' and the 'rich' are on the basis of the two sources. Then we turn to the longitudinal dimension: in this framework a comparison of income dynamics is made between the two sources with the aim of getting insight into the way in which households move along the distribution over time. The panel component of both surveys allows to study the dynamics of the relative position of households on the basis of their income by means of transition matrices. Moreover the same component offers the chance of evaluating the evolution of the geographical disparities. Special attention is always devoted to some relevant aspects of data quality such as the attrition bias.

3.2.4 Spanish economic inequality: A gender approach

Mercedes Prieto-Alaiz; Universidad de Valladolid, Spain

Abstract. This paper examines the gender differences of expenditure distribution within the last decade in Spain. In particular, the Lorenz dominance is tested in order to know the gender gap when expenditure distributions (based on the information provided by the Encuesta de Presupuestos Familiares y la Encuesta Continua de Presupuestos Familiares) are approximated by the Dagum model . The sensitiveness of the results to some conceptual choices such as the equivalence scale or the gender reference are also analysed.

3.2.5 Out of European Labour Market: What changes after retirement?

Elsa Fontainha; Technical University of Lisbon - ISEG, Portugal

Abstract. The paper aims to contribute to a better knowledge of European elderly situation and attitudes, adopting as its main research question: What are the main changes when older people leave labor market?

Paper analysis brings together information from diverse European data sources: European Social Survey (ESS), Harmonised European Time Use Surveys (HETUS - Eurostat), Luxembourg Income Study (LIS) and European Community Household Panel microdata 8 waves 1994-2001 (ECHP- Eurostat).

The study of the relationship between elderly well being, attitudes, values and retirement status is carried through multivariate analysis of variance and non parametric tests.

The expected results could shed some light on the effects on different European countries of the raising of retirement age for the target group and for all society.

Keywords: Europe, Well-being, Elderly, Retirement, Leisure

3.3 Methodological problems in epidemological and health economic survey research

Session chair:
Herbert Matschinger; University of Leipzig, Germany

This session takes place on:
Tuesday, July 19; 17:30 to 19:30

Presentations:

- Invited presentation: *Interviewer effects in estimating prevalence rates* - Herbert Matschinger (Germany)

Contributed presentations:

- *Interviewer effects in measuring expectations* - Frauke Kreuter (United States), Elisabeth Coutts (Switzerland), Rainer Schnell (Germany)

- *Six-year changes in SF-36 in the Danish Health Interview Survey* - Ulrik Hesse, Michael Davidsen (Denmark)

- *A cross-national comparison of associations between self-rated health and other health related factors/indicators* - Michael Davidsen, Ulrik Hesse, Ola Ekholm, Jeanette M. Nielsen, Niels Kristian Rasmussen (Denmark)

- *Methodological issues in comparative health promotion research in four European countries* - Fiona Chew, Sushma Palmer (United States)

3.3.1 Interviewer effects in estimating prevalence rates

Herbert Matschinger; University of Leipzig, Germany

Abstract. In an European survey on the prevalence of different mental diseases, several rates turned out to be surprisingly low. It was suspected that at least some of the interviewer tried to shorten their interviews by not adequately processing screening questions which served as a filter for later and important parts of the interview. If an interviewer really tries to avoid these screening questions, we should assume that the probability of a positive answer decreases within the course of an interviewers work. In order to evaluate the impact of the chronological order of the interviews we will show two different approaches to analyse the very effect. First, we will investigate whether the number positively answered screening questions out of the total number of interviews at each point of the sequence actually declines. The sequence effect is modelled by a generalized linear model with a logit link function and a binomial error distribution (Nelder & Wedderburn, 1972). The number of interviews for each point in the sequence serves as the binomial denominator. Unfortunately this analysis does not take into account the behaviour of a single interviewer and provides no information about the variation of the sequence effects with respect to the interviewers.

Therefore, in the 2nd step, a logit mixture model for a two-level design is adopted, to analyse the effect of the order of the interviews on the probability of a positive answer. The interviews are sorted from the first to the last interview within each interviewer and this order is employed as a predictor. Since the number of interviews carried out by the interviewer ranged from 1 to 260, the number interviews for each point in the sequence necessarily decreases. To control for the effect of this decrease on the probability of the outcome, this number is employed as a second predictor. To model the heterogeneity of the population of interviewers with respect to the sequence effect, several logit models within latent classes are estimated (Vermunt, 1997; Vermunt & Magidson, 1999).

This segmentation by latent classes allows for the identification of subgroups of interviewers, responsible for the decrease and for the low prevalence rates. Solutions for 4 and 5 classes will be compared and several problems on how to determine the adequate number of classes will be discussed.

Nelder,J.A. & Wedderburn,R.W.M. (1972). Generalized Linear Models. Journal of the Royal Statistical Society Series A-Statistics in Society, 135, 370-384.

Vermunt,J.K. (1997). Log-linear Models for Event Histories. Thousand Oaks, London, New Delhi: Sage Publications.

Vermunt,J.K. & Magidson,J. (1999). Exploratory Latent Class Cluster, Factor, and Regression Analysis: the Latent Gold Approach. In Anonymous, Proceedings EMPS'99 Conference. Lüneburg:

3.3.2 Interviewer effects in measuring expectations

Frauke Kreuter; University of Maryland, United States

Elisabeth Coutts; ETH, Switzerland

Rainer Schnell; University of Konstanz, Germany

Abstract. Measuring expectations is one core concern of three major health and retirement surveys; the US Health and Retirement Survey (HRS), the English Longitudinal Study of Ageing (ELSA), and the Survey of Health, Aging and Retirement in Europe (SHARE). The respondents' perception of the probability with which certain events will occur is meansured in all three surveys with probability scales on which the respondent is asked to provide a value on a 0-100 scale. While the use of a numerical scale has some advantages over the use of a verbal scale, such as the possibility of expressing judgments more precisely and a potenially closer match to objective probabilities, the numerical values provided are not unproblematic. Past research has shown a lack of external calibration, severe overestimation of probabilities in percent terms, as well as heaping (the high fraction of percent answers given that are multiples of ten), and high fractions of 0 and 100 percent answers.

Another phenomenon commonly observed is the relatively large number of 50 percent answers given by respondents. This paper will briefly review various theories to explain the 'mis'-use of these probability questions, and will focus on the role of the interviewer in the answer generating process. This will be done with the help of data collected as part of the DEFECT study (Schnell & Kreuter 2000) and data from experiments using various measurement scales (Coutts & Schnell 2005).

The DEFECT study offers the unique possibility of isolating interviewer effects on probability questions. The DEFECT study is the first nationwide sample to be drawn in Germany with an interpenetrated sampling design.

The same questionnaire was used to conduct five independent surveys in 160 sampling points. Four of the five surveys were conducted by professional survey institutes that are highly regarded for their good practices and reliability, while the fifth (the mail survey) was conducted by the DEFECT group itself. This design means that in each of these 160 sampling points, two face-to-face surveys with random sample selection, a face-to-face survey with quota selection, a CATI, and a mail survey (both with random selection) were carried out concurrently.

3.3.3 Six-year changes in SF-36 in the Danish Health Interview Survey

Ulrik Hesse; National Institute of Public Health, Denmark

Michael Davidsen; National Institute of Public Health, Denmark

Abstract. Introduction

Short Form 36 (SF-36) is often used in both health surveys as well as in medical studies describing self-assessed health. In Denmark SF-36 norm data derives from The Danish Health and Morbidity Survey 1994. It is well known that self-assessed health decreases with increasing age. This study aims to describe changes in general health and mental health from 1994 to 2000 in a nationally representative population.

Methods

Nationally representative Health and Morbidity Surveys of the adult Danish population have been conducted in 1994 and 2000. In 2000 the sample included persons invited to participate in the 1994-survey. SF-36 was measured using a self-administered questionnaire that persons accepting an interview in their home fulfilled after the interview. 2773 persons who participated both in 1994 and in 2000 were included in the study. Mean scores from the two surveys and the difference from 1994 to 2000 were calculated, the latter on individual level.

Results

Mean general health was 76.2 in 1994, and decreased by 2.2 (95% CI: 1.71 –3.06) during the six-year follow-up. This pattern was seen for both men and women and in all age groups, the decrease being most pronounced among the elderly. The mean score of the SF-36 mental health component was 82.8 in 1994 and increased over the six-year period with 0.28 (95% CI: -0.32 –0.88). The pattern however differed for men and women as well as for different age groups. The largest differences were found among men aged 67 or older and women aged 45-66 years of age.

Conclusion

The results from this study indicate that changes in self-perceived health tend to differ according to the measured aspect of SF-36. Results indicate that well-known problems with normality-assumption for SF36 are considerably smaller when measuring change.

3.3.4 A cross-national comparison of associations between self-rated health and other health related factors/indicators

Michael Davidsen; National Institute of Public Health, Denmark

Ulrik Hesse; National Institute of Public Health, Denmark

Ola Ekholm; National Institute of Public Health, Denmark

Jeanette M. Nielsen; National Institute of Public Health, Denmark

Niels Kristian Rasmussen; National Institute of Public Health, Denmark

Abstract. Background. Many EU-countries regularly performs health interview surveys containing information on self-rated health. In many instances the questions posed appear to be similar/identical, but data and results from these surveys seem to indicate lack of direct comparability. A north-south gradient has been found leading to discussions of cultural and linguistic differences.

Aim. The purpose of this study is to compare associations between self-rated health and lifestyle and health behaviour in 5 regions in EU and to quantify and compare magnitude of these associations..

Material and Methods. Data derives from health surveys in 4 EU-countries: The Scania Public Health Survey (Sweden, n=13715), the Danish Health and Morbidity Survey (Denmark, n=16690), Belgium Health Interview Survey (Flemish region n=3223, Walloon region n=3709) and National Health Interview Survey (Portugal,

n=23440). Prior to analysis a conceptual model depicting our understanding of the processes leading to a person's evaluation of her/his health was constructed. Based on questions having a reasonably similar wording the following 10 indicators were chosen: smoking, leisure time physical activity, obesity, alcohol consumption, activity restrictions, contact to a GP, longstanding illness, cohabital status, education and age. Statistical analysis was performed for each region and each sex separately. We used discrete graphical models to obtain a graph from which multivariate associations can be seen directly. To compare the strength of associations we used Goodman and Kruskals gamma.

Results. Among men the age-standardised prevalence of bad or very bad self-rated health was 5.0% in Scania, 5.0% in Denmark, 2.6% in the Flemish region, 4.8% in the Walloon region and 13.6% in Portugal –the corresponding figures for good or really good self-rated health were 74.0%, 80.5%, 81.3%, 76.4% and 48.6%. A similar pattern was seen for women. A clear gradient in self-rated health was for most indicators seen in all countries. For education the prevalence of bad or very bad health decreased with increasing education. The graphs obtained by thorough analysis showed a heterogeneous pattern in the association between all variables. Among women a positive association between self-rated health and leisure time physical activity, activity restrictions, longstanding illness and contact to a GP was found in all counties while a negative association to education was found in all regions but Scania. Results for other variables varied. Goodman and Kruskal's gamma showed the same direction in all countries/regions but the size varied.

Conclusions. Different prevalence of self-rated health lead to a hypothesis of lack of comparability between countries/regions. But the associations showed a similar structure indicating that the outcome variable measured the same basic concept in all the regions..

3.3.5   Methodological issues in comparative health promotion research in four European countries

Fiona Chew; Syracuse University, United States

Sushma Palmer; Central European Center for Health and the Environment, United States

Abstract.   An unfortunate by-product of state-controlled economies in Central and Eastern Europe transitioning to market-based economies, was the erosion of the preventive health care system. As an attempt to alleviate the health crisis in Russia, Hungary, Poland and the Czech Republic, a five-part television series focusing on health problems and health promotion was developed for broadcast with the input of local national health agencies in these countries. This comparative research study assessed the impact of the series on viewers' perception and practice of health behaviors.

In Central and Eastern European countries, the research infrastructure varied considerably so that comparative telephone survey research did not meet the criteria of representativeness and generalizability. Therefore, this study proposed a longitudinal field experiment and focused on qualified respondents, therefore conserving resources otherwise spent on polling unqualified respondents. The original research design outlined telephone interviewing 150 program viewers and 150 non program viewers. A year later, these same respondents would be contacted for long term effects.

The four-country research study was completed with the participation of local health agency researchers. From April 1996 through July 1998, 973 persons were surveyed/interviewed. Fieldwork contingencies arose and the original research design was modified. In Russia and Hungary, since the telephone was narrowly distributed among the population, the personal survey was confined to the main cities. In Russia, selected respondents in Moscow were provided with a videocasette recorder for three days, asked to view the programs in their homes, and polled again in three to six months. In Hungary, respondents were recruited to group view the programs at three locations – Budapest, Budapest outskirts and the city of Debrecen –and reinterviewed a year later.

In Prague, 150 Czech subjects were randomly contacted and telephone surveyed. Non viewers were also polled and both groups polled again in six months. In Poland, due to the Christmas holidays' proximity to the program broadcast, program evaluation involved screening all five episodes to a sample of 190 Poles who were interviewed after the viewing. Non viewers were contacted by telephone and both groups reinterviewed three to six months later. Key results showed that despite research design differences, viewers in all four countries reported intentions to change and health behavior changes. Viewers rated the nutrition and general health promotion television programs favorably. Media messages can be used to motivate and educate viewers. More televised health promotion programs will benefit audiences. Methodological equivalence is discussed.

3.4   Quality of Life Indicators: single or composite indices?

Session chair:
  Joachim Vogel; University of Umea, Sweden
  Heinz-Herbert Noll; Centre for Survey Research and Methodology (ZUMA), Germany
  Jeroen Boelhouwer; Social and Cultural Planning office (SCP), Netherlands

This session takes place on:
  Wednesday, July 20; 9:00 to 11:00
  Wednesday, July 20; 11:30 to 13:30
  Wednesday, July 20; 15:00 to 17:00

Presentations:

- Invited presentation: *Summarising the quality of life: an introduction to three approaches* - Joachim Vogel (Sweden), Heinz-Herbert Noll (Germany)

Contributed presentations:

- *Quality Of Life in Europe: comparing states using a composite index* - Jeroen Boelhouwer (Netherlands)

- *Comparing welfare of nations* - Hans-Olof Hagén (Sweden)

- *The Political Potential of Well-being Indicators –creating new measures of Quality of Life & Well-being for the UK policy arena* - Tim Jackson, Nick Marks (United Kingdom)

- *The Use of Composite Indicators for Ranking Countries: a Defensible Setting* - Giuseppe Munda (Spain), Michela Nardo (Italy)

- *Welfare - as we ourselves perceive it* - Claes-M. Cassel, Jan Eklöf, Lars Lundgren (Sweden)

- *Synthetic Indicators Of Quality Of Life For Europe* - Noelia Somarriba Arechavala, Bernardo Pena Trapero (Spain)

- *Using the British Household Panel Survey for composite indices of individual welfare* - Christian Brand (United Kingdom)

- *Measuring subjective well-being as an additive composite or in terms of its sub-dimensions: How robust are associations with socio-demographic variables and self-assessed health?* - Torbjorn Moum (Norway)

- *On the Impact of Different Response Scales, Different Happiness Questions, and Different Time Applications on Subjective Well-Being Research* - Mariano Rojas (Mexico)

- *Is life satisfaction influenced by comparisons with international reference groups? some evidence from Euromodule surveys* - Jan Delhey (Germany)

- *Do Spanish people perceive quality of life in work as it is defined by the European Commission? A comparison from surveys results against an aggregate social indicators approach* - Vicente Royuela, Jordi López-Tamayo, Jordi Suriñach (Spain)

- *Single-item job satisfaction measures and quality of life: A area of comparative advantage?* - Michael Rose (United Kingdom)

3.4.1 Summarising the quality of life: an introduction to three approaches

Joachim Vogel; University of Umea, Sweden

Heinz-Herbert Noll; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   Our introduction to three seminars on summarizing the quality of life will compare the three major traditions rooted in economics (national accounts), sociology (social indicators, social surveys) and psychology (subjective wellbeing). They represent parallel routes, which have matured in recent years, as data availability and conceptual consensus have increased. New constructs have been developed, combining the three approaches. The convergence of approaches is also producing larger similarity in the classification of nations.

Our introductions will:

1. compare various approaches and principles of composite index construction

2. compare the major comparative QOL indices with respect to classifications of QOL levels in developed nations.

3.4.2 Quality Of Life in Europe: comparing states using a composite index

Jeroen Boelhouwer; Social and Cultural Planning office (SCP), Netherlands

Abstract.   Because of the growing integration within the European Union, together with the interest of EU in social policies and living conditions in member states, the question of comparison between the states becomes more and more important. Current reports, like The social situation in the European Union, comprises various indicators about social issues. These reports lack, however, an integrative measuring instrument. With the European Social Survey a great source for comparison is available.

The main goal of the paper is to explore possibilities for constructing a composite index which is meaningful and useful for describing and comparing quality of life in different states of the European Union. The paper will seek for an integrative index for quality of life based on data from the European Social Survey (ESS). As it will be based on the ESS, the index will be constructed using micro-data. This is a distinction with commonly used (macro) indicators.

The starting point of the search will be a conceptual model that The Dutch Social and Cultural Planning Office (SCP) developed for measuring living conditions in the Netherlands. At the centre of this model are the living conditions, measured as a multi-dimensional concept with the life situation index (LSI). LSI combines eight domains of the living conditions into one single index (housing, health, mobility, leisure activities, consumer durables, social participation, sport and holiday activity). Income, education and work are not part of the index as they are considered resources for realising good living conditions (a causal relation is posited). LSI is based on Dutch survey-data only.

With the limits of availability of data in the ESS the new-to-develop index will combine indicators of various domains into one index. Considered for inclusion in de index are indicators of :

•prosperity and standard of living (ic income, as other indicators are not available);

•health;

•leisure activities

•social participation and social cohesion.

•crime and safety

The results are described following the LSI model, including the relationship with resources, with happiness and with perceived health.

3.4.3 Comparing welfare of nations

Hans-Olof Hagén; Statistics Sweden, Sweden

Abstract. My presentation is an attempt to show how a complex reality can be illustrated using different statistical methods. The purpose of this report is not to exhibit the actual results of analysis, but rather to show the methods used to arrive at those results. The example chosen for analysis is a comparison of the level of welfare in OECD countries and the efficiency of these countries to create a high economic standard and welfare for their citizens.

Because welfare is an extremely ambiguous concept, it is very difficult to measure. There are no given answers on the meaning of the concept of welfare, nor any explanations on how to measure it. Attempts to do so are thus much debated.

In this report, welfare has consistently been described as a multifaceted concept. The common concept gross domestic product, GDP, is difficult to calculate, because we must combine measures for activities that are difficult to compare. These activities include services and goods that are sold on a market compared with goods and services that more or less are distributed freely. Nevertheless, these measures have a common yardstick called money. However, welfare is another matter. Its components cannot be measured with the same yardstick.

In order to compare these more complicated concepts, a composite indicator has been created. Although this method is rather controversial, it has been applied increasingly.

In simple terms, a composite indicator is a way of putting apples and oranges together in order to decide which fruit basket is the most attractive. But this indicator can be problematic. For example, to someone who only likes grapes, it doesn't matter how many apples and oranges there are in the baskets.

Furthermore, many statisticians also believe that only single variables can be reported in a satisfactory way. But neither decision-makers nor the general public wants a report that looks like a huge catalogue where variable after variable is listed page after page as a base for their understanding. Even though subjectivity is inevitable, they prefer to find out which fruit basket is probably the most interesting, rather than a list that states how many twenty or so different kinds of fruit each basket contains.

3.4.4 The Political Potential of Well-being Indicators –creating new measures of Quality of Life & Well-being for the UK policy arena

Tim Jackson; University of Surrey, United Kingdom

Nick Marks; NEF, London, United Kingdom

Abstract.   The well-being programme at nef, a UK think tank, was set up in 2001 with the intention of addressing the question 'what would policy look like if promoting people's well-being was one of the main aims of government?' To this end we published three well-being reports in 2004 on objective and subjective indicators of well-being and their policy implications.

This paper will draw on our experience of creating a new objective Measure of Domestic Progress (MDP), which is an adjusted economic indicator of quality of life, as well as on-going work with the UK Sustainable Development Commission (UK-SDC) towards creating a new composite set of national 'well-being' indicators.

1. The MDP, which is based on the Index of Sustainable Economic Welfare (ISEW) (or Genuine Progress Index –as it is known in the US), seeks to differentiate between 'good' and 'bad' economic growth, thus creating a pragmatic tool for policy makers to use. nef has recently been commissioned by a UK Regional Development Agency (RDA) to investigate the feasibility of constructing a regional MDP. The RDA see the strength of the MDP is that it transparently operationalises, and communicates, the concept of Sustainable Development.

2. National well-being indicators have been recognised by the UK government as potentially useful concept. In the recent new UK Sustainable Development Strategy there was a pledge to create such indicators and nef is working with the UK-SDC on their development. This will be greatly supported by round 3 of the European Social Survey, which has recently awarded a team, that includes nef, a module on personal and social well-being across Europe.

The paper will summarise these approaches and discuss their potential as policy tools, particularly for promoting Sustainable Development.

3.4.5 The Use of Composite Indicators for Ranking Countries: a Defensible Setting

Giuseppe Munda; Universitat Autonoma de Barcelona, Spain

Michela Nardo; European Commission Joint Research Centre (JRC), Italy

Abstract.   Composite indicators are very common in economic and business statistics for benchmarking the mutual and relative progress of countries in a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation. As a consequence, improvements in the way these indicators are constructed and used seem to be a very important research issue from both the theoretical and operational

points of view. This paper starts with an analysis of the axiomatic system underlying the mathematical modelling used to construct composite indicators. Since theoretical inconsistencies are found, a new mathematical aggregation convention is developed for composite indicators aimed at ranking countries. Main features of the proposed approach are (i) the axiomatic system is made completely explicit, and (ii) the sources of technical uncertainty and imprecise assessment are reduced to the minimum possible degree. Along the whole paper, concepts mainly coming from measurement theory, multi-criteria decision analysis and social choice are used.

3.4.6   Welfare - as we ourselves perceive it

Claes-M. Cassel; Statistics Sweden, Sweden

Jan Eklöf; Statistics Sweden, Sweden

Lars Lundgren; Statistics Sweden, Sweden

Abstract. Many situations can be analysed using a method that consists of measuring a whole and decomposing it into a number of components. The decomposition will create a structure for analysis. Typically, the purpose is to find the current status of the whole and to improve the status by also measuring the status of the components and the impacts of the components on the whole. Focusing on the components that have the largest impacts will yield the most efficient way for improvement. The method has been used in economics under the name of CSI (Customer Satisfaction Index). It can also serve as a complementary approach to poverty measurement. Survey sampling, sometimes using complex sampling designs in combination with structural equation models with latent variables, is used for collecting and analysing the data. Basically the data are the responses to questions about peoples satisfaction with different aspects of the components.

The structure is solved using statistical methods such as PLS (partial lest squares) or FIML (full information maximum likelihood). The assessment of precision can be done using the super population method which takes into account both the model and the sampling. The purpose of this paper is to describe the method and give some examples of its uses

3.4.7   Synthetic Indicators Of Quality Of Life For Europe

Noelia Somarriba Arechavala; Universidad de Valladolid, Spain

Bernardo Pena Trapero; University of Alcalá, Spain

Abstract. The definition of the notion of quality of life faces us to one of the great dilemmas in the investigation of this controversial concept.. The quality of life as object of study, depends on the context in which it studies, offering a great variety of viewpoints for analyses that, without a doubt, will enrich our analysis.

The economists in the past have been sceptical and they have resisted incorporating information of subjective type in the analyses of well-being and quality of life.

However, at the present time, a great number of authors exist who approach the analysis of the well-being incorporating more subjective components and considering that the well-being not only depends on objective conditions but that are influenced by the appreciations that the own individuals have on their quality of life.

In the first part, we will approach the selection of possible indicators or attributes for the measurement and analysis of this concept, in the set of European countries, from the information provided by European Social Survey and others informational sources.

Secondly, we will apply a set of methods with the goal to construct a synthetic indicator of well-being that allows us to make space comparisons. Within this set of methods, would be:

The analysis of main components, technique that is including in the multivariante analysis, will be used in the following section with the objective to obtain a synthetic indicator of well-being.

Also, we will incorporate an approach to the concept of quality of life by means of the measurement of P2 distance, is a synthetic indicator that adds the information contained in a set of social indicators and that is designed to make inter-spatial and inter-temporary comparisons.

Finally, the methodology of the Data Envelopment Analysis (DEA), traditionally has been used for the estimation of the relative efficiency of a set of productive units. In the last years, other applications have been developed in which it includes the DEA for the derivation of synthetic indicators of well-being and quality of life. From this method, we will try to derive a synthetic indicator from quality of life.

After applying these methods, we will compare the obtained results, analyzing the advantages and disadvantages of each one of them.

3.4.8   Using the British Household Panel Survey for composite indices of individual welfare

Christian Brand; University of Manchester, United Kingdom

Abstract. The British Household Panel Survey (BHPS) is an excellent source of social data covering already 14 waves since 1990. British and comparative European social researchers have made extensive use of it. This use, however, tends to fall into a large number of 'single-issue' areas, such as research on income poverty, multiple deprivation, labour market participation, social capital, and so on. In sharp contrast to this, very little interest has been expressed in using the BHPS as a data source for the research of broader concepts of individual welfare (aka Quality of Life) and subsequent social reporting and 'index-construction' purposes. The reasons for this situation are probably twofold. Firstly, there is the UK's market-orientated, 'liberal', socio-economic culture and its often-expressed distrust for social reporting as a prelude to 'social planning'. Secondly, continental European researchers in those fields have, for obvious reasons, preferred their own national datasets. Nevertheless, the quality of the BHPS and the fact that the UK, while different on many accounts, is still a reasonably similar (post)industrial society suggest that several European approaches could be replicated in this British dataset. Hence, I will attempt to make such use of the BHPS; for the scope of this paper following the index operationalisations suggested by the Social and Cultural Planning Office of the Netherlands (Boelhouwer, 2002) and by Roland Habich (1996) of Germany. Apart from presenting comparative results as the end-product of the replication process, the paper will document the numerous methodological problems encountered in constructing a composite index in general, as well as using a dataset which –while of very high quality –is still riddled with innate insufficiencies, such as lack of one-to-one comparability, rotating variable coverage, missing values etc.

References:

Boelhouwer, Jeroen (2002) 'Quality of Life and Living Conditions in the Netherlands', Social Indicators Research [electronic journal] 58 (1/3): 115-140.

Habich, Roland (1996) 'Objektive und subjektive Indikatoren - Ein Vorschlag zur Messung des Wandels von Lebenslagen', In: Lebenslagen im Wandel: Sozialberichterstattung im Längsschnitt, edited by Wolfgang Zapf; Jürgen Schupp and Roland Habich, Frankfurt a. M.: Campus: 46-65.

3.4.9   Measuring subjective well-being as an additive composite or in terms of its sub-dimensions: How robust are associations with socio-demographic variables and self-assessed health?

Torbjorn Moum; University of Oslo, Norway

Abstract.   The focus of this presentation is on a well-established index tapping overall subjective well-being (SWB) / quality of life (Moum et al., 1990). This SWB index has been utilized in numerous previous studies of morbidity and health-related quality of life. It consists of four components, viz. (I) cognitive appraisal / life satisfaction, (II) energy / vitality, (III) overall positive-negative mood, and (IV) tenseness / anxiety. When summated to create an additive SWB index, each component contributes positively to the overall internal consistency of the summary measure, with a satisfactory alpha reliability ($\alpha > 0.75$). A large twin study (N>5000) has shown the SWB index to be highly heritable (h2≈0.5), with much of its association with self-assessed health (r=0.50) being attributable to the same genes (Røysamb et at., 2003).

Research questions:

We investigate the parallelism of the four sub-components of the SWB-index with respect to: (1) Temporal stability (test-retest); (2) Statistical associations with age, gender, and level of education; (3) Cross-sectional statistical associations with self-rated health; (5) Responsiveness to changes in self-rated health.

Material:

Responses to self-administered questionnaires from participants in two broad-population health surveys in 1984-85 (time 1) and 1995 (time 2) in the county of Nord-Trøndelag in mid-Norway (HUNT) are analyzed. Complete data are available at both points in time for 13,468 females and 11,843 males, aged 20-70 at time 1.

Results:

Test-retest correlations are comparable across components (0.36-0.44), with the summary SWB index showing the highest stability coefficient (0.51). With the exception of "tenseness / anxiety"there is an overall slight

inverse association between subjective well-being and age, but with a slight increase after age 50 for all components. Females report less energy / vitality and more tenseness / anxiety than males, whereas levels of overall life satisfaction and mood are highly comparable among males and females. The total SWB-score is slightly more favourable among males (r=-0.08). Male-female differences do not vary systematically by age for any of the components. All SWB components show slight positive correlations with level of education, with the strongest association (r=0.13) found for "energy / vitality". The associations between sub-components and self-rated health are entirely comparable at time 1 and time2, with the strongest correlations found for "energy / vitality"and the lowest for "mood"and "tenseness". Results for the overall SWB index are very comparable to those obtained for the "energy / vitality"components. Analyzing change in SWB as a function of change in self-rated health, i.e. the "responsiveness"of SWB and its sub-components to changes in self-rated health, yield results that are entirely comparable to those obtained cross-sectionally, with strongest coefficients obtained for "energy / vitality"and the overall SWB-index. With increasing age there is a slight increase in the association between self-rated health and overall SWB, but this pattern applies particularly to the sub-dimensions of "energy / vitality"and "tenseness / anxiety".

Conclusions:

Substantive conclusions regarding the association between SWB and socio-demographic variables such as age, gender and level of education would largely be the same regardless of which sub-dimension of SWB is focused upon. However, the dimension of "energy / vitality"seems to show the strongest correlations with age, gender, level of education, as well as with self-rated health.

3.4.10   On the Impact of Different Response Scales, Different Happiness Questions, and Different Time Applications on Subjective Well-Being Research

Mariano Rojas; Universidad de las Américas, Mexico

Abstract. This paper studies three main issues related to survey research and the explanation of subjective well-being. The paper uses a country-wide Mexican survey with more than 1500 observations, as well as econometric techniques, to address the following three issues:

First, whether the choice of a happiness response scale makes a difference in the explanation of subjective well-being (estimation of coefficients in assumed relationships). The same question was asked twice to each person in the survey -at different stages in the questionnaire-, while two different response scales were used. One of the questions has a 1 to 10 numerical response scale, while the other question has a 1 to 7 verbal response scale. The paper uses econometric techniques to study the sensitiveness of estimated subjective well-being relationships to the response scale used. For example, it is examined whether the response scale influences the value of the estimated parameters in the relationship between satisfaction in domains of life and happiness, as well as in the relationship between socio-economic and demographic variables and happiness. Even though there is a vast literature on response scales;, only a few studies make research comparisons across response scales.

Second, the paper studies whether the estimated parameters in the relationship between satisfaction in domains of life and happiness -as well as in the relationship between socio-economic and demographic variables and happiness- are influenced by how the happiness question is phrased. Two different happiness questions were asked to each person in the survey. One happiness question is phrased as "Taking everything in your life into consideration, How happy are you?", while the other question is phrased as "Making a recount of all your life, How would you say it has been?"The econometric exercise is also applied to a life-satisfaction question.

Third, the paper studies the sensibility of subjective well-being answers (happiness measured in a 1 to 10 numerical scale, happiness measured in a 1 to 7 verbal scale, life satisfaction, and happiness recount) to the day of the week and the time of day (morning v.s. afternoon) when the survey is applied. The sensibility of subjective well-being answers to the context of the survey has been a constant issue in subjective well-being survey research.

3.4.11   Is life satisfaction influenced by comparisons with international reference groups? some evidence from Euromodule surveys

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

Abstract.   This paper deals with subjective life satisfaction as a single-indicator approach to QOL. Among scholars there has been a lively debate between two camps, known as relative utility theory and absolute utility theory. Whereas the former has claimed that, in the first instance, life satisfaction reflects how people fare relative

to others within the same society, the latter has claimed that high or low life satisfaction mainly reflects the absolute level of need fulfilment a person enjoys. Comparative research has confirmed that absolute utility matters more than relative utility, since the low-income groups in the more affluent countries are still more satisfied with life than the high-income groups in the less well-off countries.

My paper picks up again the debate by introducing the idea of international reference groups. It questions the common assumption that people define relative utility only by comparing themselves to others within the same society. Rather, in the "global village"of today it is perfectly possible that individuals are quite aware of where they or their own country stand in the international (or at least European) hierarchy of wealth and that they take that standing into account in arriving at subjective evaluations of their personal living conditions as "good"or "bad".

Based on recent Euromodule surveys from Germany, Hungary and Turkey, the paper shows first empirical evidence that cross-border comparisons of living conditions matter for individual life satisfaction. Obviously, foreign nations do serve as a point of reference for people's subjective well-being. This holds true for such different populations as the Germans, Hungarians, and Turks. In the first line, however, upward comparisons are salient for people's well-being, not downward comparisons. The more people feel personally deprived, relative to other countries, the less satisfied they are with their lives. This corroborates the proverb "the grass is always greener over there". In contrast, to know that the grass is greyer over there does not make people more content.

The findings add a new layer to our theoretical understanding of how citizens evaluate their living conditions. Life satisfaction is not only based on absolute need fulfilment and within-country comparisons with fellow citizens, but also on cross-border comparisons.

3.4.12 Do Spanish people perceive quality of life in work as it is defined by the European Commission? A comparison from surveys results against an aggregate social indicators approach

Vicente Royuela; University of Barcelona, Spain

Jordi López-Tamayo; University of Barcelona, Spain

Jordi Suriñach; University of Barcelona, Spain

Abstract. Quality in Work is defined by the European Commission as a multidimensional concept, and thus it has to be measured considering all vectors. In our study we analyse the results coming from the survey on Quality of Life in Work, in which people is asked about their satisfaction in work, among other key questions. Additionally we compute a composite index on quality in work following the dimensional structure from the European Commission, in which we include both structural indicators and several partial results from the former survey.

Our work is developed for Spanish regions and for economic sectors. We also consider as possible sources of information two key aspects in the labour market: firms size and professional categories of workers. We analyse the period 2001-2004.

3.4.13 Single-item job satisfaction measures and quality of life: A area of comparative advantage?

Michael Rose; University of Bath, United Kingdom

Abstract. Although there is no direct method for establishing the internal consistency reliability of single-item work attitude measures, Wanous and his colleagues applied Nunnally's suggested develop-ment of the formula for the attenuation of reliability to achieve an indirect estimate. The Wanous meta-analysis suggested an internal consistency reliability of around 0.63 for SI overall job satis-faction measures. While this result is not entirely reassuring, Wanous et al. regarded it as indicat-ing acceptable performance in many situations. One reason for doing so, they argued, is that the theoretical assumption of a correlation of unity between true scores and a perfect measure of them is somewhat unrealistic and might be relaxed: assuming a maximum practically possible correla-tion of 0.90, the SI scores achieve notional Alpha values of just over 0.70. Job satisfaction data generated in most applied and organisation psychology is characterised by heterogeneity of meas-ures and relatively small samples, limiting further examination of these claims. However, since the 1980s, research mainly by teams of sociologists of work and labour economists, using large na-tionally representative employee samples, has created a wealth of data relevant to the performance of SI job satisfaction indicators among UK employees. (The samples also offer a

large number of organisational and individual contextual and control variables.) These data, like those examined by the Wanous team, show that the IRC of single-item measures can be acceptable. True enough, the IRC of composite measures is far higher, and such composite indicators yield more precise estimates of association with most outcome related measures: for example, ratings of own supervi-sor's ability; belief that employer delivered on undertakings to provide training; extent of personal involvement in decisions about how job is done; and so on. The average increase in common vari-ance explained is over 50%. However, factor analysis confirms the existence of at least 4 distinct dimensions to the job satisfaction of British workers. A SI measure of job satisfaction is most closely associated with a dimension readily interpreted as quality of working life in present job. In this sense, a SI measure of job satisfaction is quite successful at tapping some of the more im-portant quality of life related aspects of work. Moreover, and importantly for quality of life stud-ies, a single-item measure is no less successful than a composite measure in showing association with subjectively experienced stress at work.

3.5   Emotional Competence

Session chair:
   Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain
   Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

This session takes place on:
   Wednesday, July 20; 17:30 to 19:30

Presentations:

- Invited presentation: *Which is the underlying model and the appropriate analysis strategy for assessing Emotional Intelligence Competencies* - Richard E. Boyatzis (United States), Joan Manuel Batista-Foguet, Ricard Serlavos (Spain)

Contributed presentations:

- *Problems with measuring Social Intelligence* - Katarína Vasilová, František Baumgartner (Slovakia)

- *Properties of the Emotional Intelligence Questionnaire (ECI)* - Laura Guillén, Ricard Serlavos (Spain)

- *Generating a new learning environment in management education: a comparison between an Spanish and a US MBA experience* - Laura Guillén, Joan Manuel Batista-Foguet, Ricard Serlavos (Spain)

- *A longitudinal perspective to graduate management education: an integrated competency-career approach* - Arnaldo Camuffo, Fabrizio Gerli, Francesca Chiara (Italy)

3.5.1 Which is the underlying model and the appropriate analysis strategy for assessing Emotional Intelligence Competencies

Richard E. Boyatzis; Case Western Reserve University, United States

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Abstract. The measurement of EI is a big challenge that has been undertaken from different perspectives (Salovey, P., and Mayer, J.D. 1990; Bar-On, 1992; Boyatzis, 1992; Goleman, 1998). In 1999, Goleman and Boyatzis proposed an instrument, the Emotional Competence Inventory (ECI). based on a contingency theory of effective performance, that operationalise the EI concept through emotional competencies (Boyatzis, 1982; Goleman, 1995, 1998). The ECI was developed as a multi-source questionnaire; that is, a 3600 format assessment of a person's expressed competencies, or competencies in use. The current version of ECI (ECI-2) has four items per scale, resulting a total of 72 items thought to comprise the eighteen competency scales. These competences are grouped into four interrelated clusters: Self-Awareness; Self-Management; Social Awareness and Relationship Management or Social Skills (Boyatzis, R. E., Goleman, D., & Rhee, K. (1999). Among the reasons for reviewing the instrument Boyatzis and Sala (Boyatzis & Sala, 2004) pointed out that competency scales showed inter-correlations that were too high, so getting more factor differentiation/ discriminant validity among the various components, actually Competences of the EI construct.

Every effort for EIC assessment has been guided by Classical Test Theory (CTT) tradition, which assumes that an underlying factor is measured through a series of "reflective"indicators that are manifestations of the latent variable. As a consequence, covariation among the measures is attributed to their assumed common cause (i.e. the latent competence. From this assumption Factor Analysis model for validity assesment, and other measures for reliability as Chronbach's α, based on items consistency are justified.

On the one hand, Exploratory Factor Analysis (EFA) requires to only establish the number of factors and in order to reproduce the correlation (covariance) matrix the model forces the items to be invalid, since they load on every common factor. Besides classical reliability estimates, like Cronbach's, requires the items being at least Tau-equivalent.

On the other hand, the use of Confirmatory Factor Analysis (CFA) introduce genuine restrictions rather than statistical and arbitrary ones, structure the data in accordance with specific model parameters for a concrete application, achieve a single solution for these parameters, check the validity of the theory that enabled you to construct the model, obtain measurement quality indexes for the global model and for each item that are useful to check whether Cronbach's or other indices could be applied.

However, both CFA and EFA are also based on CTT, so the quality of the items will be evaluated accordingly. Those that are consistent among themselves are the ones which have higher reliability, obviously assuming they are measuring the same factor. Confirmatory Analysis of the revised ECI used for assessing in the 360ºformat more than 7000 MBA participants from Case Western Reserve have shown that the within cluster correlations among competences are not higher than between clusters, besides some of the correlations among clusters (second order factors) can not be rejected be equal to 1.

So these results could be interpreted, even taken into account common method variance or halo effect, as a confirmation of the existence of the EI construct, although with less differentiated clusters of competences as were assumed. However, this is the easiest "explanation"which would likely lead to two or three clusters. On this paper we present another possible specification for modelling the EI construct based on the eighteen competences hypothesised so far. However, in stead of being based on CTT which had lead to interdependence analysis, and consequently factor analysis models, we actually explain the magnitude of the correlations among competences analysing the dependence structure through Path Analysis correcting for measurement error in the corresponding structural equation model.

3.5.2 Problems with measuring Social Intelligence

Katarína Vasilová; Slovak Academy of Sciences, Slovakia

František Baumgartner; Slovak Academy of Sciences, Slovakia

Abstract. The paper deals with the problems of measuring social intelligence. The most appealing difficulties arise from the fact that social intelligence is a multi-dimensional construct and the solid, universally acknowledged definition is still missing. There is also the problem with overlapping constructs like: social competence, emotional intelligence, practical intelligence, etc. and a varying terminology used by researchers. Social intelligence can be seen as a psychometric or a personality construct. There can be recognized the origins of the difficulties with designing a valid measurement tool. Based on the literature review several social intelligence methods were analyzed and their common features and differences were identified. This provided a framework for the construction of the new methods concerning various aspects/components of social intelligence.

The data was gathered from the sample of university students. There were used both the new and the standardized methods of social intelligence. The results indicate that social intelligence has several facets and it is of a great importance to identify the essential components and to design the method which is concise and effective. Despite of all the difficulties it seems that this thematic orientation could be used in the survey research as well as in different organizations.

3.5.3 Properties of the Emotional Intelligence Questionnaire (ECI)

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Abstract. Emotional intelligence is an appealing construct that has attracted a broad number of disciplines, such as psychology, clinical research, management and education. To fully study its impact, further research on emotional intelligence's conceptualization and operationalization is required. Goleman, Boyatzis and McKee (2002) proposed a framework to emotional intelligence that reflects how an individual express emotion and that is linked to emotional competencies.

Based on this approach, the present paper studies the properties of the Emotional Intelligence Questionnaire (ECI), a self-report and informants measure (Boyatzis and Sala, 2003). Participants were 195 MBA students at ESADE (47 women, 147 men; 103 of the Spanish Full Time MBA, 92 of the English Full Time MBA), ranging in age from 25 to 37. 178 students completed the self-version and the total sample obtained the others' ECI-U version completed by personal and professional observers.

Results contributed to the understanding of the connections between the different sources of information provided by the ECI-U and showed the relationships of the selected demographic variables and the emotional competencies of the sample. Results also commented the measurement subscales of emotional competencies for the Spanish and English version of the questionnaire.

3.5.4 Generating a new learning environment in management education: a comparison between an Spanish and a US MBA experience

Laura Guillén; ESADE, Universitat Ramon Llull, Spain

Joan Manuel Batista-Foguet; ESADE, Universitat Ramon Llull, Spain

Ricard Serlavos; ESADE, Universitat Ramon Llull, Spain

Abstract. Among the different iniciatives to overcome the evident crisis of traditional MBA designs(Clegg & Ross-Smith, 2003; Hogan & Warrenfeltz, 2003; Friga, Bettis, & Sullivan, 2003; Starkey, Hatchuel, & Tempest, 2004;Pfeffer & Fong, 2002; Sturges, Simpson, & Altman, 2003) the competency-based training approach is proving to be a promising strategy. Although many attemps, specially in the US, have been done to introduce this approach as a part of the program design and methodology, only a few schools have undertaken the challenge to combine this innovative efforts with a line of research on the actual effectiveness of this changes (Boyatzis, Cowen, & Kolb, 1995; Boyatzis, Stubbs, & Taylor, 2002; Camuffo & Gerli, 2004). Following the pioneer experience at the Weatherhead School of Management (Case Western Reserve University, Cleveland-Ohio), we have adapted and replicated the model at the ESADE Business School in Barcelona, which will make it possible to compare the application and effects of this innovative approach in an american and a european case. At the current stage of our project (the first cohort of MBA students that came accross to the new design have just been graduated a month ago), we can present the process, the reseach design and some preliminary results. Based on this experience, we will propose a series of questions concerning the measurement of competency-based learning process and the cultural issues involved in the comparison of the application of this models in different environments.

3.5.5 A longitudinal perspective to graduate management education: an integrated competency-career approach

Arnaldo Camuffo; University of Padua, Italy

Fabrizio Gerli; Ca' Foscari University of Venice, Italy

Francesca Chiara; Ca' Foscari University of Venice, Italy

Abstract. The effectiveness of the education provided by the most important business schools has recently been questioned by various studies (Connolly, 2003; Fisher, 2004; Mintzberg, 2004; Mintzberg and Gosling, 2002; Pfeffer and Fong, 2002, 2003 and 2004), that discuss students' preparation, the impact on their career generated by the education they received, the value of Master degree, the capability of the educational processes to create value, the qualitative level of the scientific research conducted within the schools, and their actual aptitude to innovate.

Building on the research stream on competencies and managerial education, this study performs a longitudinal analysis on five cohorts of an Italian MBA and aims at showing the methodology and the results of the implementation of an integrated competency-based approach to a management education process, which uses a set of competency-based tools during all the stages of the process itself, from training needs analysis to outcome evaluation, in order to measure the effectiveness of the education and, consequently, to raise the value it produces. In addition, it proposes a competency-career framework, centered on the alumni's careers tracking and analysis, which allows to evaluate the effectiveness of the MBA program in terms of alumni's working success and to determine which competencies are mainly related to a significant career improvement.

3.6   Questionnaire design and evaluation of the quality of questions

Session chair:
    Willem Saris; ESADE, Universitat Ramon Llull, Spain

This session takes place on:
    Thursday, July 21; 9:00 to 11:00
    Thursday, July 21; 11:30 to 13:30
    Thursday, July 21; 15:00 to 17:00

Presentations:

- Invited presentation: *Question characteristics and data quality: detection and correction for measurement error in survey research* - Willem Saris (Spain)

Contributed presentations:

- *The battle for survey methods in (UK) academia: A resource to spread the word* - Julie Lamb, Martin Bulmer (United Kingdom)

- *The Panel multitrait-multimethod design and model* - Anette Scherpenzeel (Netherlands), Willem Saris (Spain)

- *Measuring Subjective Well-Being: The Choice of a Response Format* - Martin Kroh (Germany)

- *Establishing data quality: an experiment to assess sample and questionnaire design effects* - Joost Kappelhof (Netherlands)

- *Development of survey questions using linguistic information* - Irmtraud Gallhofer, Willem Saris (Spain)

- *Are personality items answered differently depending on response scale format?* - Beatrice Rammstedt, Dagmar Krebs (Germany)

- *Influencing Error of Central Tendency in Self-rating Scales* - Jozef Výrost, František Baumgartner (Slovakia)

- *Context effects in the measurement of subjective probabilities in surveys* - Elisabeth Coutts (Switzerland), Rainer Schnell (Germany)

- *Combining Visual Analog and Categorical Scales* - Anna Stangl (Germany)

- *Thinking aloud about 'fear' and 'crime': An analysis of the reception of survey measures* - Jonathan P. Jackson (United Kingdom)

- *Methodological aspects in conceptualization and measurement of the fear of crime* - Ines Ivičić, Renata Franc, Vlado Šakić (Croatia)

- *Measuring insecurity and fear of crime by means of the European Social Survey* - Joanna Tsiganou (Greece)

- *Conceptual data modeling and subject matter ontology development* - John Kallas (Greece)

- *Materialism and Post-Materialism in Great Britain: a critical analysis* - Shinobu Majima, Mike Savage (United Kingdom)

3.6.1 Question characteristics and data quality: detection and correction for measurement error in survey research

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. In 1984 Frank Andrews published a paper indicating how Multitrait Multimethod or MTMM experiments can be used to estimate random errors and systematic errors in survey research. He also showed how one can use Meta analysis of MTMM experiments to detect the relationships between question characteristics and data quality indicators of questions. Since that time the approach has been improved (Saris and Andrews 1991), further studied (Saris and Albers 2003, Coenders and Saris, 2000, Saris, Satorra and Coenders 2004, Aussems 2004) and many more experiments have been done (Költringer 1995, Scherpenzeel and Saris 1997, Wouters 2005, Corten et al 2002). At this moment a data base exist of more than 1000 questions of which the reliability and validity is determined in MTMM experiments and the question characteristics have been coded. In a paper of Saris and Gallhofer(forthcoming) the relationships between these characteristics and the data quality of the questions is evaluated. Based on these relationships a program (SQP) has been developed to predict the quality of new questions given the characteristics of the question (Saris, Van der Veld and Gallhofer,2004 and Oberski and Saris, forthcoming).

This means that the quality of a question can be evaluated before the question is used in practice and so improvements can be made in the questions. Besides that the predictions can also be used in analysis of survey data to correct for measurement error. In this way repeated observations of the same concepts in order to estimate the measurement error can be avoided. This will save a lot of costs and makes the models less complex.

At the moment the European Social Survey has built in its design 6 MTMM experiments with 9 questions in each wave that is held in 22 countries or more. So each wave provides information about the reliability and validity 1000 more questions. This information will be incorporated in the database and the analysis will be repeated so that in the future we will have a program SQP that can predict the quality of survey questions in many different languages on the basis of a much larger set of questions. At the moment we have already data for more than 3000 questions.

3.6.2 The battle for survey methods in (UK) academia: A resource to spread the word

Julie Lamb; University of Surrey, United Kingdom

Martin Bulmer; University of Surrey, United Kingdom

Abstract. Key Words: Social Surveys, Question bank, Teaching Survey Methods, United Kingdom Social Surveys

In the United Kingdom large scale social probability sample surveys such as the Labour Force Survey are carried out outside of academia, mainly in the Office for National Statistics (the UK Government statistical agency) [http://www.statistics.gov.uk] or the National Centre for Social Research [http://www.natcen.ac.uk], which have large budgets and staff numbers to carry out increasingly large, complex and costly surveys. This means that academics are often far away from the survey research process and are not acquainted with cutting edge work. In these circumstances, teaching students, especially undergraduates, about surveys can be particularly difficult. It is well known that in the UK skills in quantitative sociology are scarce (See for instance House of Commons Science and Technology Committee Select Report 'Work of the Social and Economic Research Council' 2004; http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/13/13.pdf) and that typically students are more interested in using qualitative techniques for their own research work. How can greater interest in survey research be encouraged?

This paper will discuss the development and use of the Question bank, an innovative WWW resource which is used to teach students and researchers about UK social surveys produced by survey agencies such as the Office for National Statistics and the National Centre for Social Research. Bridging the gap between academia and survey agencies is one of our key aims. The Qb contains information on the survey data collection methods and other resources used in over 50 UK surveys as well as the full facsimiles of the questionnaire instruments used.

3.6.3 The Panel multitrait-multimethod design and model

Anette Scherpenzeel; University of Amsterdam, Netherlands

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract.   The multitrait-multimethod model is very useful for evaluating the quality of survey instruments. In the past, several large methodological research projects have been done based on this model (Andrews, 1984; Költringer,1985; Saris and Münnich, 1995; Scherpenzeel and Saris, 1997; Corten et al., 2002). The major drawback of the multitrait-multimethod model, however, is that it requires repetitions of the same questions. When such repetitions occur during a single interview, problems caused by memory effects and/or the respondent's annoyance may arise. A solution to these issues is to use a panel design that spaces the repetitions of questions across different waves. We demonstrate how the general MTMM model can be adapted to fit such a panel design and how such a panel design may provide even more relevant information than the classical design. With this design, the MTMM model can also be used to monitor changes over time. These changes over time may also be useful in determining the strength of opinions.

3.6.4   Measuring Subjective Well-Being: The Choice of a Response Format

Martin Kroh; German Institute for Economic Research (DIW), Germany

Abstract.   A consistent finding of social science research is that characteristics of persons' objective situation only moderately affect their level of subjective well-being. This paper claims that the weak empirical link between objective situations and subjective well-being is partly attributable to measurement error in different satisfaction scales. Drawing on data from a split ballot multi-trait multi-method experiment carried out in a methodological pretest to the German Socio-Economic Panel (SOEP), this paper aims at identifying a survey instrument that minimizes measurement error. Among the response formats most often used to measure subjective well-being, three alternative instruments are evaluated: an 11-, 7-point and a magnitude satisfaction scale. The set of investigated satisfaction domains range from respondents' satisfaction with their jobs, health, income and their lives in general as well as their satisfaction with the environmental situation and democracy. The analysis shows that the choice of a response format makes a notable difference in terms of measurement error: The 11-point satisfaction scale produces the highest validity and the magnitude scale produces the highest reliability. The paper also reports differences between the three formats in terms of their ease of administration. An analysis of the relationship between measures of persons' objective situation (i.e., health condition and income) and their level of subjective well-being (i.e., satisfaction with health and income) illustrates that the choice of response formats also affects substantive interpretations about the nature of persons' well-being.

3.6.5   Establishing data quality: an experiment to assess sample and questionnaire design effects

Joost Kappelhof; Social and Cultural Planning office (SCP), Netherlands

Abstract.   The Amenities and Services Utilization Survey (AVO) is a longitudinal household survey conducted every four years using a self completion questionnaire. Over the last couple of waves item non response has been increasing so the questionnaire needed to be improved to reduce the item nonresponse, but also because some questions were getting backdated. After a study of possible reasons for item nonresponse and misinterpretations of questions in previous AVOs the SC questionnaire has been redesigned and improved in 2003.

Furthermore, an experiment was developed to measure the extent of questionnaire design effect, because the AVO is a longitudinal survey and we didn't want to loose continuity. The experiment was conducted among an access pool because it involved repeated measures and it's difficult (and very expensive) to reach the same respondents in the field twice and getting them to do a questionnaire again. The experiment consisted of testing two versions (old and new) of the SC questionnaire in an access pool. The 'old' version is an exact copy of the 1999 SC-questionnaire, except for minor but necessary changes, for example currency. The 'new' improved version differs in lay out, skip patterns, ordering of questions, etc.

The respondents in the access pool were randomly assigned to one of three groups. The first group completed the 'new' questionnaire first and after a period of three months the 'old' version. The second group was given the 'old' version first and the 'new' after three months. A third group was given the 'new' version of the SC-questionnaire on both occasions.

This design enabled us to assess the effect of the changes in the questionnaire after controlling for time and sampling effects by using latent class analysis for categorical data. Furthermore the quality of changes made to the SC-questionnaire were evaluated by studying the reduction of item non response.

3.6.6 Development of survey questions using linguistic information

Irmtraud Gallhofer; ESADE, Universitat Ramon Llull, Spain

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. While a lot of attention has been paid in the social sciences to the different formulation of questions the linguistic basis of the variation has not been studied in much detail. In the paper we will present linguistic knowledge that can help the designer of survey questions to generate questions that measure the concept they want to measure. For this purpose we will look at the elementary structure of assertions in general and for social science concepts in particular. Next we will show the many different ways assertion can be transformed in requests for answers. Using these two step procedure we think that social science concepts can be operationalized in a valid way with a very high certainty.

3.6.7 Are personality items answered differently depending on response scale format?

Beatrice Rammstedt; Centre for Survey Research and Methodology (ZUMA), Germany

Dagmar Krebs; University of Giessen, Germany

Abstract. The paper will present a study investigating the effect of response scale format on the answering of a well-established personality scale, the BFI-10. The BFI-10 is a short-scale version of the widely used Big Five Inventory (BFI), assessing the Big Five dimensions of personality, namely Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience by two items each. Using a dependent sample design the ten items were to be answered using end-point labelled response scales ranging from the negative (labelled "1") to the positive pole (labelled "8") at Time 1. At Time 2 (approximately three weeks later) the respondents were split in two conditions. The same ten items were administered again with an eight-point scale ranging from the positive pole to the negative pole. In the first condition the positive pole was labelled "8"and the negative pole was labelled "1"; in the second condition the positive pole was labelled "1"and the negative pole was labelled "8". For both conditions we investigated whether scale means, standard deviations, and item intercorrelations differed substantially between the administrations.

3.6.8 Influencing Error of Central Tendency in Self-rating Scales

Jozef Výrost; Slovak Academy of Sciences, Slovakia

František Baumgartner; Slovak Academy of Sciences, Slovakia

Abstract. Error of central tendency (ECT) describes the tendency of raters to avoid extreme judgments. Efforts to counteract this error are directed usually to adjust the strength verbal description of scale points, or provide smaller differences in meaning between steps near the ends of the scale than between the steps near the center, and/or use more points in the scale.

For purposes of this research a short measure of self-esteem (SELF) was developed. 6 items (good luck, success, bunce, physical strenght, mental strenght, self-confidence) were rated twice: first on a classical five-point likert scale, and second on a four point scale, where the neutral point was removed. Raters was asked to compare themselves with people they know (e.g. if these people are more/less lucky than he/she is).

Empirical data confirmed increase in discriminative potential and reliability of the 4-points scales. Results are discussed with consideration to advances and limitations of such ECT influencing in methodological, social psychological, and ethical perspective.

3.6.9 Context effects in the measurement of subjective probabilities in surveys

Elisabeth Coutts; ETH, Switzerland

Rainer Schnell; University of Konstanz, Germany

Abstract. Although subjective probabilities are of great interest for many survey researchers, measuring those subjective probabilities is difficult. The drawbacks of using categorical scales with verbal labels to capture subjective probabilities are well documented. The use of a percent scale, while offering greater potential precision and the possibility of external calibration, is also problematic since there are predictable biases in the way people make such estimates. For example, subjective probabilities of rare events obtained in percent terms are often far too high. One potential solution is the use of a relative-frequency scale instead. Much evidence exists that the use of a relative-frequency scale promotes more accurate probabilistic reasoning than the use of a percent scale. However, phrasing questions on the subjective single-event personal probabilities in which survey researchers are often interested in frequency terms can be tricky. We therefore test the possibility that survey respondents can be taught to apply frequentistic reasoning to their estimates of single-event probabilities through a prior frequency-judgment task. Specifically, we examine the effect of the scale on which a base-rate estimate is made on a later percentage estimate of a single-event probability. This test was conducted in the context of a mail-survey experiment in which approximately 1,100 residents of a German city were asked about their fear of crime. The results provide evidence for the hypothesized reasoning transfer.

3.6.10 Combining Visual Analog and Categorical Scales

Anna Stangl; Ifo Institute for Economic Research, Germany

Abstract. Visual analog scales are rating scales in which the subject ranks his preferences along a continuous line or scale. In the majority of Business Tendency Survey (BTS) three-category scales are used because they are easy to handle and require a minimum of cognitive capacity on the side of the respondent. However, in a three or even five category scale the respondent is forced to make a decision towards a predefined category. The continuous analog scale enables scores between categories, which leads to more sensitive measures. While the realization of such a scale was difficult in the traditional paper-and-pencil surveys, the Web provides optimal conditions for an easy utilization and technical implementation. The study is based on a survey of about 800 independent economic experts from over 90 countries who were surveyed in April 2005 in the context of the World Economic Survey (WES). The study reports about the first experiences with the VAS instrument and compares the results received with the two different scales.

3.6.11 Thinking aloud about 'fear' and 'crime': An analysis of the reception of survey measures

Jonathan P. Jackson; London School of Economics, United Kingdom

Abstract. The fear of crime has long been considered a significant social problem by academics, policy makers and media commentators alike. The corpus of work surrounding terms like the 'fear of crime', 'emotional reactions to crime' and 'anxieties about crime' is vast, and, in sum, suggests that the fear of crime is widespread amongst members of many contemporary westernised societies. Anxieties about crime have negative effects for the individual and for society. They may exacerbate the impact of crime by damaging an individual's quality of life. They may affect the community by deteriorating a shared sense of trust, cohesion and social control, and thus perhaps contribute to the incidence of crime itself. And in the political climate, on issues such as border controls, asylum and immigration policy, public fears exert force on policy that balances security against liberty –the need for authoritarianism to secure safety against the need for liberalism to secure freedom.

Public perceptions of crime also influence specific Government policies. Anxieties about crime make themselves felt through public demands on the police to manage crime and its concomitant causes and effects. Clamour for more police; increasing calls for the Government to tackle anti-social behaviour; the seeming refusal of many people to believe that crime rates are not rising –all these evidence the influence of public perceptions of risk. And Governments respond: witness popular punitive Law and Order sloganeering, police strategies of reassurance, and community policing of incivilities and 'quality of life' issues. Indeed, public perceptions of risk have encouraged the police to focus on reassurance, sometimes at the expense of actual risk reduction.

Yet the knowledge base that emerges from public opinion surveys may be flawed. Several commentators have raised doubts surrounding the validity of the instruments used to generate these findings. A range of methodological problems has been identified which cumulatively raise the possibility that the incidence of the fear of crime has been significantly misrepresented and exaggerated by survey research. Certainly there is some supporting

evidence of this in both the specific literature on fear of crime and the survey methodology literature more generally. Deficiencies of standard measures would appear to lead to an inflation of self-reported fear levels relative to people's actual experiences of feeling unsafe or worried.

Problems of measurement are also problems of conceptual clarity. Even a cursory survey of the criminological literature reveals a body of knowledge that lacks theoretical specification and sophistication. Criminology has failed to illuminate the wide range of possible responses to crime and the risk of crime and the idea that some responses to risk may be functional rather than detrimental. It has lacked an account of the psychology of risk and vulnerability. And it has failed to consider the social and cultural significance of crime, disorder and conditions conducive to crime –factors that may be fundamental in shaping both perceived risk and the cultural significance of crime and social order in the public mind. This has left research unable to capture the full importance of fear of crime as an indicator of social conditions across Europe.

This paper summarises outstanding conceptual and methodological dilemmas regarding 'fear' and 'crime'. Data are then reported from a series of investigations of what goes on in respondents' minds when they answer survey questions about the fear of crime. One study used Cognitive Aspects of Survey Methodology techniques in a semi-structured interview setting, exploring what standard measures are tapping into, generating methodological and conceptual insights. Another study experimentally rotated question wording to assess key issues. Still another study fielded new measures, designed using lessons from the previous empirical work; the resulting data were modelled using confirmatory factor analysis techniques and structural equation modelling. Together these pieces of research identify a range of limitations of standard measures. They suggest that survey indicators lead respondents to express a wealth of responses to and perceptions of crime and social order through the lens of 'fear' and 'crime'. The paper finishes with some suggestions on how to improve measures and concepts.

3.6.12   Methodological aspects in conceptualization and measurement of the fear of crime

Ines Ivičić; Institute of Social Sciences Ivo Pilar, Croatia

Renata Franc; Institute of Social Sciences Ivo Pilar, Croatia

Vlado Šakić; Institute of Social Sciences Ivo Pilar, Croatia

Abstract. By now numerous researchers in the fear of crime tradition pointed out inconsistent and contradictory results due to differences in the conceptualization and standardized measurement of the fear of crime concept. Since the fear of crime has frequently resulted in misspecification of models and confounded variable problems, this research paper aims to review the literature on the main methodological problems in this field, and to make suggestions that will facilitate future research.

The most basic problem related to fear of crime is definition of the concept. By now, empirical evidences support the notion that perceived crime-risk and fear of crime, while related, are conceptually and operationally distinct concepts. Due to wide public availability as well as the utility for comparison of findings, studies usually use the same, although incorrect measures of fear of crime. Thus, in this research paper the validity of the fear of crime measures routinely employed in the surveys will be assessed. Related to the problem of fear of crime operationalization, are issues related to specification of crime and level of fear related to the type of victimization. This research paper will also outline other methodological considerations such as double-barreled and multi-dimensional measures of crime, confounded measures of perceived crime-risk and fear of crime, and the problem of single or double indicators of fear of crime.

Definitional and methodological issues will be discussed by using the data generated by recent research aimed at improving the conceptualization and measurement of fear of crime, and the measures employed in our own, nationally representative crime survey. Since there is inconsistency across studies in terms of statistical significance, direction, and magnitude of effects of fear of crime, systematic review and integration of micro and macro-predictors of distinct measures of fear of crime will be also presented in this study.

3.6.13   Measuring insecurity and fear of crime by means of the European Social Survey

Joanna Tsiganou; EKKE, Greece

Abstract. The scope of this paper is to exhibit certain "uncertainties"produced through experiences from supervising the "field"during the administration of the ESS in Greece, in both previous field-work "sweeps", coupled

with inquiries emerging out of the experience in supervising the cross-cultural comparability of the conceptual and cognitive aspects involved in the translation business of a pan-European measurement tool as the ESS. It seems that certain valid questions on the items of measuring the senses of insecurity and fear of crime (questions C5 and C6 of the Core Questionnaire of the ESS) still persist and they need further theoretical (but mainly empirically tested) elaboration .

The above items serve as an illustrative example in order to examine the substantive reliability of the above questions set, not only by means of the ESS survey tool, but also by numerous crime and victimization surveys around the (western) world. I consider the measurements intended of vast importance, since the quantitative data collected and properly processed, may be used for the deployment of new public safety and criminal policy measures, or the introduction of new aspects in the exercise of social control.

The proposed paper deals with the basic question of "what do we really want to know by posing these questions"and subsequently of "what do we actually learn by posing them"which is a question of correspondence of means to ends and vice-versa. In addition to these, I intend to investigate the grounds upon which the internal consistency of the placement of these questions as core items in the ESS questionnaire is served and suggest some recommendations for their completion.

3.6.14 Conceptual data modeling and subject matter ontology development

John Kallas; EKKE, Greece

Abstract. A general research procedure

Empirical research in social sciences is a well established general research procedure, which is divided into the following steps :

I. The initial transformation of a theoretical representation of social phenomena, into formalized information;

II. The choice of variables used to characterize each of the individuals of the population examined;

III. The choice of individuals who constitute the population examined

IV. The coding of the initial data characterizing each individual by the descriptive variables.

V. The methods of analysis employed to treat and to transform the data in order to furnish formal and often statistical results;

VI. The final transformation by the researcher of formal mathematical results into the final results that are presented in a public discourse or in a scientific text;

Data modeling

The choice of variables is based on the hypothesis that a social phenomenon can be represented through a limited number of variables. We argue that we may improve formalization by introducing data modeling techniques, and at the same time, by shifting the conceptual frame of reference from the level of the research project to the level of a research program related to a subject matter field.

Until now, the formalisation of steps II, III, IV and V is based on the use of specific statistical structure as tool of formalisation, a statistical ontology, which defines three major object types: the statistical unit, the population and the statistical variable. Thus in every survey, we define the main object types (entities) according to this ontology which permits the use of the statistical methodology and tools for the analysis of the phenomenon we examine. For a better formalisation of step II we have to define the basic object types from a subject-matter point of view also, using a subject matter ontology that remains constant in the conceptual frame of a specific subject matter field.

Building a subject matter ontology

One of the main limitations of the existing metadata models is that the main units of observation and their attributes (variables) are modeled in the context of a study. If we want to establish comparability between different studies that belong to the same subject matter field, we need to establish common definitions for the same units of observations among the studies. To deal with this problem a unit of observation must be represented by an object type and its attributes, which must be element of a common subject matter ontology. This will help us to end with subject matter ontology for each research field. This ontology will be implemented in the context of a subject matter field and not of a study. When we design a new study we choose that its units of observation be described as versions of objects of observations already existing in the ontology. Otherwise we must add new objects. Thus each variable must be linked to an attribute, which is linked to ontology. Thus we have a differentiation between an attribute and a variable. An attribute is a variable that is standardized for a number of studies and that it is linked to a number of objects.

An infrastructure to support this type of design

The need for survey design in the context of a subject matter field instead of the context of a study leads to the evolution of new research practices. A number of research teams that work independently in the design of different studies may share the same concepts, the same variables and the same questions to build totally different data models and questionnaires. To achieve this goal we need a common distributed research infrastructure that support metadata sharing. The existing CESSDA Archives will be the nodes of this new, networked infrastructure, which will support the following functionality:

    o Data discovery and dissemination for secondary analysis

    o Building a subject matter ontology for specific research fields

    o Secondary data production

    o Support new study design

    o Supplementary research documentation

3.6.15   Materialism and Post-Materialism in Great Britain: a critical analysis

Shinobu Majima; University of Manchester, United Kingdom

Mike Savage; University of Manchester, United Kingdom

Abstract.   In the UK there has hitherto been relatively limited interest in exploring Inglehart's claims about the shift from materialist to post-materialist values in the post-war period. Indeed, dominant themes were raised by historians and sociologists focus on the growing significance of materialist concerns of affluence and consumerism since the 1950s. This paper provides a detailed examination of the post materialism argument in the British context.

The paper begins with a systematic analysis of the existing literature. We focus on whether materialist and post-materialist measures group together attitudes that are otherwise disparate. Using the British data, collected in 1981, 1990 and 1999 for World Values Surveys and European Values Surveys, we examine how far, in the UK context, attitudes taken to exemplify materialist and post-materialist values might be separated on different axes, such as that between libertarianism and authoritarianism; trusting and non trusting, which have been discussed in the British context.

In developing this analysis, we use a little known technique, correspondence analysis, which allows us to visualize how different attitudes are related to each other by portraying them in two-dimensional space. This method shows how attitudes do indeed break down in ways which problematise the argument that there has been a shift from materialist to post-materialist values.

The paper concludes by using qualitative data collected in the 1940s, the period of post war austerity and scarcity in the UK, to show that materialist concerns were always refracted through moral and political concerns. The paper concludes by arguing that we especially need to rethink the definition of materialism, so that it recognizes that material concerns are always wrapped in moral codes.

3.7   Behavior coding

Session chair:
    Wil Dijkstra; Vrije Universiteit Amsterdam, Netherlands
    Yfke P. Ongena; Vrije Universiteit Amsterdam, Netherlands

This session takes place on:
    Thursday, July 21; 17:30 to 19:30

Presentations:

- Invited presentation: *Behavior coding: New perspectives* - Wil Dijkstra, Yfke P. Ongena (Netherlands)

Contributed presentations:

- *Expectations of Interviewers as Determinants of Deviant Interviewer Behavior* - Johannes van der Zouwen, Johannes Smit, Marleen van der Horst (Netherlands)

- *Cross-Cultural Behavior Coding: Using the 2003 California Health Interview Survey to Assess Cultural/Language Data Quality* - Elaine Zahnd, Tammy Tam, Nicole Lordi, Gordon Willis, W. Sherman Edwards, Stephanie Fry, David Grant (United States)

- *Verbal Behaviors in Event History Calendar and Question-List Interviews: Interactions with Respondent Characteristics* - Robert F. Belli, Mario Callegaro, Dan Liao (United States)

- *Comparing Behavior Coding and Cognitive Interviewing as Assessment Tools for Survey Questions on Health Topics* - W. Sherman Edwards, Kerry Levin (United States)

3.7.1 Behavior coding: New perspectives

Wil Dijkstra; Vrije Universiteit Amsterdam, Netherlands

Yfke P. Ongena; Vrije Universiteit Amsterdam, Netherlands

Abstract.   The first applications of behavior coding concerned primarily monitoring interviewer performance, i.e. question reading, adequate probing, etcetera. Later on, questionnaire evaluation became the most important goal of behavior coding schemes. Since Brenner's (1982) scheme, developed to analyze the interaction between interviewer and respondent, including sequential analyses, this goal too became important. Of the 48 different behavior coding schemes that have been developed since Cannell et al.'s first scheme of 1968, the interviewer was the main focus in 30% of the schemes, the questions in 44%, the respondent in 5% and the interaction between the interviewer and respondent in 21% .

Because behavior coding is directed towards observable behavior, cognitive interviewing is often used to obtain insight in the more covert cognitive problems respondents may have in answering survey questions. A disadvantage of cognitive interviewing however is that respondents may report non-existing problems, and that cognitive interviewing procedures like thinking aloud may affect the cognitive process itself.

A main disadvantage of behavior coding on the other hand is that existing problems may remain undetected. In case a respondent just gives an adequate answer to a survey question, this does not mean that the respondent doesn't experience problems, or grasps the meaning of the question as intended by the researcher.

However, traditionally, behavior coding concerns only verbal behavior. If we incorporate in our behavior coding schemes paralinguistic and visual behavior, much more information will become available about the cognitive processes and problems respondents may encounter when answering survey questions. Such information does not, or to a lesser degree, suffer from the above mentioned problems of cognitive interviewing (i.e. reactivity and detecting non-existing problems). In our paper we will illustrate, how the combination of behavior coding of verbal utterances and visual behavior (like the respondent looking to or away from the interviewer) as well as paralinguistic behavior (e.g. latency times, or the use of doubt words) can provide more insight in cognitive processes than traditional behavior coding.

3.7.2 Expectations of Interviewers as Determinants of Deviant Interviewer Behavior

Johannes van der Zouwen; Vrije Universiteit Amsterdam, Netherlands

Johannes Smit; Vrije Universiteit Amsterdam, Netherlands

Marleen van der Horst; Vrije Universiteit Amsterdam, Netherlands

Abstract.   If in a standardized survey interview, interviewers have to pose difficult-to-answer questions, they will have to do their very best to guide and encourage the respondents to come up with adequate, that is, substantive, relevant and precise, answers. To do so, interviewers have different means at their disposal. Some are 'legal', like repeating the question, providing 'authorized' explanations, and neutral further probing. Others are 'illegal', that is contrary to the instructions of the researcher, like suggestive probing.

A striking form of deviant interviewer behavior is substantially rephrasing the questions of the questionnaire. For example, instead of asking a question about a whole set of household activities (the prescribed 'integral' questioning strategy), the interviewer decomposes the question into a series of sub-questions, each related to a specific task (the deviant 'partial' questioning strategy).

One possible explanation for such deviant interviewer behavior is that the interviewer is biased towards answers to be expected from the respondent. As an example from a survey among the elderly: why ask a frail looking 85-year old respondent the question "How many days did you spend on repairing or painting in the house during the past two weeks?"It is much easier to assume that this is not the case, and rephrase the question into: "So you did not do any repairing or painting in the house the past two weeks?"Or, even not asking it at all, because the interviewer expects a "No"answer anyway. Another explanation for 'illegal' behavior is that this is just a reaction to inadequate respondent behavior towards difficult-to-answer questions, which has to be solved by the interviewers with unorthodox methods.

In the paper we will go deeper into the issue whether the deviant behavior of the interviewers is a reaction to inadequate response behavior, or rather an anticipation of the answers these respondents will provide when

answering the questions of the questionnaire. For this purpose, we analyzed transcripts of fragments of 326 interviews with elderly respondents (65+) about their performance of light and heavy household tasks.

For the analysis of these transcripts we used a coding scheme with codes for different forms of deviant interviewer behavior. To investigate the relationship between interviewer behavior and respondent behavior, the coding scheme contained codes for inadequate respondent behavior (like giving 'don't know' answers, or imprecise answers), and for inadequate interviewer behavior (like suggestions, or the omission of (parts) of questions).

An important form of deviant interviewer behavior proved to be that the interviewer did not mention all tasks belonging to the set of light or heavy household tasks as printed on the questionnaire. These omissions were clearly related with gender and age of the respondent. The more physical demanding tasks were more often left out with elderly respondents; also specific gender related tasks (e.g., ironing or repairing and painting) were more often left out, dependent on the gender of the respondent.

3.7.3 Cross-Cultural Behavior Coding: Using the 2003 California Health Interview Survey to Assess Cultural/Language Data Quality

Elaine Zahnd; Public Health Institute, United States

Tammy Tam; Public Health Institute, United States

Nicole Lordi; Public Health Institute, United States

Gordon Willis; National Cancer Institute, United States

W. Sherman Edwards; Westat, United States

Stephanie Fry; Westat, United States

David Grant; UCLA Center for Health Policy Research, United States

Abstract. Background: Behavior coding has been utilized to study problems of how well particular questions perform across different surveys. Employing behavior-coding techniques cross-culturally, however, is relatively rare. This National Cancer Institute (NCI)-funded study utilized behavior coding techniques to evaluate how well particular questions work across cultures and across languages.

Methods: Behavior coding was conducted using the California Health Interview Survey 2003 (CHIS 2003), a RDD CATI survey jointly administered by UCLA, the California Department of Health Services and the Public Health Institute. CHIS 2003 was administered to 42,044 adults in six languages. For the behavior coding study, we randomly sampled five cultural/language groups (Latinos/English, Latinos/Spanish, Koreans/English, Koreans/Korean, Other/English) to record and code parts of the interview. Cases were independently re-coded to assess inter-coder reliability. Logistic regression was used to compare three language/cultural groups to determine the main predictors of respondent question answering behaviors. We eliminated Korean/Korean and English/Korean to tease out subtle differences between the other three, more comparable respondent groups.

Results: The final sample included 494 interviews with cooperation rates ranging from 88-91% for the English-language groups, 80% for the Spanish-language Latino group, and 59% for the Korean/Korean group. Between 86% and 89% of the questions were read as written in Spanish and in English, however, about one third of the Korean/Korean questions were not. Korean language interviews had the highest rate of "inadequate" respondent answers, a finding that might be expected given the high degree of Korean language misreads. Demographics (i.e., age, gender, poverty, citizenship, education), interviewer behaviors, and language (English or Spanish) were entered into four logistic regression models as independent variables with three of the cultural/language groups. A query asking about "number of days your physical health was not good in the past 30 days" did not perform as well with respondents with lower economic/educational backgrounds (p<.0001). Respondents who were older had more difficulty giving the number of days their mental health was not good in the past 30 (p<.001); similarly, altering this question's wording predicted respondent problems (p=.02). Older respondents (p=.02) and those at higher poverty levels (p=.003) were significantly more likely to have difficulty answering if poor health interfered with their usual activities. Of the four models, culture/language predicted respondent problems in only one case. Being female (p=.02), older (p=.002) and if you were Latino responding in English (versus non-Latino responding in English) (p=.0006) all predicted difficulties in answering a social support question. Demonstrating the complexity of behavior coding, all four models (questions) resulted in different explanatory factors, although age was

a fairly common predictor. Overall interpretation is also difficult due to the multiple factors that might impact the findings, such as translation, acculturation and varying cultural norms, interviewer training and supervising across languages, and varying coder performance. Many of these factors are not easily included in logistic regression models. Qualitative debriefings proved invaluable in teasing out possible factors influencing the results, such as the need for levels of politeness in asking questions of certain cultures. Cross-cultural behavior coding techniques appear to be valuable in assessing questionnaires administered in multiple languages and across different cultures. Further exploration using such methods with other surveys and language/cultural groups is needed, as is further cognitive and qualitative work, if we are to improve questionnaire cultural equivalence.

3.7.4 Verbal Behaviors in Event History Calendar and Question-List Interviews: Interactions with Respondent Characteristics

Robert F. Belli; University of Nebraska, Lincoln, United States

Mario Callegaro; University of Nebraska, Lincoln, United States

Dan Liao; University of Nebraska, Lincoln, United States

Abstract. Recent work has been showing the benefits of Event History Calendar (EHC) methodologies in promoting higher quality retrospective reports in comparison to traditional standardized question list (Q-list). Theoretically, these advantages over Q-list interviews are the result of EHC interviews promoting the greater use of effective parallel and sequential retrieval cues that are available in the structure of autobiographical memory in the context of a more motivational conversational exchange between participants. A verbal behavior coding analysis of 218 EHC and 197 Q-list paper & pencil interviews revealed significantly more behaviors in the EHC condition that indicated the use of timeline retrieval strategies and conversational engagement. In addition, better data quality was associated with a higher prevalence of retrieval cues, a greater degree of response openness, and lower levels of cognitive difficulty and rapport. The association of data quality and verbal behavior also interacted with method: retrieval cues and cognitive difficulty was directly associated with EHC response quality and indirectly associated with Q-list quality; rapport behaviors had more detrimental impact on Q-list data quality.

In an extension of this verbal behavior coding research, the respondent characteristics of age, gender, and race were included in separate analytic models to examine whether the types of verbal behaviors differed as a function of both type of interview (EHC, Q-list) and the respondent characteristics. In models that include variables to control for the length and complexity of interviews, a number of significant type of interview x respondent characteristic interaction effects were found.

Most of the interaction effects with age involved retrieval strategies in which some strategies such as those involved in sequential retrieval and in determining how many times events had happened occurred more often with older respondents in the EHC condition in comparison to the Q-list. Other retrieval strategies including parallel probing occurred more often with older respondents in the Q-list in comparison to the EHC. The presence of a third party person and respondent laughter both occurred more often among older participants in EHC interviews compared to Q-list ones.

All of the significant interaction effects with gender involved behaviors that occurred more frequently among women participants in the EHC condition compared to the Q-list. These behaviors included parallel probing, the use of examples, interviewer verifications, and respondent agreement with verifications.

The interaction effects with race of respondents all involve rapport components which occur more frequently among non-whites in EHC interviews in comparison to Q-list ones. These behaviors include the use of examples, acceptable feedback, interviewer laughter, and interviewer and respondent digressions.

In summary, results demonstrate that EHC and Q-list interviews differentially affect the manner of interviewing dependent on the characteristics of the participants. Because the complexity of interviews had been controlled in the models, these differences between the EHC and Q-list appear as a direct association with the respondents themselves. We can only speculate as to the underlying causes of these effects. For example, the greater conversational approach of EHC interviews may have permitted a less constrained approach in interviewing non-whites. Results for age and gender are not as clear.

3.7.5 Comparing Behavior Coding and Cognitive Interviewing as Assessment Tools for Survey Questions on Health Topics

W. Sherman Edwards; Westat, United States

Kerry Levin; Westat, United States

Abstract. Behavior coding and cognitive interviewing are two common tools for pretesting survey questions. The first published application of behavior coding to survey research was by Cannell et al (1968). They developed a set of codes for interviewer and respondent behavior to assess their effects on the quality of survey responses. Some 25 years later, Oksenberg et al (1991) reported on "New Strategies for Pretesting Survey Questions,"describing essentially the same coding system applied to a somewhat different purpose. The application of techniques of cognitive psychology to survey methods was first documented by Jabine et al (1984). Since then, a set of fairly standard approaches to testing survey questions with small numbers of respondents has developed under the general heading of "cognitive interviewing."A taxonomy for these approaches is described by Forsyth and Lessler (1991); Willis (2005) has described detailed approaches to implementing them.

The two techniques are similar in that they are relatively easy to implement and provide intuitively appealing information about survey questions very quickly. Behavior coding provides relatively general quantitative information on interviewers' and respondents' difficulties with question items, supplemented by the insights of the coders who observed and classified the behaviors. Cognitive interviewing provides more qualitative insights into how respondents understand survey questions and how they go about answering them. Either technique alone, or in combination with other pretesting methods, can point to the need to refine survey questions and provide some insight on how such refinements should be made.

Both behavior coding and cognitive interviewing have been developed well beyond the relatively simple approaches commonly used for pretesting survey questions to study interviewer-respondent interactions and respondents' thought processes in answering questions. This paper, however, will cover only the most common behavior coding and cognitive interviewing techniques for assessing survey questions. First, we will briefly review the literature, comparing the benefits and limitations of each approach. Then, we will provide examples of the results of each approach from similar studies funded by the U.S. National Cancer Institute (NCI).

Westat has conducted a series of studies testing survey questions related to health behaviors, cancer screening, and related topics with respondents in multiple languages. Specifically, we will compare results from a behavior coding study of sections of more than 500 2003 California Health Interview Survey interviews conducted in English, Spanish, and Korean with results from cognitive testing of questions on diet and smoking, conducted in English, Spanish, Korean, Vietnamese, and Chinese. Only a few specific items overlap between the two studies, but we will classify the item-specific findings from each approach and each language within approach, and compare the nature and frequency of problems identified and recommended changes. The objective will be to identify differences between the testing techniques, and to tie these differences to the basic differences in the techniques.

References:

Cannell, C., Fowler, F., and Marquis, K. "The Influence of Interviewer and Respondent Psychological and Behavior Variables on the Reporting in Household Interviews."Vital and Health Statistics: Data Evaluation and Methods Research, Series 2, Number 26. National Center for Health Statistics, 1968.

Forsyth, B., and Lessler, J., "Cognitive Laboratory Methods: A Taxonomy,"in Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S. (eds.), Measurement Errors in Surveys, New York: John Wiley and Sons, Inc., 1991.

Jabine, T.B., Straf, M.L., Tanur, J.M., and Tourangeu, R. (eds.). Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines, Washington DC: Naitonal Academy Press, 1984.

Oksenberg, L., Cannell, C., and Kalton, G. "New Strategies of Pretesting Survey Questions,"Journal of Official Statistics, 1991, Vol. 7, No. 3, pp. 349-366.

Willis, G., Cognitive Interviewing: A Tool for Improving Questionnaire Design, Thousand Oaks: Sage Publications, 2005.

3.8   Cognitive lab experiments

Session chair:
>   Ger Snijkers; Statistics Netherlands (CBS), Netherlands
>   Jose Luis Padilla Garcia; University of Granada, Spain

This session takes place on:
>   Friday, July 22; 9:00 to 11:00

Presentations:

- Invited presentation: *Cognitive Laboratory Methods: An Overview* - Ger Snijkers (Netherlands)

Contributed presentations:

- *The usefulness of cognitive interview to pretest adapted questionnaires in cross-cultural studies* - Alba-Saida García Negro, Jose Luis Padilla Garcia, Manuel Sainz Quevedo, Miguel Castillo Diaz (Spain)

- *Using cognitive interviews to improve the design of indirect respondent protocols on "usual and actual hours worked" in the Spanish Labour Force Survey* - Isabel Esteban Cantador (Spain)

- *What do the respondents understand self-rated health to mean? Inferring the purpose for asking the survey question* - Anja Ahola (Finland)

3.8.1 Cognitive Laboratory Methods: An Overview

Ger Snijkers; Statistics Netherlands (CBS), Netherlands

Abstract.   Cognitive pre-testing is aimed at developing a questionnaire that is respondent friendly and results in valid data. With cognitive pre-testing the question-and-answer process is investigated. Pre-test research shows where the questionnaire can result in unintended interpretations and invalid answers. In addition, the questionnaire can be improved.

To evaluate questionnaires in a cognitive laboratory a number of pre-test methods have been developed. These methods are: expert appraisal, focus groups, in-depth interviews, and behavioural coding. Within in-depth interviews techniques like thinking aloud, follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes can be used. These methods can be used to evaluate paper, CAPI, CATI, and Web questionnaires.

These methods are presented within the context of a 5-step pre-test model for data collection development. For each step in the process of developing a data collection procedure, the model indicates what methods can be used. In a full pre-test program all aspects of the survey are carefully tested in advance.

3.8.2 The usefulness of cognitive interview to pretest adapted questionnaires in cross-cultural studies

Alba-Saida García Negro; San Cecilio Hospital (Granada), Spain

Jose Luis Padilla Garcia; University of Granada, Spain

Manuel Sainz Quevedo; San Cecilio Hospital (Granada), Spain

Miguel Castillo Diaz; University of Granada, Spain

Abstract.   The main aim of questionnaire adaptation is to obtain equivalent versions in different languages. Recently, researchers are increasingly interested in studying the usefulness of cognitive pretesting to evaluate adapted questionnaires. Cognitive interview was applied to Spanish version of LittlEARS questionnaire. The LittlEARS questionnaire has been designed to describe auditory development in hearing and hearing-impaired children under 24 months after cochlear implantation or hearing aid. The LittlEARS questionnaire is "survey like". Parents are asked to respond to 35 dichotomous items answered "yes"or "no"considering the auditory behaviour of their children. Parents play as "proxies". Quality standards applied during the adaptation process have paid special attention to the translation design used, as well as to the use of empirical procedures to provide evidence of the degree of equivalence achieved between the original and adapted versions. The present study aimed to show how cognitive interviews can be useful in providing information about the meaning of translated items. Cognitive interviews are highly structured interviews designed to provide information about people's "question/response"process with respect to the questionnaire items. The data obtained from 30 cognitive interviews with parents of children with hearing deficits are presented. Parents were interviewed separately as part of the evaluation protocol of the Hearing Loss Intervention Programme of the San Cecilio University Hospital (Granada, Spain). The interviews were conducted in accordance with a protocol designed to provide information about the meaning of different items. The items to be analysed were selected by a panel of experts. The changes made to the items in light of the data obtained from the cognitive interviews are described. Finally, the conditions for including cognitive interviews into the process of test and questionnaire adaptation are discussed.

3.8.3 Using cognitive interviews to improve the design of indirect respondent protocols on "usual and actual hours worked"in the Spanish Labour Force Survey

Isabel Esteban Cantador; Statistics Spain, Spain

Abstract.   The main objective of the Statistics on working time is to reveal the characteristics of the labour market. The Spanish Labour Force Survey is the principal source of information on working hours in Spain. The LFS asks respondents to provide direct and indirect (proxy) information on "usual and actual hours worked", as well as overtime during the survey reference week.

Accordingly to the changes that have been in the Spanish LFS in 2005, various pre-tests have been carried out to improve the quality of working time measurement. The present paper takes into account results of using

cognitive interviews,they were realized to direct and indirect (proxy) respondents, to improve the formulation of questions concerning actual working time and usual working time. This research has consisted on the study of 57 cognitive interviews, whose respondents were selected on the basis of socio-demographic and educational characteristics of particular interest for the study: immigrants, young people, employees, housewives and people aged over 65. The research was developed a similar interview protocol, it was designed to obtain information on the understanding of the concepts and the formulation of answers to questions about "usual and actual hours worked", comparing indirect and direct discourses.

Because of proxy report is allowed in several surveys conducted by the Spanish Statistics National Institute conduct surveys, the goal of this paper is to present approaches to the design of cognitive interview protocols for indirect respondents.

3.8.4   What do the respondents understand self-rated health to mean? Inferring the purpose for asking the survey question

Anja Ahola; Statistics Finland, Finland

Abstract.   The paper sets out from the premise that factors which affect the way questions work cannot be understood in isolation from the context in which they are employed. Besides the respondents' ability to understand the question, their answers are also influenced by their ideas of the purpose for the question: why is it asked, what will the answers be used for and by whom (see Schwarz 1996; Alasuutari 1998; Briggs 2002). Thus, answers to survey questions may reflect both what the respondent presumes the researcher wants to know, and what he or she assume the researcher will do with the information they give.

I studied how the interviewees' understand the meaning of the question that is used as an international indicator measuring self-rated health. How do they decide what to answer? Does explaining in advance that the data are collected for an EU survey or for a national study influence their replies? How do the respondents define the data collection situation?

Twenty-three persons were interviewed for the study, drawn from the area of Helsinki to represent as evenly as possible men and women from all socio-economic groups. The sample was divided in two groups. One was given to understand that the data would be used by the EU and the other that they would be used for national purposes. Two questions used as indicators, one about self-rated health and the other about pay, were embedded into a structured questionnaire inquiring about other living conditions. Understanding of the questions was tested with the cognitive interviewing method (retrospective think-aloud and probes). Defining of the data collection situation was studied with the semi-structured interviews. The transcribed data was analysed with a qualitative content analysis.

The question measuring self-rated health read: "How would you rate your health in general? Is it excellent, good, average, poor or very poor?"It was easy for the respondents to understand. The respondents justified their choice of response alternative by absence or presence of ill-health, but also by personal experiences and life situation, and by certain action. Many also based their assessment on their fitness or health behaviour.

Norenzayan and Schwarz (1999) have presented empirical results about how the institution the researcher represents influences inference of the meaning of a question. However, according to this study the respondents did not differentiate between the given reasons for the data collection, i.e. EU indicators or national indicators, when explaining why they had decided to participate. By contrast, their ideas about what the data would be used for were clearly influenced by how important and confidential they regarded the data collection situation. Whether the respondents believed that the collection of the data and the purpose for which they would be used were important was clearly influenced by the surveying organisation. Thus, the factor influencing the perceived importance of the interview was the organisation conducting it –an official government agency –and the mental images associated with its data use. Many regarded participation in an interview conducted by Statistics Finland as almost a civic duty.

The results depict the general trust people in Finland have in collections of data for official indicators. The replies were characterised by a general belief that such indicators promote development in society or at least do no harm. They had trust in the confidentiality of the data collection situation and anonymity in the use of the obtained answers. It may be that answering is bound by different orientations in different cultures. Because the indicators are used for international comparisons, it would be important to know how respondents in different countries define the situation where data are collected for the purpose of compiling an indicator on health.

3.9   Measurement of social networks

Session chair:
  Anuska Ferligoj; University of Ljublana, Slovenia

This session takes place on:
  Friday, July 22; 11:30 to 13:30

Presentations:

- Invited presentation: *Reliability and Validity of Social Support Measurement Instruments* - Anuska Ferligoj, Tina Kogovsek, Valentina Hlebec (Slovenia)

Contributed presentations:

- *Data quality of generated names in a self-administered survey on social capital* - Bettina Langfeldt, Dagmar Krebs, Juergen H.P. Hoffmeyer-Zlotnik (Germany)

- *Who is missing? A study on relational characteristics of nonrespondents to social network surveys* - Daniëlle De Lange, Filip Agneessens, Hans Waege (Belgium)

- *Social Network Measures for "Nosduocentered"Networks, their Predictive Power on Performance* - Luis Coromina (Spain)

3.9.1 Reliability and Validity of Social Support Measurement Instruments

Anuska Ferligoj; University of Ljublana, Slovenia

Tina Kogovsek; University of Ljublana, Slovenia

Valentina Hlebec; University of Ljublana, Slovenia

Abstract. The paper presents an overview of several studies dealing with the quality of complete social support networks and ego-centered social support networks, collected by surveys.

The most important findings of these studies for complete networks are:

1. The binary scale as a measure of tie strength is the least reliable, and the five pont ordinal scale with labels the most reliable.

2. Free recall and recognition data collection yield equally reliable data.

3. Original and reciprocated questiones also yield equally reliable network data.

In the case of ego-centered social support networks the most important results are:

1. The telephone mode gives more valid data than the face-to-face mode and also quite reliable network data.

2. The data collection 'by alters' gives more reliable data than the data collection 'by questiones'.

3. Larger networks have lower validity than smaller ones.

4. The ties in the core network are measured more reliably and more validly than those in extended networks.

5, Some demographic ego's characteristics (gender, age) also affect social support network data quality.

3.9.2 Data quality of generated names in a self-administered survey on social capital

Bettina Langfeldt; University of Giessen, Germany

Dagmar Krebs; University of Giessen, Germany

Juergen H.P. Hoffmeyer-Zlotnik; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract. Researchers interested in ego-centered networks usually collect data by asking respondents about the alters they count on for multiple kinds of support to ensure that the various dimensions of social support situations are represented and that the resulting measures of characteristics of the ego-centered network represent a complete enumeration of network members. But name generators as one instrument for the measurement of social capital vary widely in the reliability of the measures they produce and most of them are tainted with one or the other problem. This is especially true for name generators used in self-administered modes where respondent's burden is very high.

On the basis of a study on social capital of PhD-students we aim to shed light on what happens when respondents use name generators and list alters without interviewer help or interference in a self-administered questionnaire. We point out typical shortcomings of this method by analysing the data as well as by reporting the results of a cognitive pretest conducted in advance of the main study. In a next step we will suggest different ways of data preparation for lists of names won by a self-administered name generator and make some comments on the impact of data preparation on data quality of ego-centered networks.

Finally we will show what kind of analysis is possible with our data and report some findings of the study on social capital of PhD-students.

3.9.3 Who is missing? A study on relational characteristics of nonrespondents to social network surveys

Daniëlle De Lange; University of Gent, Belgium

Filip Agneessens; University of Gent, Belgium

Hans Waege; University of Gent, Belgium

Abstract. Little is known about the characteristics of unit nonrespondents in surveys dealing with social network questions. The aim of this paper is to get an idea about the extent to which nonrespondents differ in characteristics from respondents. The main idea is based on the social isolation hypothesis (Goyder, 1987; Groves and Couper, 1998). It states that people isolated in society or peripheral to society will be less inclined to cooperate with a

survey request than people being more central. Two approaches will be used to look for differences in centrality-based network characteristics between respondents and nonrespondents. First, as nonrespondents can be regarded as similar in characteristics to late or reluctant respondents, social network characteristics of people who respond only after a second or third reminder will be compared with characteristics of people responding in the first phase of the follow up. Second, we will examine differences in network characteristics between respondents from all follow up phases and the final nonrespondents. Complete network questions offer the unique possibility to estimate some relational characteristics of these nonrespondents. Based on the social isolation hypothesis, we expect that people being less central in different types of social networks (advice, cooperation, work support, friendship and trust) will be more inclined to answer late in the follow up procedure, or will be more prone not to respond. On the group level, it is expected that more dense and less centralised networks will encourage survey cooperation for all team members. For both approaches, we find that individual and group level effects are significant. As such, network researchers should be aware that the people who are missing in a social network survey have different network characteristics from those people who do respond.

3.9.4   Social Network Measures for "Nosduocentered"Networks, their Predictive Power on Performance

Luis Coromina; University of Girona, Spain

Abstract.   Our purpose in this article is to define a network structure which is based on two egos instead of the egocentered (one ego) or the complete network (n egos). We describe the characteristics and properties for this kind of network which we call "nosduocentered network", comparing it with complete and egocentered networks. The key point for this kind of network is that relations exist between the two main egos and all alters, but relations among others are not observed. After that, we use new social network measures adapted to the nosduocentered network, some of which are based on measures for complete networks such as degree, betweenness, closeness centrality or density, while some others are tailor-made for nosduocentered networks.

We specify three regression models to predict research performance of PhD students based on these social network measures for different networks within a research group such as advice, collaboration, emotional support and trust. Data used are from Slovenian PhD students and their supervisors.

The results show that performance for PhD students depends mostly of the emotional network, because it is significant for all three models. Trust and collaboration networks are significant for two models and advice is not significant for any model.

As regards network measures, classic and tailor-made measures are about equally good. Measures related to the total intensity of contacts (e.g., density, degree centralization and size) seem to work best to predict performance.

Finally, we used regression models in order to know whether egocentered or nosduocentered networks explain performance best. One regression was done including nosduocentered network variables, another egocentered network variables and a third model including both networks. Nosduocentered network alone leads to a higher R2 and thus has a higher predictive power for performance.

3.10   Measurement of Social Capital

Session chair:
    Beate Völker; University of Utrecht, Netherlands

This session takes place on:
    Friday, July 22; 15:00 to 17:00

Presentations:

- Invited presentation: *Better a good neighbor than a faraway friend? Trends in neighborhood social capital in several countries* - Beate Völker (Netherlands)

Contributed presentations:

- *How Much Social Capital Do You Have? Measuring social capital at the individual level in social surveys* - Keming Yang (United Kingdom)

- *Measuring Social Capital in Europe* - Christian Haerpfer, Florian Pichler, Claire Wallace (United Kingdom)

- *Converting a CAPI module on social capital into a postal format* - Gerry Nicolaas, Sarah Tipping, Abigail Dewar, Amanda Wilmot (United Kingdom)

- *Social Capital and Academic Performance among First Graders: Individual vs. Class Social Capital* - Ronan Van Rossem (Belgium), Marjolijn Vermande, Chris Baerveldt, Beate Völker (Netherlands)

3.10.1   Better a good neighbor than a faraway friend? Trends in neighborhood social capital in several countries

Beate Völker; University of Utrecht, Netherlands

Abstract. Since the writings of Robert Putnam (e.g. 1995, 2000), the issue on the decline in social capital is back on the research agenda. In this paper, this alleged trend is studied, while focussing on a particular relationship, i.e. among neighbors. In the last decade, a number of arguments have been provided on why neighborhood relationships might have become less important: mobility destroys local bonds, there is a shift from ascribed relationships like family and neighbors to achieved ones like colleagues and friends, and, last but not least people spend less time in their neighborhood due to demanding and unpredictable work schedules. The following questions are sought to be answered: How did the role of neighbors in personal networks change? Do neighbors (still) provide social capital for each other? How can we explain the changes: via individual differences, regional influences or countries? The data used stem from the International Social Survey Program (ISSP) on personal networks in 1986 and 2001.

3.10.2   How Much Social Capital Do You Have? Measuring social capital at the individual level in social surveys

Keming Yang; University of Reading, United Kingdom

Abstract.   Social capital has been popularly taken as a collective property and therefore analysed at aggregate levels. Empirical research following this line of thought measures social capital at the individual level in social surveys but the ultimate objective is to derive rates and proportions for a collective unit. In contrast, others have argued that it is more promising to study social capital at the individual level, but empirical research based on this principle has made little advance. In this paper I advance the idea that social capital should and can be measured at the individual level (although this should complement rather than preclude research at the aggregate level), and an important question that naturally follows is how we can measure the amount of social capital possessed by an individual, similar to how economists measure financial capital and human capital. Hence the title question. I will show that the items currently used in the World Bank Integrated Questionnaire for Social Capital, the European Social Survey, and the British Youth Lifestyle Survey can indeed be employed for evaluating an individual's amount of social capital. If the general idea can be accepted, it will be only a technical issue of how to best measure social capital as an individual property in social surveys, although this may entail many debates.

3.10.3   Measuring Social Capital in Europe

Christian Haerpfer; University of Aberdeen, United Kingdom

Florian Pichler; University of Aberdeen, United Kingdom

Claire Wallace; University of Aberdeen, United Kingdom

Abstract. One problem with measuring social capital is that there is no agreed definition of the concept as such. Another problem is that social capital is measured usually by a series of indicators (trust, civic mind-edness etc.) that are not really social capital itself, since social capital is a potential set of ties. It is par-ticularly difficult to measure this cross nationally since the amount of "noise"in the models is further increased by the fact that there are variations in cross-cultural meaning. Also, given the problem of disentangling what is a prerequisite of social capital as opposed to what is a consequence of it, means that we never really know what should be the dependent and what the independent variable. Rather, we can look only at the various associations between them.

The most common way of measuring social capital until now was the World (European) Values Sur-vey, which contains a number of questions that can be used to assess social capital. The European So-cial Survey (ESS 2002) and the Eurobarometer social capital survey in autumn 2004 recently have emerged and contribute to the lively debate on measuring social capital. Both of these sources are in many ways preferable. The ESS aims to be the best-designed survey with the largest sample (2000 in most countries). The Eurobarometer (EB) bears the advantage that it covers the whole of the European Union (25 countries) and the accession candidate countries Bulgaria and Romania. The EB for 2004 is a great improvement on the traditional EB and is in fact the best source

for the measurement of social capital today. This comparative survey has been developed by Christian Haerpfer and Claire Wallace.

The indicators that are normally used span a wide field of single elements of what is usually referred to as social capital. These indicators focus on levels of trust within a society, membership and partici-pation of voluntary organizations, qualities of good citizenship or mutual networks of support (recip-rocity).

However, the indicators are themselves problematic. Often they bear multiple meanings and do not indicate the intensity of people's involvement in community life. Take generalized trust for example. Asking whether people generally trust other people could be interpreted in many different ways. We would also expect critical citizens to exhibit a healthy scepticism of social and political institutions rather than trusting everything they come across. But such an expectation necessarily goes along with decreased levels of social capital.

Similar problems arise concerning the measurement of people's involvement in voluntary organiza-tions. Formal participation, i.e. membership, does not tell much about people's engagement. Member-ships come into existence through paying subscription fees alone. Whilst it is relatively straightfor-ward to measure formal participation (usually people know if they are members of organisations) it is extremely difficult to measure informal participation.

Trying to gauge the extensiveness and quality of social networks is extremely problematic in ques-tionnaire type research as the complex nature of personal communities can differ in number, intensity, type and so on.

Last not least, comparative research relies on national aggregates of individual measures of social capital. Even though it is a commonly used methodology, it is inherently flawed. It is especially flawed with regard to social capital because social capital is a social relationship existing between in-dividuals rather than the property of any one individual. Linked to this problem we can identify an-other. Social capital inheres in communities rather than nations or individuals.

We enrich the debate revolving around designing indicators of social capital. We refer to the state of the art and mention advantages and disadvantages of survey designs and several data sources. Finally we strive for contributing to the enlargement of survey research. We argue for the inclusion of meas-ures on the community aspect in order to derive a better understanding of social capital as a property of a group (community, society) rather than a single individual

3.10.4   Converting a CAPI module on social capital into a postal format

Gerry Nicolaas; National Centre for Social Research - NatCen, United Kingdom

Sarah Tipping; National Centre for Social Research - NatCen, United Kingdom

Abigail Dewar; Office of National Statistics, United Kingdom

Amanda Wilmot; Office of National Statistics, United Kingdom

Abstract.   In the UK, an interdepartmental working group has developed a harmonised set of social capital questions to be used in government surveys. These questions are also intended for use in local authority surveys. However, the harmonised question set has been designed for CAPI whereas most local authorities do not have the resources to carry out CAPI surveys. In response to Local Authorities' needs for a less expensive method for collecting information about social capital, the Office for National Statistics has commissioned a study to convert the harmonised set into a postal format.

It is important that the postal questions produce comparable results to the CAPI questions, so that local au-thority estimates of social capital can be compared with national estimates. In order to achieve this aim, we have adopted a three-stage approach:

(1) The Health Survey for England includes a module on social capital (not the harmonised set) which was interviewer-administered (CAPI) in 2000 and self-administered since 2001 (CASI). The comparison of results from 2000 and 2001 provided us with some indication of problems that could arise when converting the har-monised set into a self-completion format.

(2) The harmonised set was converted into a postal format based on the results from the comparison between CAPI and CASI, and a literature review on the design of postal questions. The postal questions were subsequently tested using cognitive methods.

(3) Finally, the social capital questions were included in a postal survey and a face-to-face survey, run in parallel. The data from both surveys will be compared to assess the extent to which the different modes will produce different estimates.

We will present results from stages one and two and some preliminary results from stage three.

3.10.5 Social Capital and Academic Performance among First Graders: Individual vs. Class Social Capital

Ronan Van Rossem; University of Gent, Belgium

Marjolijn Vermande; University of Utrecht, Netherlands

Chris Baerveldt; University of Utrecht, Netherlands

Beate Völker; University of Utrecht, Netherlands

Abstract. Social capital plays an important role in pupils' school adjustment. In the child development literature there is ample evidence that children with better peer relations (i.e., higher social capital) perform better at school. These authors perceive social capital primarily as an individual attribute. Another school of thought, however, argues that social capital is primarily an attribute of collectivities and benefits all the members of the collectivity. This paper examines how both individual level and class level social capital affect the academic performance and the well-being of first grade pupils. The sample consists of 1036 children in 60 first grade classes in 46 Dutch elementary schools. Multilevel regression techniques were used for the analysis. The results indicated that a substantial proportion of the variance in the school adjustment variables could be attributed to the class/school level, and that both the child level and the class level social capital have substantial effects on school adjustment.

4   Miscellaneous Topics (Aula 008)

4.1   Models for longitudinal survey data

Session chair:
  Han Oud; University of Nijmegen, Netherlands

This session takes place on:
  Tuesday, July 19; 9:00 to 11:00
  Tuesday, July 19; 11:30 to 13:30

Presentations:

- Invited presentation: *Continuous Time State Space Modeling of Panel Data by Means of SEM and its Relationship to Conventional Longitudinal SEM Modeling* - Han Oud (Netherlands)

Contributed presentations:

- *Continuous Time Models for Panel Data* - Hermann Singer (Germany)

- *Authoritarianism and Anomie by Means of Stochastic Differential Equation Modeling* - Eldad Davidov, Peter Schmidt (Germany), Han Oud (Netherlands)

- *Modeling Age-Related Development of Delinquency during Adolescence and Early Adulthood with an Auto-Regressive Growth Curve* - Johannes A. Landsheer, Cor van Dijkum (Netherlands)

- *Dynamic Analysis of Multivariate Panel Data* - Kees van Montfort (Netherlands)

- *Core Value Change and Party Allegiance: Evidence from the British Household Panel Study 1991-2001* - Patrick Sturgis, P.W. Smith, A. Berrington, Y. Hu (United Kingdom)

- *Analysing the Structure of Causality Connections between socio-economic indicators* - Yurii Gavrilets, Andrey Lebedev (Russia)

- *Systems Methodology and Social Issues* - Igor Timchenko, Ekaterina Igumnova (Ukraine)

- *The nature of measurement error in panel data: Estimating opinion stability in panel surveys* - William van der Veld (Netherlands), Willem Saris (Spain)

- *Two Different Formulations of Dynamic Linear Model for Preliminary Estimation* - Alessia Naccarato, Fabrizio Solari (Italy)

4.1.1   Continuous Time State Space Modeling of Panel Data by Means of SEM and its Relationship to Conventional Longitudinal SEM Modeling

Han Oud; University of Nijmegen, Netherlands

Abstract.   First, SEM state space modelling of panel data (Oud, 2004) will be compared to the latent growth curve (LGC) model. In contrast to the descriptive nature of LGC models, having time as the independent "explanatory"variable, the state space model has a Laplacian causal starting point. Nevertheless, LGC parameters are easily incorporated into the measurement part of the state space model as shown by several authors over the last 25 years (Goodrich & Caines in 1979; Jones in 1993; Shumway & Stoffer in 2000). The LGC random subject effects, defined by means of basis functions in the measurement part of the model, cause serious interpretation problems, especially as a result of their strong dependence on the time scale origin (zero point). In contrast, traits in a state-trait model are also random subject effects, but defined as dynamic components in the latent structural part of the state space model.

The cross-lagged panel design has become an increasingly popular tool to disentangle the over-time influences between variables. However, discrete-time models, such as the discrete-time state-trait model, have important disadvantages. First, the cross-lagged panel design offers two types of effects: cross-lagged effects and instantaneous effects, which often give contradictory results. Second, results from discrete-time analysis will be shown to be highly dependent on the length of the interval between measurements, different time-intervals giving different and contradictory results. The problems are solved by means of the continuous time state space model, using stochastic differential equations. By means of the nonlinear SEM program Mx, survey data sets from different sources, longitudinal with varying time intervals between waves as well as cross-sectional, can be utilized to estimate the same underlying state space model in continuous time. An application assessing bi-directional influences between family relationships and adolescent problem behavior will be presented.

4.1.2   Continuous Time Models for Panel Data

Hermann Singer; FernUniversität Hagen, Germany

Abstract.   Continuous time stochastic processes are used as dynamical models for discrete time measurements (time series and panel data). Thus causal effects are formulated on a fundamental infinitesimal time scale. Interaction effects over the arbitrary sampling interval can be expressed in terms of the fundamental structural parameters. It is demonstrated that the choice of the sampling interval can lead to different causal interpretations although the system is time invariant. Maximum likelihood estimates of the structural parameters are obtained by using Kalman filtering (KF) or nonlinear structural equation models (SEM).

4.1.3   Authoritarianism and Anomie by Means of Stochastic Differential Equation Modeling

Eldad Davidov; University of Giessen, Germany

Peter Schmidt; University of Giessen, Germany

Han Oud; University of Nijmegen, Netherlands

Abstract.   In this paper we are going to apply differential equations in order to delineate the dynamic process underlying the development of authoritarianism and anomie from 2002 - 2004 in Germany. The advantage of differential equations modelling over other methods to analyse data over time such as autoregressive or latent curve modelling is the higher sensitivity these model offers to track changing processes over time. Data was taken from the Group-Related Enmity Survey, which is a representative sample of the German population. The survey was conducted in three time points between 2002 and 2004. We compare in detail the results for a second order growth curve model, an autoregressive cross-lagged panel model and a differential equation model using the same data set of our three-wave panel study. Finally we evaluate the methods and their results in a comparative way.

4.1.4   Modeling Age-Related Development of Delinquency during Adolescence and Early Adulthood with an Auto-Regressive Growth Curve

Johannes A. Landsheer; University of Utrecht, Netherlands

Cor van Dijkum; University of Utrecht, Netherlands

Abstract.   To investigate the development of human beings the collection of longitudinal data over the full time scope of interest is an appropriate way to test theories and models of such time dependent behavior. Nevertheless, in many situations of research practice, data is collected covering a shorter interval than the full time scope of interest from groups that each deal with their own period of investigation.

In this study an accelerated panel study has been used with multiple age-cohorts to overcome the fragmentation of such data. The data are analyzed with an auto-regressive growth model on the basis of discrete and continuous time, using the structural equation modeling (SEM) program Mx.

Data from a research of delinquency during adolescence and early adulthood is the test case. For this research it is tried to reconstruct the time-varying process of the development of delinquency, commonly depicted in the age-crime curve, eventually differentiated according to gender.

4.1.5   Dynamic Analysis of Multivariate Panel Data

Kees van Montfort; Vrije Universiteit Amsterdam, Netherlands

Abstract.   Many models for multivariate data analysis can be seen as special cases of the linear dynamic or state space model. Contrary to the classical approach to linear dynamic systems analysis, in which high dimensional exact solutions are sought, the model presented here is developed from a social science framework where low-dimensional approximate solutions are preferred. Borrowing concepts from the theory on mixture distributions, the linear dynamic model can be viewed as a multi-layered regression model, in which the output variables are imprecise manifestations of an unobserved continuous process. An additional layer of mixing makes it possible to incorporate non-normal as well as ordinal variables.

4.1.6   Core Value Change and Party Allegiance: Evidence from the British Household Panel Study 1991-2001

Patrick Sturgis; University of Surrey, United Kingdom

P.W. Smith; University of Southampton, United Kingdom

A. Berrington; University of Southampton, United Kingdom

Y. Hu; University of Southampton, United Kingdom

Abstract.   Core values and beliefs have been invoked as fundamental explanatory variables in models of vote choice (Rokeach 1979; Feldman 1988). Although scholars have diverged in their views regarding the exact number and definitional characteristics of these social mores, treatments of British value systems have focused predominantly on the two key dimensions of 'socialist/laissez-faire' and 'libertarian/authoritarian' (Heath, Evans, and Martin 1993; Heath, Evans, and Lalljee 1996). Developing early in the life-course through long-term process of socialisation, core values have generally been conceived as highly stable guiding principles which allow citizens to make evaluative judgments about the relative merits of candidate and policy alternatives. In its simplest form, this perspective posits that parties compete to converge on the median of fixed voter preferences. More recently, however, research has begun to focus on the extent to which parties are able to shape voter positions on these dimensions to their subsequent electoral advantage (Dunleavy and Ward 1981; Stauberger 2003). Right-wing incumbents might, for instance, increase home or share ownership in an effort to shift voter preferences towards stronger support for economic liberalism, which would play to their advantage in subsequent elections. To date, however, there is little evidence to support the existence, let alone the effectiveness, of such strategies.

In this paper we apply multiple indicator Latent Growth Curve models (Meredith and Tisak 1990; Muthén 2000) to the British Household Panel Study to explore value change on the 'socialist/laissez-faire' core value dimension between 1991 and 2001. In addition to fitting models for the population as a whole, we also examine core value change as a function of party support in 1991. Lastly we turn to an evaluation of the effect of value change at the individual level on subsequent vote choice. Our results support the hypothesis that the 1990s was a period of ideological convergence, with supporters of the Labour and Conservative parties moving toward the centre from the left and right wing poles, respectively. There is little evidence to suggest, however, that value change at the individual level had much subsequent influence on vote choice in the 2001 general election.

4.1.7 Analysing the Structure of Causality Connections between socio-economic indicators

Yurii Gavrilets; Russian Academy of Sciences, Russia

Andrey Lebedev; Russian Academy of Sciences, Russia

Abstract. This work provides the forecast of multidimensional Markov chain for components of a vector that represents socio-economic indicators of population in Russian Federation. The vector is built on the CESSI data and varies in time. The proposed method of forecasting is based on the two notions: the one of Markov's three characters and the one of multidimensional random value structure. Also information and entropy approach was used as the basic method of measuring the Causality Connections. Static and dynamic causal graphs were built.

The research allows to consider the stability of the causal graph in the sense of conservation of revealed connections, the possibility of forecasting itself, the elimination of the dimension of the space of parameters due to throwing out of the consideration non-connected components of parameters vector.

4.1.8 Systems Methodology and Social Issues

Igor Timchenko; National Academy of Sciences Ukraine, Ukraine

Ekaterina Igumnova; National Academy of Sciences Ukraine, Ukraine

Abstract. A new systems methodology presented in this paper was developed was aimed on analysis of socio-economic processes of sustainable development.[1]. This methodology consists of systems principles, thinking, modeling and management technology. We employ six systems principles (goals relativity, integrity, causality, subordination, adaptive balance and information unity [1]) to construct a new Adaptive Balance of Causes (or ABC) method for complex systems modeling and to create a new systems thinking technology of decisions support.

The ABC method has several advantages in contrast with the famous system dynamics method (Forrester and his followers) due to standard ABC module equation for all modeling processes, the possibility of objective determination of the model's coefficients and current observational data assimilation in the model. To achieve these features we introduced the dynamic-stochastic approach based on the optimal filtering theory [2].

The core of the proposed decision support technology is a standard multi-agent "cost-benefit"ABC model. It permits development scenarios predictions under various changing external conditions. In accordance with the systems thinking concept it focuses on the goal-seeking movement in a specified environmental system.

We applied the ABC AGENT technology in many case studies of economic, ecological-economic and socio-economic systems. Some effective results were obtained in an integrated management of natural-industrial systems which conjoin socio-economic and environmental managerial objectives [3]. The most recent implementations of this technology were made in cognitive and educational systems [4]. We proposed an ABC negotiation support model helping to take advantage over opponents by the systemic study of their behaviour. Various ABC AGENT control technologies were suggested for learning and educational systems. In particular, conceptual model of students' communication skills development was built.

A number of model parameters, representing the required skills, correspond to the significant personal qualities that determine the communicative or lingua-cognitive personality types of students. Being the basic elements of general communication "culture"of a person, these notions characterize personality of each student individually, revealing the levels of the student's speech ability/skills and communication ability/skills. Mutual influences between all the model parameters, describing the communication "culture"development system are studied, with application to a student individually, or on the level of a whole academic group.

The learning cycle management model allowed to evaluate the lecturers' roles in the cycle, as well as to study various optional scenarios of processes within the cycle. The necessary decisions to improve the quality of teaching/learning were made on the basis of numerous simulation experiments.

REFERENCES

1. Timchenko I.E., Igumnova E.M. and Timchenko I.I. Systems Management and ABC technologies of Sustainable Development. Sevastopol. "ECOSEA-hydrophysics". 2000. 220 p. (in Russian)

2. Timchenko I.E. Stochastic Modeling of Ocean Dynamics. Harwood Academic Publishers, Chur-London-Paris-New-York, 1984. 240 p.

3. Eremeev V.N., Igumnova E.M. and Timchenko I.E. Modeling of Ecological-Economic Systems. Sevastopol. "ECOSEA-hydrophysics". 2004. 320 p. (in Russian)

4. Timchenko I.I., Igumnova E.M. and Timchenko I.E. Education and Sustainable Development. Systems Methodology. Sevastopol. "ECOSEA-hydrophysics". 2004. 527 p. (in Russian)

4.1.9   The nature of measurement error in panel data: Estimating opinion stability in panel surveys

William van der Veld; University of Amsterdam, Netherlands

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract. This study is about the assessment of reliability of single survey measures. Reliability of single survey measures can only be assessed with several designs. We will discuss two of the most common designs: the quasi-simplex design (panel) and the parallel test design (cross section). The quasi-simplex design should have been the solution for the assumed deficiencies of the parallel test design. However, the quasi-simplex design seems to have problems of its own, resulting in the underestimation of the reliability. This follows from correction for attenuation leading to corrected correlations larger than 1, which is theoretically impossible. In addition, the assumed effect of memory (the biggest deficiency) in the parallel test design is not so bad that implausible correlations appear after correction. In order to show what exactly is wrong with the reliability estimate from the quasi-simplex model –to show the true nature of measurement error in this panel design –we have compared the model with the VAS model. We have recently (Van der Veld and Saris, 2004) developed this model, and it follows from the VAS model that the reliability in the quasi-simplex model is the product of attitude crystallization and instrument quality.

4.1.10   Two Different Formulations of Dynamic Linear Model for Preliminary Estimation

Alessia Naccarato; University of Rome 3, Italy

Fabrizio Solari; ISTAT, Italy

Abstract. As prescribed by European Community timeliness of statistical information, i.e. the length of time occurring between its availability and the reference period of the event or phenomenon it describes, is a key issue for improving survey process quality.

The usual practice is to report preliminary estimates of the variable of interest for the current reference period and to produce final estimates when all the information is available. In literature different approaches have been proposed, typically based on either a special sampling design or a different way to collect data or simply using early respondents and modelling for preliminary non responses treatment.

In this paper the problem of constructing preliminary estimates is addressed adapting dynamic linear models to the context of short-term surveys. We refer to two different contexts: the former related to panel survey data, for which information for each sampling units is available for all the surveys carried out in the previous periods of time; the latter more general context concerning situations where sampling units (preliminary respondents or not) are not necessarily the same over the time. In the first case the natural choice is to use a unit level model with all the information derived from the previous periods of time used for non-responses treatment. Unfortunately, in the second case the unit level model is not feasible since auxiliary information available for units not always in the surveys is not coherent with the model. One possible solution is to use a model obtained aggregating sampling units, on the basis of an auxiliary variable correlated with the phenomenon of interest.

The proposed methodology has been applied to data from the monthly survey of industrial turnover conducted by the Italian National Statistical Institute. The turnover index is one of the most relevant short-term indicator of business activity, in terms of the demand for industrial output. The aggregated model has been defined on the basis of the classification of economic activities. The above specified models have been compared by means of a simulation study. More precisely, the properties of both models have been assessed in the panel survey framework, while the aggregated model has been also applied in the more general situation with sampling units changing over the time.

4.2 Social Attitudes Time Series

Session chair:
    Roger Jowell; City University, United Kingdom

This session takes place on:
    Tuesday, July 19; 15:00 to 17:00

Presentations:

- Invited presentation: *Unresolved (and unresolvable?) issues in the long-term measurement of social attitudes* - Roger Jowell (United Kingdom)

Contributed presentations:

- *The German General Social Survey (ALLBUS)* - Michael Blohm, Achim Koch, Martina Wasmer (Germany)

- *Social-Science Research and the General-Social-Survey Model: The American Experience* - Tom W. Smith (United States)

- *Poles Apart? Some Methodological Reflections on the South African Social Attitudes Survey* - Ben Roberts, Udesh Pillay (South Africa)

- *The advantages and disadvantages of panel studies for studying the change in objective and subjective indicators of living conditions –example of russian socio-economic panel* - Anna Andreenkova (Russia)

4.2.1 Unresolved (and unresolvable?) issues in the long-term measurement of social attitudes

Roger Jowell; City University, United Kingdom

Abstract.   The British Social Attitudes surveys started in 1983 and are still running annually some 22 years later. As with several similar national time series (eg in the USA, Germany, Australia, South Africa, Brazil and Mauritius), and in common with analogous cross-national time series (such as the European Social Survey, the ISSP, the Eurobarometers and the European Values Surveys), their measurements are based on repeat cross-sections rather than panels. Their primary purpose is to measure changes in societal values over time.

Given this common aim, it might seem reasonable to assume that, after settling on their sampling methods, questionnaires and so on at the start, such time series then require few decisions on measurement issues between successive rounds. Surely these aspects of methodology are set in stone to ensure that any trends they then uncover reflect changes in outputs rather than inputs? At any rate it would be comforting to believe that this was so.

In reality, however, these time series are faced with regular convincing arguments at more or less every round to adjust their methods or instruments for the sake of some larger and more important considerations. One such consideration might be to help consolidate or expand funding opportunities; another might be to accommodate new developments in survey methods. Other changes might be prompted by shifts on the ground in social or political circumstances, or by the emergence of new theories or ideas that need testing. Some changes might even be occasioned by new knowledge arising from the survey's own findings.

How flexible can such times series afford to be in response to these sorts of persuasive pressures to change their inputs? Ought they to have strategies in place for aiding their decisions at the margin? Experience from various national and cross-national time series will be presented to illustrate the sorts of dilemmas involved, and examples will be adduced to show when and in what circumstances even damaging interruptions to these time series have been considered legitimate.

4.2.2 The German General Social Survey (ALLBUS)

Michael Blohm; Centre for Survey Research and Methodology (ZUMA), Germany

Achim Koch; Centre for Survey Research and Methodology (ZUMA), Germany

Martina Wasmer; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   The German General Social Survey (ALLBUS) is a biennial survey that has been conducted since 1980. In each ALLBUS survey, a representative cross-section of the population is interviewed. The ALLBUS questionnaire usually covers a wide range of topics relevant for Sociology and Political Science. Substantial parts of the questionnaire are replicated at intervals in order to facilitate the analysis of stability and change in attitudes, behaviour and social structure of the German population. ALLBUS is oriented toward academic use, the data are accessible for all persons and institutions interested as soon as data preparation and documentation have been finalized. The data are widely used in research and teaching.

In the present paper we will firstly provide information on several basic features of ALLBUS, like content of the questionnaire, sampling procedures and sample sizes, organisational structure, and data usage. Secondly, we will focus on the issue of data collection. From its very beginning, ALLBUS has been fielded as a face-to-face survey. For each round of ALLBUS a commercial survey organisation has been selected to carry out the fieldwork. In recent years, data collection has posed major challenges for ALLBUS, both with respect to fieldwork results as with respect to costs and time needed. In our presentation we will discuss the relationship between response rates, sample composition and sample bias, length of fieldwork period and survey costs.

4.2.3 Social-Science Research and the General-Social-Survey Model: The American Experience

Tom W. Smith; University of Chicago, United States

Abstract.   During the last generation a new measurement instrument has emerged in the social sciences, the general social surveys (GSSs). Traditionally, empirical research in the social sciences had been intermittent, varied greatly in data quality and generalizability, focused on a narrow set of issues and/or hypotheses, and led by a senior researcher or principal investigator pursuing his or her own research agenda. The GSSs embraced a new model of social-science research.

This paper discusses 1) the GSS-model of social-science research including a) the creating of a social-science infrastructure, b) reliable, valid, and generalizable measurement, c) broad coverage of topics, d) a collective, community focus, and e) equal and widespread access and use; 2) the initial development of the GSS-model in the United States, Germany, Great Britain, and Australia, and 3) recent developments, especially in East Asia.

4.2.4   Poles Apart? Some Methodological Reflections on the South African Social Attitudes Survey

Ben Roberts; Human Sciences Research Council (HSRC), South Africa

Udesh Pillay; Human Sciences Research Council (HSRC), South Africa

Abstract. Compared to the long-established American General Social Survey (since 1972), the German General Social Survey (ALLBUS, since 1980) and the British Social Attitudes Survey (since 1984), the South African Social Attitudes Survey (SASAS) is in its relative infancy. Conducted annually by the Human Sciences Research Council since 2003, the survey gathers information on the attitudes, beliefs, behaviour patterns and values of a representative sample of 7000 individuals aged 16 and older. The intention is to measure specific attitudes and opinions on a diverse set of topics of national relevance, including democracy and governance, poverty, social identity, and moral issues. It will thus provide a unique long-term account of the social fabric of modern South Africa, and of how its changing political and institutional structures interact over time with changing social attitudes and values. This paper examines key aspects of the survey's design, notably data collection mode, continuity versus change in content, and translation.

In common with most national surveys in South Africa, SASAS uses face-to-face data collection. The preference for this mode of interviewing is influenced by a number of factors. These include the relatively low level of literacy of the South African population and the positive influence that the presence of an interviewer has on response rates. Telephonic and mailed surveys are also constrained respectively by insufficient household access to telephones and the absence of postal addresses for a sizable share of the population. Nonetheless, although response rates in the two SASAS rounds have been good, a lingering problem is the ability to reach respondents in the upper end of the income distribution. Escalating violent crime has led to the rising use of access control to restrict access to properties and suburbs over the last decade. The increasing presence of high walls, ferocious dogs, security guards and gated communities poses a real challenge for gaining physical access to selected visiting points. This is compounded by the difficulty of finding people at home and a higher propensity for refusal due to questionnaire length and the sensitivity of certain topics. Recent SASAS planning meetings have begun to discuss experimenting with alternate or mixed mode approaches of data collection for these difficult to reach localities in future rounds.

The process of determining item or topic continuity or change between rounds is an area where there is scope for improvement. The thematic content of the baseline and subsequent round is currently informed by a list of key perennial topics identified through a series of workshops with academics. However, there are mounting concerns about questionnaire length and, associated with this, whether an appropriate balance between core and rotating items has yet been achieved. In a multilingual society such as South Africa, which has eleven official languages and is commonly referred to as the 'rainbow nation', it is important that respondents are interviewed and answer questions in a language they feel comfortable. This implies that questionnaires require translation into each of the major languages. Securing linguistic equivalence in a multilingual environment is extremely challenging. The diversity of cultures means that attempts at finding shared meanings for concepts such as 'family', 'household', 'married', 'employed' are fraught with difficulty and that there exists a high risk for misrepresentation due to misunderstanding. Consequently, much time and resources need to be devoted to translation procedures.

The paper concludes by arguing that, in spite of the demands that prevailing social conditions in developing countries like South Africa pose for the design of social surveys, there remain similarities with the approach adopted in developed countries and cross-national surveys. There are also parallels in the problems faced, though it is recognised that the frequency and severity of these vary considerably. This presents an important opportunity for enhanced knowledge-sharing and lesson-learning in the continuing process of improving the quality and accuracy with which social attitudes are measured in South Africa.

4.2.5   The advantages and disadvantages of panel studies for studying the change in objective and subjective indicators of living conditions –example of russian socio-economic panel

Anna Andreenkova; Institute for Comparative Social Research, Moscow, Russia

Abstract. The paper will discuss the main advantages and some disadvantages of using the panel studies instead of cross-sections to study the change in objective and subjective indicators of living conditions in the time of rapid socio-economic and political transformations on the example of RUSSET (Russian socio-economic panel) - panel study of Russian households conducted in 7 annual waves from 1993 to 1999 (subsidized by Dutch Organization for scientific research NWO for a period of 7 years).

Among the advantages of the panel the paper will show that analyzing cross-section data we see no change or only minor changes in most objective and subjective indicators of living conditions even in the situation of basic and rapid changes in economic conditions (in particular in housing conditions, working conditions, distribution of work in the household, family conditions, etc.), but the changes on individual level which was tracked in the panel study were in opposite very large.

Among the problems specific for panel studies, RUSSET faced the problem of sample attrition (from 3727 respondents in 1 wave to 2808 in second, then the attrition decreased to 5% a year and increase to 11% in the last wave). This problem was accompanied by the change in sample structure because of two factors –ageing of the sample and systematic attrition of the most mobile part of population which experienced the most changes (housing, job, marital status). The solution of this problem used in RUSSET was the reconstruction of the sample to the initial size adding the new independently drawn national sample in the middle of the survey (wave 5). This allowed to check how large was the biases related to sample attrition of initial panel comparing some indicators with new independence sample.

Except of the sampling problem, RUSSET also faced the problem of question wording for measuring the individual change in objective living conditions. Two approaches were used in RUSSET: exactly the same questions in different survey waves (repeated questions) and direct question about the change in particular parameter (Did you get any additional education for the last 12 month? etc.). The paper will discuss the advantages and disadvantages of both of these approaches.

4.3   Comparative longitudinal surveys

Session chair:
> Peter Lynn; University of Essex, United Kingdom

This session takes place on:
> Tuesday, July 19; 17:30 to 19:30

Presentations:

- Invited presentation: *Comparative Longitudinal Surveys: What are the distinctive methodological issues? -* Peter Lynn (United Kingdom)

Contributed presentations:

- *Adapting an existing UK survey to meet the longitudinal requirements of EU-SILC: Challenges and Opportunities -* Roeland Beerten (United Kingdom)

- *Implementing EU-SILC in the Czech Republic -* Simona Měřinská, Martin Zeleny (Czech Republic)

- *The Cross-National Equivalent File (CNEF) of Household Panel Surveys: A product of cross-national research -* Gert G. Wagner (Germany)

- *Ensuring cross-national and cross-cultural comparability in SHARE –management, processes, and survey design -* Axel Boersch-Supan, Karsten Hank, Hendrik Juerges (Germany)

4.3.1 Comparative Longitudinal Surveys: What are the distinctive methodological issues?

Peter Lynn; University of Essex, United Kingdom

Abstract. Longitudinal surveys are methodologically distinct from cross-sectional surveys in many ways. And comparative (cross-national) surveys are rather distinct from single-nation surveys. In both cases, the distinguishing features can broadly be summarized as considerable extra analysis potential, accompanied by considerable extra design complication in order to realize that potential. Comparative longitudinal surveys might therefore be thought of as the "holy grail"of the survey world. They have tremendous potential to further our understanding in ways that other surveys cannot. But to achieve this, there are very considerable challenges of design and implementation that must be faced.

This presentation will set out some objectives for the design and implementation of comparative longitudinal surveys. A set of criteria for achieving those objectives will then be outlined. These criteria will in turn lead to the identification of the features of design and implementation that require particular attention. It will be seen that some of these features are either unique to comparative longitudinal surveys or play a distinctive role in the case of comparative longitudinal surveys.

The ideas will be illustrated with examples from surveys that are either cross-national or longitudinal (or both).

4.3.2 Adapting an existing UK survey to meet the longitudinal requirements of EU-SILC: Challenges and Opportunities

Roeland Beerten; Office of National Statistics, United Kingdom

Abstract. In April 2005 the ONS General Household Survey started a new life as a longitudinal survey with a modified content. Both the longitudinal sample design and the new survey content were introduced to comply with the European Regulation for Statistics on Income and Living Conditions (EU-SILC). Under this regulation all EU Member States are required to provide Eurostat with longitudinal data on income and living conditions of households.

The UK decided to integrate the new data requirements with an existing survey, partially for efficiency reasons. After an assessment of the scope and content of existing household surveys, the General Household Survey was considered to be the survey which best matched the EU-SILC requirements. Subsequently the survey was redesigned to accommodate the EU-SILC components.

The new GHS(L) differs from the current GHS in various aspects, and this paper will present the solutions ONS developed to meet the challenges of the new design. The presentation will focus on the new sample design, the changes to the questionnaire (including the methodology for longitudinal data collection) and the design of a system for tracing respondents throughout the panel. It will discuss the benefits of the new design and its opportunities for future survey work in ONS.

4.3.3 Implementing EU-SILC in the Czech Republic

Simona Měřinská; Czech Statistical Office, Czech Republic

Martin Zeleny; Czech Statistical Office, Czech Republic

Abstract. This paper describes some of the methodological issues that had to be faced in implementing the EU-SILC survey in the Czech Republic. The requirements of the survey posed several challenges, especially the requirements for longitudinal data collection. We describe which aspects of the requirements were particularly demanding, and why. We summarise relevant outcomes of two stages of field piloting. And we highlight some implications for the cross-national comparability of data.

In the realm of social surveys, the Czech Statistical Office (CZSO) had no prior experience with panel surveys of this kind. Therefore, the sample design and following rules for subsequent waves required considerable attention. Additionally, there was no similar previous survey which could be used to help predict longitudinal response rates, and hence the required initial sample size. In the paper, we describe our experiences with developing a sample design, testing the response rate (both for wave 1 and wave 2 of the Pilot) and operationalising the following rules. Some constraints on the fieldwork operation will also be outlined, as these are pertinent to field outcomes such as response rate and success at following movers.

Several of the concepts and definitions underpinning the data to be collected were found to be divergent from standard definitions used within the CZSO and, in some cases, difficult to apply to the Czech situation. We will describe some examples, the questionnaire solutions that we adopted and lessons we learnt from the two pilots. In some cases, we believe there may be implications for cross-national comparability of data.

4.3.4 The Cross-National Equivalent File (CNEF) of Household Panel Surveys: A product of cross-national research

Gert G. Wagner; DIW Berlin and Berlin University of Technology (TUB), Germany

Abstract. I describe the project CNEF that builds on the LIS model (Luxembourg Income Study) which contains cross-sectional data sets. But CNEF uses multiple waves of longitudinal data from Canada, Germany, Great Britain, and the United States (and Australia). What distinguishes this project from other standardization projects is that the development of the data is driven by research questions. All equivalently defined variables flow from the research of experienced cross-national researchers who have developed cross-nationally comparable measures for their own analyses. Thus, the standardized data CNEF maintains is an amalgam of the knowledge of many researchers answering a diverse set of questions.

Researchers at Cornell University along with colleagues from the Institute for Social and Economic Research (ISER) at the University of Essex, the German Institute for Economic Research (Deutsches Institut für Wirtschaftsforschung –DIW Berlin) in Berlin, Statistics Canada in Ottawa, and the Survey Research Center at the University of Michigan have developed and tested algorithms that place information from four panel surveys into a framework of comparably defined variables for use in cross-national research. The panel surveys include the British Household Panel Survey (BHPS), the German Socio-Economic Panel (GSOEP), the Canadian Survey of Labour and Income Dynamics (SLID) and the United States Panel Study of Income Dynamics (PSID). Using these panel surveys, researchers created a longitudinal micro-database known as the Cross-National Equivalent File 1980-1998 (CNEF). The CNEF is administered at Cornell University. It was funded from 1990 through 2000 by the National Institute on Aging and is now funded by DIW Berlin and Cornell University.

In the presentation I first provide an overview of the equivalent data file and the four data sets on which our standardized data are based. I then describe the methodological approach to the standardization of variables and I provide a list of the comparable variables we have created. Finally, I provide an example of how these variables can be used in cross-national comparative analyses by showing how economic well being after labor force exit varies across countries.

4.3.5 Ensuring cross-national and cross-cultural comparability in SHARE –management, processes, and survey design

Axel Boersch-Supan; University of Mannheim, Germany

Karsten Hank; University of Mannheim, Germany

Hendrik Juerges; University of Mannheim, Germany

Abstract. The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national data base of micro data on health, socio-economic status and social and family networks of some 22,000 Continental European individuals over the age of 50. Eleven countries have contributed micro data to the 2004 SHARE baseline study. They are a balanced representation of the various regions in Europe, ranging from Scandinavia (Denmark and Sweden) through Central Europe (Austria, France, Germany, Switzerland, Belgium, and the Netherlands) to the Mediterranean (Spain, Italy and Greece). Further data are currently being collected in Israel.

SHARE has been designed after the role models of the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). Compared to HRS and ELSA, SHARE has the advantage to encompass cross-national variation of public policies, cultures and histories in a variety of European countries. This advantage makes SHARE a unique and innovative data set.

We describe the strategy –in management, processes, and survey design –taken by SHARE to ensure strict cross-national and cross-cultural comparability of the data: (1) a 'matrix structure' of working groups on survey content and country teams, (2) centralised (final) decision-making enforced by central funding, (3) a single generic CAPI survey instrument, (4) tightly controlled translation following the ESS model, (5) objective health measurements and vignettes to gauge subjective, culturally-biased, survey responses.

4.4   Reduction of measurement errors in surveys / effects of dependent interviewing

Session chair:
    Annette Jäckle; University of Essex, United Kingdom

This session takes place on:
    Wednesday, July 20; 9:00 to 11:00

Presentations:

- Invited presentation: *Dependent Interviewing: A Framework and Review of Current Practices* - Annette Jäckle (United Kingdom)

Contributed presentations:

- *Dependent Interviewing and Data Quality in the Survey of Income and Program Participation* - Joanne Pascale (United Kingdom)

- *Using Dependent Interviewing to Measure Changes in Household Composition on the Millennium Cohort Study* - Lisa Calderwood (United Kingdom)

- *The Effect of Dependent Interviewing on Measures of School Progression* - Alison Aughinbaugh (United States)

- *Does Dependent Interviewing Make You Sick?* - Carli Lessof, Hayley Cheshire (United Kingdom)

4.4.1   Dependent Interviewing: A Framework and Review of Current Practices

Annette Jäckle; University of Essex, United Kingdom

Abstract. Many panel surveys have introduced dependent interviewing (DI), the feeding forward of substantive information from previous waves of data collection to formulate questions or determine routing (proactive DI) or prompt post-response edit checks (reactive DI). Proponents of dependent interviewing claim that it reduces respondent burden, increases the efficiency of data collection and improves data quality. The main purpose, however, varies across survey organisations and across surveys resulting in a variety of designs and applications. As a consequence it is not straightforward to evaluate the effects of different design features or compare the use of dependent interviewing in different surveys.

This article reviews current applications and develops a conceptual framework of dependent interviewing, in an attempt to disentangle design features, channels through which they take effect, and implications for burden, efficiency and data quality. The scope of design features is broad. With proactive DI respondents can be reminded of previous answers (1) to aid their memory and provide a boundary before asking the standard independent question, (2) to ask respondents to check and confirm previously recorded answers, or (3) explicitly to ask about changes. If circumstances have not changed, proactive DI may in addition provide the means to route around follow-up questions (perhaps with subsequent imputation of previous data). Reactive DI can be used to follow-up item non-response (don't knows or refusals), or as a corrective edit which can be displayed (1) always, for example to check consistency of verbatim answers with previous reports, (2) to clarify reports that are inconsistent with previous reports, or (3) selectively, if reports differ from previous reports beyond a defined threshold. The aim of the conceptual framework is to aid the understanding of the effects of dependent interviewing, about which empirical evidence has been very limited until recently.

4.4.2   Dependent Interviewing and Data Quality in the Survey of Income and Program Participation

Joanne Pascale; National Centre for Social Research - NatCen, United Kingdom

Abstract. Since 1983 the U.S. Census Bureau has been conducting the Survey of Income and Program Participation (SIPP), a large panel study designed to measure the dynamics of economic well-being among U.S. household residents. Core topics include demographics, health insurance and income from a broad range of sources including jobs and businesses, government assistance programs and assets. The survey is administered every 4 months over the course of 3-4 years (depending on the panel) and each time the survey is administered (called a 'wave') month-level data are collected on these topics. Beginning in the late 1990s the Census Bureau launched a major effort to review and improve on the survey. Among the many data collection techniques explored was 'dependent interviewing.' This is a technique whereby data collected in one wave of the survey is fed back during the next wave in order to reduce repetitiveness and burden, aid recall, reduce spurious change, and generally provide a sense of continuity over the life of the survey.

This paper will present an overview of several different evaluations of dependent interviewing (DI) in the SIPP. Cognitive interviewing was conducted to explore respondents' general reaction to the technique and their understanding of confidentiality issues related to feeding back data collected from one respondent in one wave of the survey to a different respondent in a later wave. Interviewer debriefings and surveys were carried out to assess strengths and weaknesses of DI relative to control conditions from the interviewers' perspective. On the quantitative side, three large split-ballot field surveys were conducted in which the control treatment employed only minimal DI and the test treatment reflected a much more comprehensive use of DI. Evaluation of these field tests included unit nonresponse rates and attrition, item non-response rates for income questions, and 'seam bias' analysis. The 'seam' in a panel survey is the point at which one wave meets another. Seam bias is observed when there is an unevenness in the pattern of month-to-month changes observed over the life of the panel. For example, across two waves of SIPP data, there are 8 months and hence 7 month-to-month transition pairs. In theory, the number of changes (for example, on-and-off welfare, or on-and-off employment) should be roughly equal across each of the 7 month-to-month pairs. However, in the SIPP and many other panel surveys, many more changes are observed at the seam than at other month-to-month pairs in the timeline. DI has the potential to reduce this seam bias by reducing spurious change and hence producing a pattern of change more evenly distributed over time.

4.4.3   Using Dependent Interviewing to Measure Changes in Household Composition on the Millennium Cohort Study

Lisa Calderwood; University of London, United Kingdom

Abstract. Dependent interviewing is becoming increasingly common in longitudinal survey research and is used as a way of attempting to reduce both measurement error and respondent burden. This method of data collection involves using data from a previous interview to inform the choice and/or wording of questions.

The UK Millennium Cohort Study follows nearly 19,000 children born in 2001/2 and is funded by the UK Economic and Social Research Council (ESRC) and a consortium of UK government departments led by the Office for National Statistics (ONS). Unlike the previous British birth cohort studies (National Survey of Health and Development, National Child Development Study and 1970 British Cohort Study), the initial sweep of data collection took place when the cohort children where aged 9 months rather than during the first week of life. The second sweep of data collection took place when the cohort members were aged 3 years and further surveys are currently planned for Age 5 and Age 7.

This paper will focus on the use of dependent interviewing to measure change in household composition as many studies use dependent interviewing in this area. It will look at changes in household composition and relationships between household members between the first and second surveys of the Millennium Cohort Study and reflect on how these findings may inform the choice of interviewing method in this area.

4.4.4   The Effect of Dependent Interviewing on Measures of School Progression

Alison Aughinbaugh; US Bureau of Labor Statistics, United States

Abstract.   This paper sets out to explore whether and how reports of progression through school, as measured by highest grade attended and highest grade completed, vary under independent versus dependent interviewing techniques. To examine this question, I will use two longitudinal data sets across which the schooling questions have been asked differently. The first is the National Longitudinal Survey of Youth 1979 (NLSY79), which began as a nationally representative sample of youth ages 14 to 21 (birth years 1957 to 1964) and living in the United States in 1979. The respondents of the NLSY79 were interviewed annually through 1994 and have been interviewed biennially since. Along with questions on employment, marriage, income, and fertility, at every interview the NLSY79 respondents have been asked about their progression through school. Each respondent is asked whether s/he attended school since the date of the last interview, and those who have attended school are asked to report highest grade attended and highest grade completed.

The National Longitudinal Survey of Youth 1997 (NLSY97), the second data set employed in this study, is composed of youth ages 12 to 16 (birth years 1980 to 1984) who were living in the US in 1997. These youth continue to be interviewed annually with the first seven rounds of data currently available. Because of apparent inconsistencies in the NLSY79 schooling data, dependent interviewing is used in the schooling section of the NLSY97 and an event history of schooling is collected. After enrollment dates and grades attended and completed at each school are collected, the reports of highest grade attended and highest grade completed provided during the current survey round are compared with the reports from the previous interview. If the information seems inconsistent across the two rounds, the respondent is asked to explain the pattern of schooling. For instance, if a respondent has advanced two grades over the course of a single year, he is asked why the highest grade he currently reports attending is more than one grade higher than the highest grade he reported attending at the previous interview. This then gives the respondent a chance to explain, or to realize that a mistake was made in his reporting or in the interviewer's entering of his report which can then be corrected.

The difference in the way the questions are asked in the NLSY79 and NLSY97 provides an opportunity to examine whether dependent interviewing impacts the frequency of apparent skipping and repeating of grades. For this analysis, the sample will be limited to youth of the same ages. That is, the progression through school reported by the 1980 to 1982 birth cohorts in the NLSY97 will be compared to that of the 1962 to 1964 birth cohorts from the NLSY79. In both surveys, each birth cohort contains over 1500 respondents, providing large sample sizes which should permit even small differences to be detected. Admittedly, a number of factors that influence retention and advancement through school may have changed between the time in which the two groups attended school; however, the broad array of topics covered in the surveys should help control for at least some of the potential differences.

4.4.5   Does Dependent Interviewing Make You Sick?

Carli Lessof; National Centre for Social Research - NatCen, United Kingdom

Hayley Cheshire; National Centre for Social Research - NatCen, United Kingdom

Abstract. The English Longitudinal Study of Ageing (ELSA) is a major longitudinal study of people aged 50 and over and their younger partners funded by the National Institute on Ageing (US) and a consortium of government departments in the UK. Over time, it will explore the way in which economic, social and health circumstances affect each other as individuals age.

ELSA has been developed through a collaboration between University College London, the Institute of Fiscal Studies, and the National Centre for Social Research (NatCen). The sample was drawn from households that had previously responded to the Health Survey for England (HSE) in 1998, 1999 or 2001 and were living in private households in England at the time of their first ELSA visit. As a result, pre-baseline information from the HSE is available about most sample members, with particularly detailed health data. The first 'follow-up' wave of ELSA was carried out in 2002/03 and resulted in 12,100 interviews. The second wave began in 2004 and is nearing completion. We aim to re-interview sample members every two years.

ELSA is modelled closely on the Health and Retirement Study (HRS), based in the United States, and both studies have adopted the use of dependent interviewing in some sections of the interview (ELSA's approach will be outlined in the seminar). For the most part, the use of dependent interviewing is uncontroversial, for example checking information about household composition and other rosters. However, during the health module of HRS and ELSA participants are reminded about health conditions reported at the previous wave, and asked to confirm whether they still have the condition. This form of proactive dependent interviewing is less commonly cited in the literature and the implications of using this method on reported prevalence rates has not been fully researched.

A recent paper entitled "Do panel surveys make people sick? US arthritis trends in the Health and Retirement study"Social Science & Medicine (2004), compares age-specific arthritis prevalence in the HRS with cross-sectional findings from the National Interview Survey (NHIS) in the United States. HRS was found to report a higher prevalence of arthritis than NHIS. The authors suggested that panel conditioning was a possible explanation for this. It was argued that participation in periodic health surveys may induce some participants to seek diagnostic testing or ask doctors about symptoms they would not have otherwise mentioned. A number of other explanations are, however, possible. In particular, there is a need to assess whether dependent interviewing could lead to over-reporting of specific conditions (through acquiescence). Possible analysis presented at the seminar could be used to:

1. Make an initial assessment of whether questioning techniques in longitudinal studies appear to influence reporting of specific health conditions for those aged 50 and over by comparing ELSA (which adopts this form of dependent interviewing) with the British Household Panel Study (which does not).

2. Make an initial assessment of whether panel conditioning influences reporting of a variety of health conditions for those aged 50 and over, by comparing each wave of ELSA and BHPS against the equivalent year of the Health Survey for England (HSE). Principally, are respondents in panel suveys more likely to report health conditions (regardless of method of questioning) compared to those in cross-sectional studies, and what explanations, including differential attrition, could be offered? How much of a part might dependent interviewing play?

4.5   Cognitive Models Of Survey Research

Session chair:
    Jaap Murre; University of Amsterdam, Netherlands
    Bregje Holleman; University of Utrecht, Netherlands

This session takes place on:
    Wednesday, July 20; 11:30 to 13:30
    Wednesday, July 20; 15:00 to 17:00

Presentations:

- Invited presentation: *Cognitive Models of Survey Research* - Bregje Holleman, Jaap Murre (Netherlands)

Contributed presentations:

- *The Memory Chain Model for question answering behaviour in surveys: Theory and application* - Antonio G. Chessa, Bregje Holleman, Stasja Draisma (Netherlands)

- *The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties* - Allyson L. Holbrook (United States)

- *Origins and Developments of the Cognitive Models of Answering Questions in Survey Research* - Mario Callegaro (United States)

- *Cognitive factors of stability of answers to attitudinal questions* - MichałWenzel, Krzysztof Zagórski (Poland)

- *A Hybrid, Affective-Cognitive Computational Model of the Survey Response Simulating the Dynamics of Candidate Evaluation in the 2000 U.S. Presidential Election* - Milton Lodge, Sung-youn Kim (United States)

- *Affect and False Beliefs About Candidates* - Dan Cassino (United States)

- *Analyzing cognitive processes in CATI surveys with response latencies –an empirical evaluation of the consequences using different baseline speed measures* - Jochen Mayerl, Piet Sellke (Germany)

- *An Overview of the Respondent-Generated Intervals (RGI) Approach to Sample Surveys* - James Press, Judith Tanur (United States)

- *Enhancing recall accuracy in surveys by cognitively tailored timeline methods: a record check study* - Wander van der Vaart, Tina Glasner (Netherlands)

- *The reminiscence bump and the telescoping effect* - Steve M.J. Janssen, Antonio G. Chessa, Jaap Murre (Netherlands)

4.5.1   Cognitive Models of Survey Research

Bregje Holleman; University of Utrecht, Netherlands

Jaap Murre; University of Amsterdam, Netherlands

Abstract.   The book by Tourangeau, Rips and Rasinski (2000) introduced a verbally stated model of survey research, consisting of four stages: (1) comprehension of the question, (2) retrieval of information, (3) deriving a judgment, and (4) formulating a response. This model takes an important step towards a deeper understanding of the cognitive and other process involved in answering survey questions. In this talk we will present new developments since that model, including some of the work by the project group that under the direction of Willem Saris spent one year on the development of new models and experiments. We will discuss a number of research methods from cognitive psychology and even cognitive neuroscience that can be of great benefit to survey researchers. In particular, when these methods are combined with existing theories and models in cognitive science, survey research can be put in a broader perspective and at the same time use these advances to remove sources of error. We will review a number of useful methods, such as reaction time studies and evoked response potentials, and a number of modelling formalisms, such as connectionist models and rule-based models. We will also address the following specific topics and we will invite the audience to discuss these with us, such as explicit versus implicit processes in question answering, and the neural basis of attitudes.

4.5.2   The Memory Chain Model for question answering behaviour in surveys: Theory and application

Antonio G. Chessa; University of Amsterdam, Netherlands

Bregje Holleman; University of Utrecht, Netherlands

Stasja Draisma; Vrije Universiteit Amsterdam, Netherlands

Abstract.   The question answering processes that underlie response behaviour in surveys are very complex, as these encompass neural processes behind memorisation of information about a survey issue and the response selection process. Mathematical models that integrate memory and response processes may be very helpful for gaining insight into these processes when such models are used to analyse observed answering behaviour of respondents.

The Memory Chain Model is a model that integrates both processes. This model formalises memory representations in the brain as a point process, which is a probabilistic class of models used for the location of individual features of such memory representations in different areas of the brain in time. The model describes four stages: (1) Initial memory encoding; (2) Storage of memory replicas in different parts of the brain, forming short-term and longer-term memory representations after a series of memory consolidation and decline processes (McGaugh, 2000); (3) At a test, memory retrieval is attempted, where the number of retrieved memory features depends on the quality of the cue(s) derived from the question asked; (4) Finally, recall takes place when the number of retrieved features exceeds a threshold. The model has been fitted successfully to several hundreds of memory retention data sets (e.g., see Chessa and Murre (2004) and Janssen, Chessa and Murre (2004)).

Stages (3) and (4) represent the 'survey part' of the process, which starts with question asking and ends with response or non-response. The response selection process requires additional attention in survey applications, because it may differ from survey to survey and because of competition between different answering options.

Two applications of the Memory Chain Model to answering behaviour will be discussed, regarding a factual survey and an attitudinal survey. Probability distributions of reaction times and of response options were derived from the model and fitted to survey data. The application to the factual survey shows that the model gives indications about the strategy adopted by respondents that lead to answering questions correctly or incorrectly. The application to the attitudinal survey indicates that wording effects caused by replacing the question word 'allow' by 'forbid' result from an increase of the response threshold when answering forbid-questions. This means that respondents find a 'yes forbid' answer more difficult to endorse than a 'not allow' answer, which offers an explanation for the so-called forbid-allow asymmetry, leading to longer answering times and lower scores for 'yes forbid' than 'not allow'.

References

Chessa, A.G., and Murre, J.M.J. (2004). A memory model for Internet hits after media exposure. Physica A, 333, 541-552.

Janssen, S.M.J., Chessa, A.G., and Murre, J.M.J. (2004). The reminiscence bump in autobiographical memory: Effects of age, gender, education, and culture. Memory. In press.

McGaugh, J.L. (2000). Memory –A century of consolidation. Science, 287, 248-251.

4.5.3 The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties

Allyson L. Holbrook; University of Illinois, Chicago, United States

Abstract. Behavior coding is one technique researchers use to detect problems in survey questions, but it has been primarily explored as a practical tool rather than a source of insight into theoretical understanding of the cognitive processes by which respondents answer survey questions. The latter is the focus of the current investigation. Using the data from a large study in which face-to-face interviews were taped and extensive behavior coding was done, we tested whether sets of respondent behavior codes could be used to distinguish respondent difficulties with comprehension of the question from difficulties associated with mapping a judgment onto the response categories provided, and whether characteristics of the survey questions and respondents could be used to predict when and for whom such difficulties would occur. Sets of behavior codes were identified that reflected comprehension and mapping difficulties, and these two types of difficulties were associated with different question and respondent characteristics. This evidence provides greater insight into the links between behavior coding and the processes by which respondents answer survey questions.

4.5.4 Origins and Developments of the Cognitive Models of Answering Questions in Survey Research

Mario Callegaro; University of Nebraska, Lincoln, United States

Abstract. During the last thirty years, several models have been proposed to explain how respondents process information in a survey situation. They refer to a standardized interview situation where the respondent is asked generally closed-ended questions and has to provide an answer selecting one the response options provided. The movement called Cognitive Aspects of Survey Methodology (CASM) gave stimulus to study how the respondent processes a standardized survey questions. It brought together cognitive psychologist and survey researchers to study cognitive factors that influence survey responses and the ways in which survey questions might be designed to take into account these factors. This paper presents thirteen models, three precursors, and two spin-offs models showing how a convergence is reached over the last years toward a model (four steps) that became the paradigm of the survey response process. Now the "four steps model" is taught in survey research handbooks. A comparison of the models is conducted to highlight commonalities and differences and to show how some ideas originated at the early stage of survey research and left off, are subsequently introduced again. The paper is concluded with an evaluation of the cognitive modeling idea, its empirical evidence, and its applications and some unresolved problems.

4.5.5 Cognitive factors of stability of answers to attitudinal questions

MichałWenzel; Public Opinion Research Center - CBOS, Poland

Krzysztof Zagórski; Public Opinion Research Center - CBOS, Poland

Abstract.   The paper will discuss the analyses of the unique panel data collected by CBOS in three waves in Poland, in 2004-2005. The uniqueness of the panel results from the fact that its second wave was conducted only one month after the first one, while the third wave was conducted a year later. Since one month is not a sufficient time for the substantial change in the attitudes on both individual and societal scale, our assumption is that inconsistency between the answers given in one month interval should be attributed to general instability of opinions and low quality of responses rather than to factual attitudinal change. This may be less true in respect of long-time changes, as reflected by inconsistency in the answers obtained in one year time. The paper tests the hypothesis that such cognitive factors as general skill (measured in years of education), ability to understand the questions (assessed by the interviewers) and self-declared interest in politics influence the stability of the answers both directly and indirectly, through shaping the respondent's attitude toward participation in the survey. Moreover, the hypothesis seems to be justified that this influence is stronger in a case of short-time inconsistencies than in those appearing in a longer time. An additive scale of answers' instability will be used as the main data quality indicator at the individual (not aggregate) level. Similar measure of positive response set, called also "Yes Sir"syndrome, will be used as an additional indicator of data quality.

4.5.6 A Hybrid, Affective-Cognitive Computational Model of the Survey Response Simulating the Dynamics of Candidate Evaluation in the 2000 U.S. Presidential Election

Milton Lodge; Stony Brook University, United States

Sung-youn Kim; Stony Brook University, United States

Abstract. In this study, we test two models of survey response –the on-line (Lodge et al 1995) and the memory-based process models (Zaller and Feldman 1992, Tourangeau, Rips and Rasinski 2000) - against empirical data via a computer simulation of the 2000 U.S. presidential election.

We represent the 1) on-line model, 2) 3 possible versions of memory-based process models, and 3) a hybrid model that integrates affect and cognition in a computational model based on the ACT-R cognitive architecture. Each of these models are built atop 4 common mechanisms:
    * the classic representation of knowledge as a semantic network in
    long-term memory and such classic recall mechanisms as practice, recency, via spreading activation;
    * an affective mechanism based on the 'Hot Cognition' (Abelson 1963)
    and 'Primacy of Affect' hypothesis (Zajonc,1983) and attitude priming effects (e.g., Fazio 1986);
    * automaticity (e.g., Bargh 1997, Lodge, Steenbergen, & Brau,1995); and
    * a mechanism converting internal representations into survey
    responses.
Employing each of the models as multiple agents (survey respondents), we simulate the empirical dynamics of candidate evaluation in the US 2000 presidential election captured by the National Anneberg Election Survey (NAES) 2000. The NAES 2000 was conducted almost daily for more than a year, between November,1999 and January, 2001. The total number of respondents are over 100,000 (For details, see Romer et al, 2003).
    Specifically, the survey responses (candidate evaluations) of
    5 self-identified ideological groups over the course of 5 major campaign events in the election were simulated.
    The results show that the hybrid model is not only theoretically necessary but also empirically outperforms the other models. Specifically, the hybrid model more closely tracks the empirically observed changes in the survey responses during the election.

4.5.7 Affect and False Beliefs About Candidates

Dan Cassino; Stony Brook University, United States

Abstract. My research, carried out experimentally at Stony Brook University in 2004 under the direction of Dr. Milton Lodge, deals with how individuals cope with differences between their implicit and explicit evaluations of a candidate. In the experiment, participants evaluated a candidate on the basis of twelve statements attributed to him. After each statement, participants made an implicit evaluation of the candidate, through a battery of response latency items, and after all twelve statements, participants made an explicit evaluation of the candidate, through a number of Likert and open-ended questions. The results of this study showed that, as expected, the implicit evaluation was primarily a result of the first few things learned about the candidate, while the explicit evaluation was largely a result of the most recent –and thus most accessible –statements attributed to the candidate. Due to the design of the study, the statements from which these evaluations were made were ideologically at odds with each other. Thus, participants were left with a great deal of conflict between their explicit and implicit evaluations.

The open-ended questions during the explicit evaluation section of the study give us an ideal window to study how participants rectify these opposing evaluations. Results show that those individuals who hold strong implicit affect (in either direction) for the candidate during the early stages of the experiment (and, thus, are expected to have the opposite reaction to the statements attributed to the candidate in the later stages of the experiment), attribute issue positions to the candidate that he did not hold in order to support their views. That is, when a participant who has a strong implicit evaluation of the candidate after the first few statements is called upon to make an explicit evaluation of the candidate will, quite unknowingly, make up reasons why they would like the candidate. Approximately forty percent of subjects made such an error in the open-ended evaluations, and these errors were neither related to the political knowledge of the participant, nor to the case with which they attended to the statements attributed to the candidate.

Furthermore, the open-ended explicit evaluations had a substantial impact on later implicit evaluations of the candidate. The evaluation made of the candidate, including the made up statements regarding his issue positions was used as information in making further implicit evaluations. The study not only uncovers a major bias in survey responses, but demonstrates the importance of considerations –the reasons individuals give for their evaluations.

4.5.8 Analyzing cognitive processes in CATI surveys with response latencies –an empirical evaluation of the consequences using different baseline speed measures

Jochen Mayerl; University of Stuttgart, Germany

Piet Sellke; University of Stuttgart, Germany

Abstract. Measuring reaction times, i.e. the time respondents need to answer a survey question, is a promising path to detect cognitive processes underlying the answering process. However, "There may be nothing scientifically less meaningful than the simple observation that subjects respond in x milliseconds"(Fazio 1990: 89). The interpretation of reaction time is only feasible with a solid theoretical background. Reaction time is a multidimensional construct and influenced by a vast number of cognitive processes, e.g. the mode of information processing or the cognitive accessibility of attitudes. The identification and monitoring of these different processes, therefore, is indispensable for a solid interpretation of reaction times. Additionally, reaction times are influenced by a vast number of biasing effects that have to be taken into account. Besides biasing effects triggered by the measurement instrument, the respondent's individual baseline speed has to be controlled for, as well. The baseline speed can be understood as the general speed of responding that is independent from the question content. Without controlling the baseline speed, one cannot compare and interpret respondents' differences in reaction time.

Several approaches have been suggested in order to control for baseline speed. For instance, Fazio (1990) proposed to compute response latency indices, e.g. by subtracting the baseline speed from the "raw"reaction time. Alternatively, Huckfeldt/Sprague (2000) proposed to control for baseline speed through using baseline speed as a covariate in multiple regression analysis.

In literature, several different methods to calculate the individual baseline speed are proposed. Firstly, the individual baseline speed can be computed by the mean reaction times of so called "filler latencies"–reaction times to questions, that are not connected theoretically or thematically to the interesting "target latencies"(Fazio 1990b). Secondly, some researchers use the mean reaction times of all items, including the target latencies (e.g. Bassili 2003). Thirdly, the level of difficulty of the filler latencies is an unsolved question: should the baseline speed be computed only with response latencies of very "easy"questions requiring little cognitive effort like sociodemographical issues (Bassili 1993, Shrum/O'Guinn 1993), thus controlling the general speed needed to access memory? On the other side, should the filler latencies imply more difficult questions, hence controlling additionally for factors like education and intelligence?

Consequently, there are unexplored inconsistencies concerning the use and measurement of baseline speed in the present research on response latencies. Thus, empirical investigations of different baseline speed measures are needed in order to analyze possible biasing effects of different baseline speed measures on empirical studies using response latencies.

In this paper, we explore different methods to measure individual baseline speed. Specifically, we compare reaction times of very "easy"questions versus more complex self-reflecting questions versus overall mean. These different methods are evaluated empirically by the following criteria:

1) Validity of the measurement;

2) Consequences of the use of different baseline speed measures, e.g. concerning the moderational power of response latencies controlled with different measures of baseline speed as well as interpreting response latency as a measure of attitude accessibility. To do so, the temporal stability of attitudes as well as attitude-behavior-relationships are investigated.

To do so, measures of behavior and attitudes in the context of money donation and health conscious nutrition were used. For this analyses, we used a nation-wide panel-survey (CATI) conducted in 2005 in Germany.

4.5.9   An Overview of the Respondent-Generated Intervals (RGI) Approach to Sample Surveys

James Press; University of California, Riverside, United States

Judith Tanur; State University of New York, United States

Abstract. This paper provides an overview of research on the Respondent-Generated Intervals, or RGI, protocol for asking recall-type questions in sample surveys. The RGI protocol for asking questions in sample surveys involves asking each respondent not only for a basic answer to a recall-type question, but also, for a smallest value his/her true answer could be, and a largest value his/her true answer could be. We have developed a Bayesian hierarchical model for analyzing such RGI data from a survey. With the development of the model came the need to explore how well the model might work in practice, with real people and real empirical data. We examined pencil-and-paper classroom surveys, and a telephone survey using Census data and provide results of these surveys.

4.5.10 Enhancing recall accuracy in surveys by cognitively tailored timeline methods: a record check study

Wander van der Vaart; Vrije Universiteit Amsterdam, Netherlands

Tina Glasner; Vrije Universiteit Amsterdam, Netherlands

Abstract.   In order to reduce recall error in survey data, social researchers at times employ aided recall techniques. By providing respondents with cues, landmarks, time frames and the like, these techniques aim to help them to gain better access to long-term memory. This paper examines the effect of a graphical timeline procedure that integrates a number of such measures, and aims to shed more light on the way timeline techniques relate to cognitive processes involved in answering retrospective questions. Our timeline was designed to fit the commonly assumed hierarchical network structure of memory and especially Conway's (1996) multi-level model of autobiographical memory. This model reflects the fact that autobiographical events are imbedded in a context of ongoing life experiences. It assumes memory structures that consist of autobiographical information at three levels of abstraction which are highly interrelated (thematically organized lifetime periods, generic events and specific-events respectively).

In the present study the hierarchical design of the timeline intended to incorporate both high-level retrieval cues about lifetime periods as well as more specific cues. Also, the timeline aimed to promote the use of sequential associations of what happened earlier and later among events within lifetime themes (e.g. regarding residences, relationships, etc.), and the use of parallel associations among contemporaneous events and/or landmarks between such themes.

We performed a split-ballot field experiment on our (paper-and-pencil) timeline procedure in a Dutch telephone survey on consumer purchase behavior; record check information was available as gold standard for recall accuracy (N=214). In two Dutch towns a random sample of consumers was asked to report the purchases of optical devices like lenses and glasses. The timeline covered a recall period of two-and-a-half years (lenses) and seven years (glasses) respectively. It was applied as a separate device in addition to a standardized questionnaire and was sent to the respondents beforehand. The retrospective reports of consumers were compared to purchase database information regarding the number of products bought, the purchase dates and the price of the products. The paper explores the expected beneficial effects of the timeline on recall accuracy and evaluates the application of cognitive principles in it's design.

4.5.11 The reminiscence bump and the telescoping effect

Steve M.J. Janssen; University of Amsterdam, Netherlands

Antonio G. Chessa; University of Amsterdam, Netherlands

Jaap Murre; University of Amsterdam, Netherlands

Abstract. In this article, we examined the interaction between two robust cognitive phenomena: the telescoping effect and the reminiscence bump. The telescoping effect is the displacement of recent events backwards in time and the displacement of remote events forward in time. The reminiscence bump is the effect that people recall more personal events from their early adulthood than from the periods before (childhood) or after (adulthood). According to our memory-for-time model people should displace events from their early adulthood less than events that originate from the periods before and after. Furthermore, we examined the effect of dating formats, presentation order of the dating formats, and age of the participants on the size of the telescoping effect and on the moment that backward telescoping shifts to forward telescoping.

4.6 Data linking and integration

Session chair:
    Mariano Torcal; University Pompeu Fabra, Spain

This session takes place on:
    Wednesday, July 20; 17:30 to 19:30

Presentations:

- Invited presentation: *Introduction* - Mariano Torcal (Spain)

Contributed presentations:

- *Social Inequality and Environmental Pollution: Empirical Results from a Survey Linking Subjective and Objective Exposure Data* - Reto Meyer (Switzerland)

- *Problems in the Integration of Comparative Survey Data Sets: The ISSP "DataWizard" Software as a (Partial) Solution* - Markus Quandt (Germany)

- *Issues of relating composite surveys on higher education and labour market* - Andromachi Hadjiyanni, Dimitra Kondyli (Greece)

4.6.1 Introduction

Mariano Torcal; University Pompeu Fabra, Spain

4.6.2 Social Inequality and Environmental Pollution: Empirical Results from a Survey Linking Subjective and Objective Exposure Data

Reto Meyer; ETH, Switzerland

Abstract. Some environmental surveys conducted by sociologists in German-speaking countries have found a positive relationship between the SES of the respondent and his or her perception of exposure to several kinds of environmental pollution. This is a surprising result, since one would expect SES to be positively associated access to less polluted neighborhoods. However, not all research has produced this unexpected result. For example, various epidemiological studies have instead found the expected negative correlation between SES and reported exposure to pollution. The root of this discrepancy may be methodological differences among the studies. We are therefore conducting a methodologically rigorous study on environmental justice in the form of a survey on exposure to environmental pollution in an attempt to resolve this discrepancy. We also intend to examine the relationship between subjective and objective exposure to pollution. One such effort will involve linking subjective data on exposure to noise pollution with objective data on traffic noise exposure. We are currently conducting several large-scale pretests of our instrument and linking techniques. The results presented here are from three pretest surveys conducted with large samples (n=1200, n=990, n=110) in two Swiss cities. These survey responses have been supplemented with official noise cadastral data from the city administrations. The relationship between the objective noise exposure and subjective perceptions as a function of the respondent's SES will be discussed.

4.6.3 Problems in the Integration of Comparative Survey Data Sets: The ISSP "DataWizard"Software as a (Partial) Solution

Markus Quandt; University of Köln, Germany

Abstract. Over the last decades, a number of large international survey projects have been established. Apart from much discussed problems of comparability of survey questions etc., such programmes are also confronted with particular organisational and technical tasks. One of these tasks is that of integrating data sets from the different participating countries into one common analysis file. This task has so far been left to ad hoc procedures, relying on the 'on board' tools of the usual statistical software packages.

The ad hoc approach has contradictory implications for the ideal distribution of work between partners in a comparative project: On the one hand, full quality control of data harmonisation procedures is easiest if all stages of the data cleaning and recoding process are done centrally. On the other hand, a central institution often lacks the familiarity with each country's characteristics that may be necessary for a valid recoding of the national data to a common standard.

The International Social Survey Programme (ISSP) as a large, long-standing comparative survey project has therefore adopted a partly distributed, partly centralised structure to produce its annual integrated data sets. To further increase the efficiency of this structure, the ISSP is presently implementing a specialised software tool, the ISSP DataWizard. Technically, the software streamlines the process of adapting several different source data sets to a predefined reference definition of variable descriptions and coding schemes. Such reference definitions or 'Standard Setups' can be centrally prepared and distributed for processing to cooperating countries or parties, with processing documentation and results being later returned for checking and final integration steps to the central institution.

In spite of its origin from a particular survey programme, the DataWizard promises benefits for other comparative projects. Usage scenarios that involve only single workplaces are equally possible. Its application is of course not limited to international comparisons, but the software can also handle structurally similar settings like matching data sets from the same population but different points in time. The software is written in Java and thus platform independent and can accommodate different database backends. By implementing the Data Documentation Initiative's (DDI) standard for data description, the DataWizard is prepared for integration into the rapidly emerging software environments following that standard.

The presentation will then give a glimpse on core working steps of data processing with the DataWizard.

The DataWizard is a joint development by the Information Centre for the Social Sciences, Bonn, and the Central Archive for Empirical Social Research, Cologne.

4.6.4   Issues of relating composite surveys on higher education and labour market

Andromachi Hadjiyanni; EKKE, Greece

Dimitra Kondyli; EKKE, Greece

Abstract.   A subject matter base on tertiary education and its connection to the labour market was designed to collect, process and document primary and secondary data derived from various sources, so as to provide researchers with easy and user-friendly access to scientific data, information and methodological tools (such as questionnaires, codebooks etc.) that are often spread, inaccessible or even unknown. The specific subject matter base developed within the framework of the infrastructure "Node –Development for Secondary Analysis and Administration".

The development of the object matter base relied on the experience from the research project "Integrating young people in the labour market: The case of social sciences university graduates"which was implemented by the National Centre for Social Research in the period 1998-2000, and included:

- An empirical research using a structured questionnaire administered to 620 social sciences graduates. The initial empirical research was designed to enable recording of both the educational and occupational progress of tertiary education graduates, as well as their views on their occupational experience

- A labour market study on the demand for high-skilled work force by using face to face interviews.

- A labour market research study based on secondary data derived from the National Statistical Service of Greece as well as the Ministry of Education.

The value of the methodological undertaking of developing this object matter base consisted s in its goal to:

- Document composite research on particular subjects, which allows the simultaneous recording of multiple objects of observation as for instance graduates students , and their education paths and career.

- Carry out longitudinal recording of subjects and enable comparabilities over time and space.

- Enable systematic connection of empirical research (or field research) both to secondary processing results and results from related surveys conducted in Greece.

- Provide an appropriate environment both for secondary data analysis of the existing datasets and designing or enhancing future research.

- Enable navigation between datasets through links based on common concepts, questions and/or variables in order to enhance our knowledge on the research procedure

- Provide University Departments access to the object matter base as a tool for regular outcome assessment through following up students' entry into the labour market.

4.7   Nonattitudes and informed opinions

Session chair:
>   Peter Neijens; University of Amsterdam, Netherlands

This session takes place on:
>   Thursday, July 21; 9:00 to 11:00
>   Thursday, July 21; 11:30 to 13:30

Presentations:

- Invited presentation: *The Dutch referendum on the European Constitution: (non-) attitudes, information and choice* - Claes de Vreese, Peter Neijens, Andreas Schuck (Netherlands)

Contributed presentations:

- *A unified model for the survey response process: Estimating the stability and crystallization of public opinion* - William van der Veld (Netherlands), Willem Saris (Spain)

- *Knowledge Deficit? The Effect of Science Information on Attitude to Genomic Science* - Patrick Sturgis, Ian Brunton-Smith, Helen Cooper, Chris Fife-Schaw (United Kingdom)

- *The Information-Choice Questionnaire versus a Traditional Questionnaire: The influence of expert information on lay people's opinions* - Marjolein de Best-Waldhober, Dancker Daamen (Netherlands)

- *What is the use of asking citizens for feelings of safety complementary to police figures: Effects of providing feedback* - Jeroen Slot, Willem Bosveld (Netherlands)

- *Give the people a voice - the application of the choice questionnaire in a local decision making process* - Joyce Neys, Robert Voogt (Netherlands)

- *Public Opinion on the European Constitution: non attitudes, framing, and informed opinions. A Joint Presentation of Four Interrelated Papers.* - Claes de Vreese, Andreas Schuck, Peter Neijens (Netherlands)

4.7.1   The Dutch referendum on the European Constitution: (non-) attitudes, information and choice

Claes de Vreese; University of Amsterdam, Netherlands

Peter Neijens; University of Amsterdam, Netherlands

Andreas Schuck; University of Amsterdam, Netherlands

Abstract.   In the paper we bring together different strands of research from public opinion research, political psychology, and communication science. Our starting point is that citizens hold non-attitudes (or non-crystallized attitudes) on a number of issues. Framing affects citizens' opinions by making certain considerations more salient. Framing can take the form of a survey question which affects respondents. However, framing is also part of larger societal discussion of issues, especially in the media which is the key resource for many citizens for political issues. Moreover, we extend the notion of framing to also look at the concept's potential for improving opinions by leading to (better) informed opinions and choices. We hereby bring together work on framing and on questionnaire design research. The topic of our research is the public opinion on the European Constitution. This topic is particularly interesting because of its social, economic, and political implications, but most importantly because it represents an area in which opinions are probably volatile and easily susceptible to influences.

Research Questions

First, we investigate antecedents and structures of public opinion about and support for the European Constitution. We specifically look at change in public opinion and how that relates to exposure to (media) information during the campaign. Second, we study the differences between naïve and informed opinions and the conditions under which they differ.

Data

We collected survey data on public opinion in The Netherlands (representative two wave panel) and media content analysis data. Embedded in the survey was an experiments designed to study the effect of information on the European Constitution.

4.7.2   A unified model for the survey response process: Estimating the stability and crystallization of public opinion

William van der Veld; University of Amsterdam, Netherlands

Willem Saris; ESADE, Universitat Ramon Llull, Spain

Abstract.   In the past, survey researchers assumed that the respondents held opinions which could be obtained by asking the proper questions. This idea - the "file drawing model"for survey response - was criticized strongly by Converse in 1964, but also strongly defended by Achen (1975), Erikson (1978, 1979), and Judd et al. (1981). Zaller (1992) introduced the idea that people create their response on the spot when they are asked a question on the basis of the most salient considerations. The saliency of the different considerations can vary with the context in which the question is asked and the formulation of the question itself. This approach has gained a lot of attention in the last 20 years (Tourangeau et al., 2000; Tourangeau & Rasinski, 1988; Wilson & Hodges, 1991).

In this study a model has been developed that unifies these ideas into one model (van der Veld & Saris, 2004). For this purpose a theory is needed that describes how a question leads to a response. This theory should account for the fact that the response could be different from the opinion, due to the (bad) quality of the measurement procedure. Furthermore, the model should allow that an opinion is made up of pre-existing or stable considerations and temporary or unique considerations. A structural equations model that follows from the mathematical description of the survey response for an individual has been developed. With this structural equations model it is possible to decompose the variance of responses into three different components that also can be linked to commonly used concepts in survey research. The variance of the responses will be decomposed in variance due to pre-existing considerations, temporary considerations, and measurement error. These three variance components can be linked to opinion stability, opinion crystallization, and measurement quality respectively.

4.7.3   Knowledge Deficit? The Effect of Science Information on Attitude to Genomic Science

Patrick Sturgis; University of Surrey, United Kingdom

Ian Brunton-Smith; University of Surrey, United Kingdom

Helen Cooper; University of Surrey, United Kingdom

Chris Fife-Schaw; University of Surrey, United Kingdom

Abstract. Scholars of public opinion are justifiably reticent about the use of survey measures of issue preferences to represent the 'will of the people'. Evidence abounds that citizens in modern capitalist democracies have a shallow grasp of matters politic and provide labile and ephemeral attitudinal responses as a result. Various methods have been proposed as solutions to this 'nonattitude' problem, all of which, by various means, aim at eliciting what has been termed 'informed opinion' (Price and Niejens, 1997; 1998). Some of the more prominent examples in this regard are Deliberative Polling (Fishkin 1997); Planning Cells (Dienel 1978); Choice Questionnaires (Neijens 1987); and simulation modeling (Bartels, 1996; Althaus, 1998). In this paper we report on a study using a new approach to estimating informed opinion; information about genomic science is provided to respondents as a short film incorporated into the survey interview. The basic design was as follows: in June 2003 a probability sample was drawn from the target population and respondents administered a questionnaire via Computer Assisted Personal Interview (CAPI). Approximately six months later respondents were re-contacted and randomly allocated to treatment and control groups. Treatment groups were presented with factual information about genomic science in the form of a short film as part of the CAPI interview, the control group receiving no information. All respondents were then re-administered the original questionnaire, with the causal effect of the information being equivalent to the average of the difference in opinion between the two groups. A third wave of data collection, via Computer Assisted Telephone Interview (CATI), was then conducted approximately three months after second interview.

While this approach has clear limitations in the depth of information that can be communicated relative to other methods of estimating informed opinion, its strengths lie in its use of random sampling and the high penetration of information provision across the sample as a whole. Our analysis evaluates the effect of viewing the film on attitudes to eight different areas of genomic science, using a multi-level model for repeated measures. We also examine respondents' own appraisals of the accuracy and relevance of the information provided and how the effect of the information on attitude interacts with educational attainment. We conclude by considering the general utility of the approach for estimating informed opinion.

4.7.4   The Information-Choice Questionnaire versus a Traditional Questionnaire: The influence of expert information on lay people's opinions

Marjolein de Best-Waldhober; University of Leiden, Netherlands

Dancker Daamen; University of Leiden, Netherlands

Abstract.   If one wants to know opinions of the general public regarding a specific policy issue, it is often a fruitful approach to "just ask them". That is, to design a questionnaire on the issue and administer it to a sample representative for the population. However, such conventional mass opinion surveys may fall short if the issue concerns a policy problem for which the options for solution are new to the public. Respondents may then simply lack the knowledge to have opinions. Part of them may refrain from answering but a significant part of them may respond with "pseudo-opinions"or "non-attitudes"(cf. Converse, 1964). Several studies show respondents' inclination to give an opinion even on topics they know nothing about (Bishop, Oldendick & Tuchfarber, 1986, Schuman & Presser, 1981). Other research showed that such pseudo-opinions are unstable and easily changed by contextual information (e.g., Strack, Schwarz & Wänke, 1991).

As the risk of pseudo-opinions holds especially for new and complicated issues, the issue of modern technologies, global warming and their relationship might well be the kind that is at risk for pseudo-opinions. We tested the knowledge of the Dutch public on these issues. Specifically, we presented a sample of the Dutch population (n=327) with knowledge questions and opinion questions about global warming and about six modern technologies which use fossil fuels for energy combined with carbon capture and storage as a way to reduce CO_2 emissions. Our data show that a substantial part of the respondents lacks even the most basic knowledge that is needed to have (or construct) a well considered opinion on these issues. Our data furthermore show that only part of the respondents who state their lack of knowledge withhold themselves from giving their opinions. For instance, on average half the respondents who just admitted to having never heard of a specific modern technology, did give an overall evaluation of this technology. The stability of these evaluative judgments was low: Within twelve minutes, we asked again for the overall evaluations of each of the six technologies, after adding very little information. The correlations between the first and the second evaluation of each of the six technologies ranged around a mere 0.35.

These results all suggest a necessity for more informed opinions than can be obtained with traditional surveys on new and complicated issues. One technique that aims for more deliberated and better informed opinions is the Information-Choice Questionnaire (ICQ) (Neijens, 1987; Van Knippenberg & Daamen, 1996). In an ICQ, respondents are confronted with a policy-relevant decision problem. However, before choosing, respondents are provided with well-balanced expert information on the most important consequences of each option. This information is compiled by various energy experts with different backgrounds, checked by other experts and then translated for lay people. The processing of this information is facilitated. We developed an ICQ on topics identical to those in the traditional survey we described earlier. Contrary to this traditional survey (TS) though, the respondents in the ICQ were provided with information. This ICQ was administered to a representative sample of the Dutch population (n= 995). In our presentation, we will discuss ICQ results as well as TS results and discuss factors that can influence the effects of the differences between these types of surveys.

References

Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. Public Opinion Quarterly 50, 240-250.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (ed.) Ideology and discontent. New York: Free Press.

Neijens, P. (1987): The Choice Questionnaire. Design and Evaluation of an Instrument for Collecting Informed Opinions of a Population. Amsterdam, Free University Press.

Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys. New York: Academic Press.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. Social Cognition, 9, 111-125.

Van Knippenberg, D., & Daamen, D. (1996). Providing information in public opinion surveys: Motivation and ability effects in the information-and-choice questionnaire. International Journal of Public Opinion Research, 8, 70-82.

4.7.5 What is the use of asking citizens for feelings of safety complementary to police figures: Effects of providing feedback

Jeroen Slot; Amsterdam Bureau for Research and Statistics, Netherlands

Willem Bosveld; Dienst Onderzoek en Statistiek, Amsterdam, Netherlands

Abstract.   Since 2004 the 'objective' and 'subjective' safety of Amsterdam citizens is studied. The objective measure is derived from police figures and the subjective measure is derived from surveys among Amsterdam citizens.

We study the association between both measures in different ways. Firstly, we look at associative measures, secondly we study the effect of events and thirdly we study the impact of providing respondents in the survey with feedback on both objective and subjective outcomes.

In one condition respondents are first explained how objective information is gathered and then they are asked to estimate this measure for their own neighborhood. Then they are provided with feedback on the discrepancy between their estimate and the sample information. In a second condition this procedure is followed for subjective figures.

In both conditions respondents are then asked to attribute the discrepancy: was their estimate false or were the figures provide false. We hypothesize that no differences in attributions occur between both conditions, suggesting the independent effect of information about on of respondents perceptions and the relevance of this measure in analyzing safety in cities.

4.7.6 Give the people a voice - the application of the choice questionnaire in a local decision making process

Joyce Neys; University of Amsterdam, Netherlands

Robert Voogt; Ministry Of Social Affairs and Employment, Netherlands

Abstract. In The Netherlands there is no tradition of referenda, like for example in Switzerland. Still, Dutch law makes it possible for local communities to organize referenda, to ask their inhabitants for their opinion about an issue, before making a final decision about this issue. In the last few years, some referenda have been organized,

especially in larger Dutch cities, like Amsterdam and Groningen. An important disadvantage of a referendum is that it is a very expensive way to ask the people for their opinion. Another disadvantage is that is does not take into account the possibility that many people do not have the information to make a deliberate choice. The choice questionnaire offers a solution for both problems. Instead of asking everybody, only a representative sample of the population is asked to participate and people are presented the information they need to make a deliberate choice.

In the Dutch municipality of Loenen (which contains five villages with a total of 8500 inhabitants) the choice questionnairy is used to give the inhabitans a voice in the decision about cuts on spending on welfare facilities –like community centers, sport facilities, child care and grants for local associations and clubs. This decision to involve the inhabitans was a reaction of the municipality council on large scale protests against an earlier proposal of the local government to cut all spending on welfare with 50% –without making any distinctions. The municipality council wants to come to a more balanced and better argued proposal for the cuts of the welfare budget. The opinion of the inhabitants, as measured in the choice questionnary, is an important element in the process that will lead to this final proposal.

On the conference we will present the design of the choice questionnaire, the way in which we gathered the information needed to compose the questionnaire and some first results.

4.7.7 Public Opinion and the European Constitution: does it matter if a conflict is about issues, persons or strategies?

Claes de Vreese; University of Amsterdam, Netherlands

Andreas Schuck; University of Amsterdam, Netherlands

Peter Neijens; University of Amsterdam, Netherlands

Abstract. Conflict and disagreement are features of information exchange that are emphasized as potentially beneficial to the vitality and health of democratic processes. However, while these features are celebrated in the context of interpersonal discussion, they are pundits when it comes to information mediated through news media. News media often present issues and events in terms of conflict, but the type and focus of the conflict may vary, and this variation can cause differential effects.

This paper assesses the differential effects of conflicts about issues, persons or strategies in the context of a referendum on the European Constitution on the balance of respondents' thoughts, their argument repertoire and perception of the referendum campaign.

An experiment was embedded in a national survey providing respondents with information about the EU constitution. Respondents were randomly assigned to one of three types of conflict (issues, persons, or strategies). In each condition respondents received information from 'a recent newspaper report' focusing on either a content-based differences of opinion (issue conflict), personal disagreement (person conflict) or disagreement about campaign strategies (strategy conflict).

It was hypothesized that exposure to issue conflict would result in more balanced considerations about the Constitution, lead to a 'deeper' argument repertoire and result in a more positive perception of the campaign compared to exposure to either person or strategy based conflict. It was further expected that this relationship would be moderated by the individuals' level of political sophistication so that in particular highly sophisticates would benefit from exposure to issue conflict.

The results are discussed in the light of the literature that highlights both the positive and negative impact of conflict-driven information on opinion quality and formation.

4.8   Event data across Europe: a measurement problem

Session chair:
  Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

This session takes place on:
  Thursday, July 21; 15:00 to 17:00
  Thursday, July 21; 17:30 to 19:30

Presentations:

- Invited presentation: *Events across Europe: Why and how to collect media-reported events* - Ineke Stoop (Netherlands)

Contributed presentations:

- *Identifying National and International Events Using Computer-Assisted Content analysis* - Cornelia Zuell, Juliane Landmann (Germany)

- *Data events and attitudes in Slovakia* - Katarína Vasilová, Michal Kentoš (Slovakia)

- *The impact of terrorist acts on Dutch society: The case of the Van Gogh Murder* - Enny Das (Netherlands), Brad Bushman (United States), Marieke Bezemer (Netherlands)

- *Analysing and Measuring Event Data Across Europe: a comparison between weekly and monthly event reporting* - Theoni Stathopoulou (Greece)

- *Event Data: an useful way of accessing European's attitudes and opinions* - Anália Torres, Tiago Lapa (Portugal)

4.8.1 Events across Europe: Why and how to collect media-reported events

Ineke Stoop; Social and Cultural Planning office (SCP), Netherlands

Abstract.   In the early stages of the preparation of the European Social Survey (ESS), during the final years of the last century, when the construction of an event data inventory was planned as a corollary to survey data collection, the most disruptive events most people had in mind that could influence individual reactions to certain questions were national elections. The 9-11 terrorist attacks in the USA in 2001 and its aftermath showed that events could have an infinitely more far-reaching effect on values and opinions. When designing the actual data collection for the event inventory in early 2002, it was still assumed that similar high impact events would occur only very sporadically. Sadly, this has not been the case. Now, in the Spring of 2005, the importance of collecting event data during fieldwork seems more pertinent than ever before.

The fact that the fieldwork of the first waves of the ESS took place against the background of a war in Iraq (the preparation in Round 1, the aftermath in round 2), when the influence of international organizations was under discussion, when European countries were seriously divided among themselves and new countries entered the EU, is something future users of the ESS should be aware of, especially as the study has been set up as an important asset for historical micro analysis. The impact of international events, such as the war in Iraq, on the social and political fabric of Europe is something substantive researchers should especially take into account as this international event translated itself or was related to many distinctive national events, such as attitudes towards the Blair government in the UK, the overthrow of the Aznar government in Spain, the involvement of national armies in the Iraq war, hostage situations involving different countries, and so on. The single event of the war in Iraq and the impending threat of terrorism thus may not have influenced societies in a way that was uniform across countries.

To pinpoint the impact of the national 'Iraq events' or 'tsunami events', the involvement of local reporters is required who can relate events to the content of the ESS questionnaire. During fieldwork they forwarded perceptive overviews of what kept people –or more precisely –newspapers busy during the fieldwork of round 1 in monthly reports and of round 2 in weekly reports. From the inventory of national media-reported events it can be seen that both waves of the ESS were conducted in turbulent times: reports mention regime changes, terrorist attacks, major floods, heated public discussions on immigration and asylum seekers, political wranglings, financial scandals, planned revisions of social protection regimes, economic decline and national elections. When analyzing attitudes, the event inventory should provide the necessary background. The complete overview of the events that took place in the participating countries during fieldwork is publicly available at the ESS websites.

4.8.2 Identifying National and International Events Using Computer-Assisted Content analysis

Cornelia Zuell; Centre for Survey Research and Methodology (ZUMA), Germany

Juliane Landmann; Centre for Survey Research and Methodology (ZUMA), Germany

Abstract.   Since the beginning of the first data collection period of the ESS, the events occurring during data collection have been documented in all countries of the ESS. The assumption is that respondent behaviour or answers to some questions are influenced by significant events in various areas. Therefore, the impact of an event must be considered and, whenever possible, controlled across the countries in the ESS. Unfortunately, manual documentation of events is a very time-consuming task and therefore we will discuss ways of using computer-assisted content analysis for identifying such events.

The most frequently used approach in computer assisted content-analysis is the dictionary-based approach. This approach requires an a-priori developed dictionary defining all possible events and, additionally, when different countries participate in the ESS, these dictionaries must be available in the appropriate languages. To slightly reduce the efforts of building new dictionaries, one can test the usability of already developed dictionaries, for example those in programs like Keds/Tabari, or take advantage of technical aids for computer-linguistics like lemmatisation, part of speech tagging, or synonym handling. Obviously, building new dictionaries would require a great deal of time for development.

Another approach is based on co-occurrences of words. The co-occurrences of words in a text unit define a matrix of similarities between words and this matrix can be further analysed by classification procedures like cluster analysis or multidimensional scaling. Classified words found with this analysis could be used to identify the major events in newspapers. Nevertheless, some pre-analysing steps are necessary to get interpretable results:

First, which text types in newspapers form the best data base (title of articles, titles and the first sentence, editorials or reports, etc.) to identify events? Secondly, how can meaningful words within texts be traced or meaningless words eliminated (defining stop words, using reference corpora)? Finally, is it necessary to handle synonyms and lemmas before conducting such an analysis?

We will discuss these aspects in our paper, using articles published in newspapers in Germany (FAZ) and Great Britain (Guardian) between August and October 2002 and will compare our findings with the reports of the National coordinators from these countries during the respective period.

4.8.3   Data events and attitudes in Slovakia

Katarína Vasilová; Slovak Academy of Sciences, Slovakia

Michal Kentoš; Slovak Academy of Sciences, Slovakia

Abstract. The presented study is focused at the event data reports carried out within the framework of the ESS 2nd round. Significant events recorded during the fieldwork in Slovakia were analyzed. Although it is difficult to express the extent of the influence of exogenous factors on the response tendency, the events are evidently related to the content of the questionnaire. The analysis of the potential influence of exogenous factors –significant events on the change of attitudes in selected aspects was done. The research sample consisted of 1524 respondents participating in the 2nd round of European Social Survey and was selected by random sampling from all the residents of Slovakia.

4.8.4   The impact of terrorist acts on Dutch society: The case of the Van Gogh Murder

Enny Das; Vrije Universiteit Amsterdam, Netherlands

Brad Bushman; University of Michigan, United States

Marieke Bezemer; Vrije Universiteit Amsterdam, Netherlands

Abstract. In an experimental study on the effects of news on terrorism, a representative sample of 100 Dutch participants saw a news program about either Islamic terrorist acts or the Olympic Games (control condition). Participants subsequently read either a positive article about the integration of Arabs into Dutch society, or a negative article. During data collection, Dutch filmmaker Theo van Gogh was assassinated in Amsterdam by an Islamic extremist. The murder was included as a naturally occurring factor in the design, allowing us to assess reactions before and after the murder. To assess psychological defense mechanisms against terror, participants' accessibility of death-related thoughts was assessed by a word completion measure. The main dependent measures were attitudes about the integration of Arabs into society and toward Arabs (and Asians as a control). Consistent with terror management theory, terrorism news and Van Gogh's murder increased death-related defense mechanisms, which in turn predicted prejudiced attitudes toward Arabs. Terrorism news and Van Gogh's murder also increased negative attitudes about the integration of Arabs into Dutch society. News on terrorism may inadvertently increase negative attitudes toward Arabs when it confronts viewers with their own mortality.

4.8.5   Analysing and Measuring Event Data Across Europe: a comparison between weekly and monthly event reporting

Theoni Stathopoulou; EKKE, Greece

Abstract. The collection of media-reported event data for the European Social Survey (ESS) aimed at adding timing and context to the formation of survey respondents' opinions and attitudes. In an effort to extract the most out of the database created by the first round of the ESS, we proposed a new model for events analysis. This model used a combination of text mining, time series, and cluster analysis techniques to monitor the evolution of the events in time and space and map the impact of events in various countries, thus tracking the formation and flow of "event identities"across the participating countries of the survey.

While the event data of ESS Round 1 were collected on a monthly basis, the event data of ESS Round 2 were collected weekly. On the basis of the methodology developed after ESS Round 1 outlined above, this paper will compare the weekly data with the monthly data in order to:

1. study the quality of incoming information on a weekly basis;

2. study the degree to which the weekly information creates clearly defined groups over time as regards content and quality; and

3. study the degree to which the weekly collection of events increases or decreases noise levels in the analysis as compared to monthly collection.

Keywords: ESS, event data, text mining, cluster analysis.

4.8.6 Event Data: an useful way of accessing European's attitudes and opinions

Anália Torres; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Tiago Lapa; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Abstract. With the help of some examples, we try in this paper to show how the analysis of event data can be useful as an instrument more for a better understanding of attitudes, opinions and the formation of public opinion. Event data is the data about socio-political events (such as the proximity of an election, economic or political crisis, or even a natural disaster), collected from the headlines of national newspapers during the fieldwork of the "European Social Survey".

Based on the registration of the main events in newspapers headlines, event data might be very useful in several ways. It can give us a wider vision of how main events were (or not) shaping public opinion in the period of the application of the questionnaire (2002, 2004) at national or international level. Are newspapers, and other media, really shaping and inflecting opinions? Are some people more "permeable"than others? We can compare the impact of some events with the opinions people revealed in the survey and analyse possible effects. We can also assess the wide possibilities of the event data, comparing the answers people gave in 2002 and outline some expectations for what they will answer in 2004 and then corroborate or not our hypothesis.

International events like the war on Iraq or the Buttiglione's crisis in the European Parliament, in 2004, might have influenced attitudinal data like the trust in the United Nations or in the European Institutions. National events like the dismissal of a government might affect an array of political issues, like the satisfaction with the government's performance, the economy or the state of democracy. We will analyse some examples and we will try to establish some interpretations based on the event data analysis made in 2002 and in 2004.

4.9   Measurement of social classification: which variables should be used to define social classification?

Session chair:
   David Rose; University of Essex, United Kingdom

This session takes place on:
   Friday, July 22; 9:00 to 11:00
   Friday, July 22; 11:30 to 13:30

Presentations:

- Invited presentation: *The Measurement of Social Classifications: Classes and Scales* - David Rose (United Kingdom)

Contributed presentations:

- *How good are International Standard Classifications? The case of ISCO88* - Peter Elias (United Kingdom)

- *The reliability of stratification measures* - Anette Scherpenzeel (Netherlands)

- *Class Demography and Gender Inequality in Europe: Alternative Interpretations* - Hakon Leiulfsrud, Ivano Bison (Norway)

- *Specificity and Universality in Occupation-based Social Classifications* - Paul Lambert, Ken Prandy (United Kingdom), Max Bergman (Switzerland)

- *The Development of a new European Socio-economic Classification* - David Rose, Eric Harrison (United Kingdom)

- *Classification systems and contextual difference. Some comments on occupational hierarchies from a South European angle* - Thomas Maloutas (Greece)

- *Measurement errors and consequences in student's proxy reports about the socio-economic status of their parents* - Frauke Kreuter (United States), Kai Maaz, Rainer Watermann (Germany)

4.9.1 The Measurement of Social Classifications: Classes and Scales

David Rose; University of Essex, United Kingdom

Abstract. There are two basic approaches to socio-economic classification (SEC): those using categorical measures and those employing continuous measures. That is, some SECs divide the population into a discrete number of categories or social positions. Others allow for an unlimited number of graded distinctions. The first may be called the class approach and the second the status approach. However, each form relies heavily on occupational information for its construction. The earliest continuous measures combined information on occupation, education and income, i.e. summarised social and economic variables relating to occupations. More recently, measures such as the Cambridge Scale have been constructed on the basis of occupational information from friendship and marriage patterns. However, many social scientists prefer to explore the effects of occupational position alone. Thus, there is another tradition in socio-economic classification. This one sees individuals as occupying common class positions in the social structure. Typical of modern approaches are the EGP class schema and the Wright class model. All of these approaches will be discussed in the papers being presented in the sessions on the measurement of social stratification. This paper will provide an overview of approaches and related measurement issues.

4.9.2 How good are International Standard Classifications? The case of ISCO88

Peter Elias; University of Warwick, United Kingdom

Abstract.   In 1990 the International Labour Office introduced a new international standard for the classification of occupations (ILO, 1990). Like other classifications in the 'International Family of Social and Economic Classifications' (UN, 2002) the new occupational classification (ISCO88) was designed to improve international comparability of information on occupational structure in two ways. First, it was intended to help foster a common understanding of occupational structures between countries, possibly via mapping from national classifications to ISCO88. Second, for those countries without a national classification of occupations, it would provide an 'off-the-peg' solution to their need for such a statistical instrument.

It has been argued elsewhere (Ganzeboom and Treiman, 1996) that comparative research on social mobility and social status stood much to gain much from adopting ISCO88 as the standard tool of classification. This paper considers the merits of this argument in the light of 10 years of experience using ISCO88 as the common standard for reporting occupational information within the European Union and from current attempts to use ISCO88 as a means of creating a harmonised European Social Classification. The paper concludes by describing work now underway to remedy some of the main problems with ISCO88 –the development work for ISCO2008.

References

Ganzeboom, H. B. G. and D. J. Treiman (1996) 'Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations' Social Science Research, 25, 201-239.

International Labour Office (1990) International Standard Classification of Occupations, 1988. Geneva: International Labour Office.

United Nations (2002) List of international family of economic and social classifications
http://unstats.un.org/unsd/cr/family1.asp

4.9.3 The reliability of stratification measures

Anette Scherpenzeel; University of Amsterdam, Netherlands

Abstract. In the Swiss Household Panel survey, the module containing measures of work classification and social stratification is the largest module of the individual questionnaire. The longitudinal nature of the survey allows an estimation of change and stability in these measures over time. The distribution of the stratification classes seems very stable when we treat the data as if they were cross-sectional. However, finding the same number of people in each class in each panel wave, does not necessarily mean that the classes contain the same people. To distinguish real change from random variation, or unreliability, a Simplex model is specified, which shows there is very little true change over the four waves but the amount of random variation, or unreliability, in the measures is very high. It is shown that, even though this unreliability does not affect the distributions across the social classes, it does affect longitudinal models or longitudinal analysis. The measures of social stratification and of work characteristics contain so much random variation over time, that they loose all explicative power in a longitudinal causal model of satisfaction

4.9.4 Class Demography and Gender Inequality in Europe: Alternative Interpretations

Hakon Leiulfsrud; Norwegian University of Science and Technology (NTNU), Norway

Ivano Bison; Norwegian University of Science and Technology (NTNU), Norway

Abstract. Interpretations of ongoing changes of social structure and class formations are highly dependent on data available and class models used. Different theoretical rationales may result in alternative interpretations of class and inequality structures. Instead of restricting ourselves to one model we introduce and evaluate three theoretical class models developed by Erikson, Goldthorpe and Portocarero (EGP-classes), Wright, and Esping-Andersen.

In a first step we present comparative data on the European social structure based on the above classifications. In a second step we look at how these models enable or hinder us from giving a realistic picture of men and women's positions in the labor market between European countries. In a third step we discuss what is gained and lost using alternative class approaches. The data is based on European Social Survey 2002-3 and includes in our case 21 European countries.

4.9.5 Specificity and Universality in Occupation-based Social Classifications

Paul Lambert; University of Stirling, United Kingdom

Ken Prandy; Cardiff University, United Kingdom

Max Bergman; University of Basel, Switzerland

Abstract. The Cambridge Social Interaction and Stratification Scales (CAMSIS) occupation-based social classifications differ from most alternatives in the field on the basis of their specific character. They are derived separately for every country, time period and gender to which they are applied. Scores for occupational unit groups are calculated on the basis of empirical patterns of social interaction between the holders of different occupational units. As such, CAMSIS measures indicate the relative position of social stratification advantage or disadvantage typically associated with a particular occupation within the given society.

Thus, CAMSIS scales allow, first, that equivalent occupations may well be assigned different relative positions in different countries, time periods or gender groups. They are also amenable to the fact that occupational categories may differ in their coverage or definition (i.e., who precisely gets coded to them) between different structures. This contrasts with most other occupation-based social classifications, which are 'universalist', in that all occupations are intended to have exactly the same relative locations regardless of context. Moreover, for universalist occupation-based schemes, variations in the membership of occupational groups between societies (particularly between countries) are often problematic, seldom explicitly documented, and frequently the source of inconsistencies and uncertainties in subsequent analyses.

This paper reviews some of the strengths and weaknesses of the CAMSIS approach to specificity in occupation-based social classifications. The CAMSIS strategy can be considered as both theoretically and empirically more appealing. It can be evaluated favourably in terms of 'criterion' and 'construct' validity, exhibiting strong associations with other related individual level measures. However the CAMSIS approach encounters some pragmatic problems in terms of its practical operationalisations. It also raises questions on the theoretical limits of specificity, namely the issue of how fine a detail of specific derivation we should pursue. We present contemporary evidence on these issues from a number of national contexts (using in particular the ESS and LIS databases), and argue resolutions to some of the apparent shortcomings.

4.9.6 The Development of a new European Socio-economic Classification

David Rose; University of Essex, United Kingdom

Eric Harrison; University of Essex, United Kingdom

Abstract. This paper will explain and describe a proposed new European SEC currently being developed through a grant from the EU 6th Framework Programme. ESeC is based on the EGP class schema but has some detailed differences. The conceptual basis of ESeC will be discussed along with issues relating to its measurement and operationalisation using other European harmonised variables.

4.9.7 Classification systems and contextual difference. Some comments on occupational hierarchies from a South European angle

Thomas Maloutas; University of Thessaly, Greece

Abstract. The paper will focus on the importance of contextual difference for the performance of systems of social classification, related in particular to a number of specificities of South European labour markets and welfare systems/regimes. The three following issues will be mainly discussed:

1. The reduced pertinence of the managerial / supervisory
hierarchy as a classifying element within contexts where
large corporations and bureaucratic hierarchies are not predominant.
2. The importance of the social mobility dynamic in the
determination / consolidation of class positions within the socially
fluid context of increased mobility (related to the comparatively
recent and massive transition from agrarian to service economies,
from manual to non-manual labour etc).
3. The particular importance of citizenship within contexts of recent,
abrupt and mainly illegal and 'uninvited' immigration in reshaping
occupational hierarchies.

Relevant data from ESS Round 1 will be used to substantiate the general claim that contextual difference may confer a variety of social statuses within the same occupational positions, and that these statuses may not always be pertinently disentangled by correlation to the variation in employment contracts and/or to employment statuses while they are more closely related to other variables. The need for a common European social classification system being undeniable for a variety of reasons, the paper will conclude on the importance of this system to avoid blindness towards contextual variety.

Depending on the length of presentations, brief reference may be made to an application for the full exploitation of census data, currently developed by the National Centre for Social Research and the National Statistical Service of Greece, and the ways this application permits to retrieve and elaborate on data about occupational categories

4.9.8 Measurement errors and consequences in student's proxy reports about the socio-economic status of their parents

Frauke Kreuter; University of Maryland, United States

Kai Maaz; Max Planck Institute for Human Development, Germany

Rainer Watermann; Georg-August-Universität, Göttingen, Germany

Abstract. The measurement of socio-economic status is an important background variable for school achievement research. Questions on parent's education and occupation are often times part of children's questionnaires and therefore measured via proxy-reports (e.g. studies such as PISA, TIMSS). The quality of such proxy-reports of parental socio-economic status was discussed in the early years of school achievement studies. Past research focused only on answer comparison between student and parent's responses but did not evaluate any correlation with the dependent variable such as school achievement. Systematic measurement errors in proxy-reports that are correlated with school achievement will lead to an underestimation of SES-effects on school achievement. Variation in the quality of proxy-reports across nations will disturb international comparisons of SES-effects on school achievement. In this paper cognitive theories about the response process are used to derive hypothesis about systematic errors in proxy-reports and their relationship to the response variable school achievement. Data from PISA 2000 (OECD Programme For International Student Assessment) are used to test the hypothesize.

4.10 Other

Session chair:
Rolf Steyer; University of Jena, Germany

This session takes place on:
Friday, July 22; 15:00 to 17:00

Presentations:

- Invited presentation: *A new approach to the analysis of conditional and average mean differences between groups* - Rolf Steyer (Germany)

Contributed presentations:

- *Classic and Symbolic Methods for Clustering with an Empirical Exercise from the European Social Survey* - Seppo Laaksonen (Finland)

- *Methodological Argumentation Of Expert Judgements In Sociological Research* - Irina Kurdina (Ukraine)

- *Concept For Solving The Missing Data Problem In Social Surveys* - Ekaterina Markova (Bulgaria)

- *Social-cultural effects in delinquent and non delinquent symbolic decisions: an exploratory multidimensional analysis to define a social classification* - Helena Carvalho, Eduardo Ferreira (Portugal)

4.10.1 A new approach to the analysis of conditional and average mean differences between groups

Rolf Steyer; University of Jena, Germany

Abstract. Can there be anything new in the analysis of mean differences after almost a whole century of analysis of variance, regression models and comparison of means? The intriguing answer is "yes". The first example in case is non- orthogonal analysis of variance. I will show that none of the programs available in the standard statistical packages such as SAS, SPSS, Systat, etc. provides a correct analysis of the main or average effect of a treatment variable even though all these programs offer three or more ways of computing the sums of squares (type I, II, III, and IV), at least if "main"or "average effect"is defined in a way compatible with the Neyman-Rubin approach to causality. A second example is analysis of covariance. I will show that Ancova does not, in general, yield correct results, if there is an interaction between the covariate and the treatment variable. Although this finding will not be a surprise to many, because, the model assumes "no interaction", the question remains "How to analyse such a model?"A linear regression analysis with interaction terms also does not give the correct results either, if the covariate is stochastic, which is the standard situation in observational studies. This is my third example. And what about the fourth example, in which the covariate is latent and there is an interaction between the covariate and the treatment? Just take the manifest proxis? Wrong again! The new approach presented provides correct solutions for the analysis of conditional and average mean differences in all these examples. It uses standard programs for structural equation modelling that can handle nonlinear constraints.

4.10.2 Classic and Symbolic Methods for Clustering with an Empirical Exercise from the European Social Survey

Seppo Laaksonen; Statistics Finland, Finland

Abstract. There are a number of clustering methods based on micro data, but this paper does not try to explain these. At contrast, we have here macro data and we wish to cluster these macro units or aggregates and thus create different clusters so that these would as homogeneous as possible. But this is not as simple as looks, since the results depend on a macro data structure. Classic analysis is usually based on 'point values' so that in the case of aggregates the initial variable values are to be point estimates such as averages, medians and maybe other quantiles. But there exists a newer methods family, called symbolic data analysis, that allows for variable values to be more complex such as intervals and frequencies. This paper compares some classic clustering techniques with symbolic ones. The micro data are from the European Social Survey (ESS), covering those 20 countries in which the Schwartz life value questions were asked using the supplementary questionnaire of round I (2002/2003). The aggregates here are just these 20 countries and the clusterings are based on the four variables, that are coming from the four factors (naturally aggregates could be more complex). Thus we have first calculated factor scores of each ESS respondent and then aggregated these at country level usign different strategies, and then started the clustering. The paper is mainly methodological but we present the key results from each strategy as well.

4.10.3 Methodological Argumentation Of Expert Judgements In Sociological Research

Irina Kurdina; National Taras Shevchenko University of Kiev, Ukraine

Abstract. Currently, expert judgment methods are indispensable component of any social research. Their popularity and widespread application have caused the need for thorough scientific investigation aimed at methodological reasoning and elaboration of optimal procedure for intuitive analysis combined with quantitative methods of evaluation and processing. Such need is endorsed by the fact that the overwhelming majority of sociological research applying expert judgments is deprived of theoretic methodological basis. Instead, this research is conducted in spontaneous and disorganized manner and subsequently proves ineffective. Sociological scientific literature discusses methodological aspects of expert judgment application skin-deep; some of these aspects are left behind the methodology researchers' attention at all. To the author's mind, there is enough evidence to consider expert judgments as peculiar approach to social reality analysis rather than merely instrument for information acquiring. Hence, such problems as distinctive features of expert judgment object and experts status and criteria for their selection with regard to specificity of subject of inquiry, are of current importance. Also, there is a need to further develop such issues as grading of application spheres for this method in sociology; capabilities and restraints; scientific status, validity and reliability of survey conclusions. Current research is focused on methodological argumentation of expert judgments in sociological research.

4.10.4   Concept For Solving The Missing Data Problem In Social Surveys

Ekaterina Markova; Bulgarian Academy of Sciences, Bulgaria

Abstract. The paper presents an author's concept for solving the missing data problem in social surveys. Its main goal is systematization of the criteria for choosing approaches and methods, and the possible strategies, which could be undertaken in various research situations. This analysis is made according to the Bulgarian research experience.

The activities, directed toward decreasing the uncoverage (missing data), are generally presented as follows: 1) preventive –at the stage of survey preparation; 2) current–at the time of fieldwork, and 3) subsequent, with which the data, obtained from the sample, is treated. The effectiveness of the procedures before and at the time of the data collection is, to the highest extent, under researchers' control. The subsequent activities are used when "no other choice is available", i.e. at the presence of missing data.

We developed criteria for choosing the optimization approach (method), and as a methodological background we accept the conclusion that there is not and could not be only one method, which could be placed in a privileged position. However, what has been said up till now we cannot conclude that the methods for sample optimization are "equal in rights". On the contrary, they all have different, relative advantages and limitations.

The criteria for choosing approaches and optimization methods could be classified at least in three groups: 1) accuracy criterion; 2) financial and other expenses; and 3) necessary time for obtaining the final results. In the paper we make critical analysis of the Bulgaria's most frequently used (most preferable) methods for sample optimization. Special attention is paid to the factors, determining the choice of subsequent optimization approach depending on: 1) percentage of cases, for which data is missing; 2) actuality of the information about the observed characteristics of the population; and 3) availability of correlation between weighting indicators and the part of the sample, for which the data is obtained.

Special place in designing of a strategy against missing data problems is devoted to the issue of costs –financial expenses, qualified staff, time, as well as (in particular) the requirements for timely submission of the results, organizational problems, etc.

By defining of three main (principal) strategies the base is given of complete concept for solving the missing data problem. With these strategies we try to find out what is "moving threshold" in missing data, beyond which the problem should bother the researchers.

The "movable threshold" is defined by the following main factors: 1) decreasing the sample size and, as a consequence, increasing variable (sampling) error; 2) violation of representativeness of estimates, related to the appearance of additional component, i.e. bias; 3) value of the bias. Additional reason to call this admissible missing data threshold "moving" is the various research tasks.

We present examples from the Bulgarian social survey practice illustrating the three strategies for solving the missing data problem.

Probably in future we should expect this problem to aggravate, because of two essentially important tendencies: first, the field "fatigue" and respondents' lowering interest in something very well-known (answering in surveys); second, the world perspective to collect more and more personal data by Internet

4.10.5   Social-cultural effects in delinquent and non delinquent symbolic decisions: an exploratory multidimensional analysis to define a social classification

Helena Carvalho; Instituto Superior de Ciências do Trabalho e da Empresa - ISCTE, Portugal

Eduardo Ferreira; Instituto Superior de Polícia e Ciências Criminais, Portugal

Abstract. The persistent difficulties of classical sociological theories to explain why social, economic or cultural tensions do not always induce delinquent behaviours led, among other consequences, to the emergence of the rational choice and the routine activities approaches. The theoretical and empirical attractiveness of these approaches covered up, for some time, that delinquent or non-delinquent decisions can be affected by subjective variables as social shame, fear of significant others reactions, or even "guilt". On the other end, these subjective variables can be affected by more objective and structural variables, such as gender, age or social origin. This paper describes the results of an exploratory research that tries to define social-cultural typologies through a methodological model that stresses their symbolic space. In a first phase this model gives a description of different symbolic social-cultural configurations, which is supported by a multidimensional structure of categorical variables, using

Multiple Correspondence Analysis. In a second phase the model gives an effective definition of different types of what we called delinquent or non delinquent symbolic cultures. In order to formalise a classification system, the model uses an articulation between two methods of multivariate analysis –Multiple Correspondence Analysis and Cluster Analysis.

Index of Authors