

**SEVENTH FRAMEWORK PROGRAMME
THE PEOPLE PROGRAMME
MARIE CURIE ACTIONS – NETWORKS FOR INITIAL TRAINING (ITN)**

**ELECDEM
TRAINING NETWORK IN ELECTORAL DEMOCRACY
GRANT AGREEMENT NUMBER: 238607**

**Deliverable D4.1 – Measuring Citizen Preferences in an
Globalised World Final Report**

Experienced Research fellow (ER): Steve Schwarzer

Host Institution: TNS opinion

1 Abstract

This report discusses some of the methodological challenges that scholars face in cross-national survey research. Multi-country research, as a complex phenomenon, adds some additional layers to public opinion and electoral research projects.

To ensure that the cross-national variations observed result from true variation between countries and their different institutional arrangements, researchers have to ensure the measurement quality of their instruments by controlling for possible measurement errors. Measurement error in comparative cross-national/cross-cultural research may occur due to a variety of problems regarding design and sampling, the questionnaire design and the operationalisation of theoretical concepts, modes of questionnaire translation and survey administration.

Most statistical techniques that have been developed to test for the quality of measurement (reliability, validity, equivalence) in survey-based research are usually applied *a posteriori*.

To a great extent, cross-cultural comparability can be secured by targeted international cooperation in the developmental phase of the survey (multiple translation verification, extensive pre-testing phase, following strict rules for survey administration in all national (cultural) contexts etc.).

However, cross-national (cross-cultural) equivalence may only be achieved when the indicators applied also reflect the systemic level of the countries concerned. The legal environments and political traditions (in terms of democratic practices, institutional variations, and socio-demographic differences) may vary significantly between countries.

Thus, by discussing three different examples with a multi-country or at least international dimension, the report will highlight the theoretical, analytical and cultural complexity of conducting empirical research projects. The underlying motive is of course the fact that research needs to ensure comparability and equivalence in cross-national and cross-cultural survey research.

Table of contents

1	Abstract.....	2
2	Executive Summary.....	4
3	Objectives and Aims – Workpackage 4.....	6
3.1	Conceptual linkage: Evaluating Electoral Democracy.....	7
4	Cross-national electoral research in practice	9
5	Example 1 - Cross-national electoral research and equivalence	12
5.1	Theoretical basis and assumptions	12
5.1.1	Context and individual-specific causes of over-reporting	13
5.2	Experiment	15
5.3	Results.....	16
5.4	Discussion of the results in the light of this report	19
6	Example 2 - Opinion polls in former authoritarian regimes	21
6.1	Theoretical and empirical basis – Armenia, an emerging democracy	21
6.2	Conducting an opinion poll in Armenia.....	24
6.3	Discussion in the light of the report	29
7	Example 3 - Tackling national variations in the analytical phase.....	30
7.1	Theoretical basis of this example	31
7.2	Modelling	33
7.3	Results.....	34
7.4	Discussion of the results in the light of this report	40
8	Conclusion.....	42
9	References	47

2 Executive Summary

This report aims at highlighting the very different dimensions which evolve in cross-cultural research. The special difficulties of multi-country studies include challenges to measurement equivalence and the use of multiple languages, as well as conceptual aspects and the basic question of obtaining good sample frames.

Every element of a survey research project has the potential to influence the results. Difficulties that are often emphasized include challenges to conceptual and indicator issues, obtaining good sample frames and entirely practical problems found in fieldwork organization and data collection.

Cross-national electoral studies such as the European Election Study are multi-country projects, conducted in several countries. A survey implemented in several different settings needs a comparative and reliable instrument in order to address a wide range of features: conceptual coverage, level and choice of vocabulary used, adaptation of wording and other technical features.

As survey questions are measurement tools and language vehicles used by researchers to formulate enquires about indicators chosen to measure specific constructs in order to gain insights into theoretical or empirical concepts, care is required in preparing the questionnaire for its use in different countries.

Different standards and *modi operandi* are employed to deal with these kinds of challenges and research institutes specialized in multi-country research have developed their own approaches to conducting research across multiple countries.

The following report uses three examples to address major issues in questionnaire-based research projects.

The first example addresses the meaning of a specific and widely used concept in electoral research, "voter turnout". The project was carried out as a survey experiment in an online setting and mainly focuses on question wording. Nevertheless, the results indicate that questions work differently in different settings, which directly leads us to the question of measurement equivalence.

The second example discusses the transfer of concepts widely used in western European electoral research to a country in transition, Armenia. In order to prepare pre-election surveys in Armenia, a questionnaire was developed and pre-tested that allowed successful implementation of the project. Nevertheless, important lessons can be drawn from this project about the importance of the cultural environment in multi-country projects.

The third example does not focus on the preparation or implementation phase, but focuses on the analysis of data from a multi-country setting by using multi-level analysis to differentiate between individual and country-related variance. The research question in this third example sought to establish which factors affect the political engagement of young people in different countries. For this example the data of the International Civic and Citizenship education study were used.

All three examples fall into a frame that deals with the question on how to organize cross-cultural research projects. The report highlights some pitfalls that might not be seen when analysing the data. But on the other hand, the analysis of data is heavily affected by these aspects.

Analyzing data from a multi-country research project also means to think about the conditions under which data are collected and whether one can assume that measurements are equivalent across countries. This is not only a matter of procedures and standardized processes, but foremost related to cultural settings. As soon as the question tackles a problem that relates questions asked in a questionnaire and cultural contexts, we might introduce bias into our surveys and also in the analysis.

At this stage, survey research and comparative research are somehow interlinked and need to reflect on the specific requirements that come along with both methodological streams of discussions.

Of course, under the condition that researchers established measurement equivalence in the data from different countries, researchers can make use of all the advanced statistical techniques available. But again, one has to bear in mind that these techniques are able to tackle specific structures within our data, e.g. hierarchical data structures, but on the other hand only work under certain assumptions.

Nevertheless, there are solutions to overcome this problem. On the analytical side it seems necessary to focus more on analysing equivalence of measures rather than only focussing on predictive theoretical models. With respect to the implementation of surveys, cross-cultural survey research should try to implement survey in the manner of a true comparative survey design to achieve at least functional equivalence.

3 Objectives and Aims – Workpackage 4

"It can help the researcher go into geographical and cultural settings with which he or she is relatively unfamiliar without committing egregious errors of interpretation" (Peters 1998: 101).

European diversity - reflected in a variety of cultures, languages, institutional arrangements and political opportunity structures - entails incredible richness, but also poses particular challenges for comparative social research.

Besides technical problems, cross-national survey research faces theoretical and methodological challenges which also have to be addressed by the research teams. The methodology within the quantitative paradigm of research rests on a general assumption about secure knowledge which defines the role of theory in the research process. Basically, empirical researchers test theoretically deduced hypotheses.

Different kinds of measurement instruments are applied to gather data and to test theoretical assumptions about social reality. This leads to the fundamental dilemma that we do not know whether our theoretical assumptions are false or incorrect or whether the indicators that we use to measure theoretical concepts are inappropriate.

However, statistical methods cannot handle the necessity for control within a study design. Usually, tests for equivalence, validity and reliability are applied after the collection of data. The problem arises, because we assume a relatively secure knowledge over the world and this lead us to assume stable relationships between different social phenomena and different levels of analysis.

Whereas experimental designs can be seen as a research design allowing for control, and therefore able to prove causal linkages between dependent and independent variables, quasi-experimental and non-experimental research designs need to apply statistical analysis to *a posteriori* control for third variable effects.

This problem is even more complicated if the research project is cross-nationally or cross-culturally designed, because the assumed relationships are going to be tested in different cultural, societal and structural settings.

An experimental design can be seen as the best approximation to the ideal of control and statistical techniques allowing for approximation to experimental methods of control: "it entails the conceptual (mathematical) manipulation of empirically observed data – which cannot be manipulated situationally as in experimental design

- in order to discover controlled relationships among variables. It handles the problem of control by means of partial correlations" (Lijphart 1971: 684).

Yet statistical control can only be as good as the quality of the measures applied to test the hypotheses. Thus, quality of measurement is crucial for the predictive power of explanatory models.

3.1 Conceptual linkage: Evaluating Electoral Democracy

The ELECDEM network aimed at analysing the effects of globalization, communication and institutions on the quality of electoral democracy.

The work in workpackage 4 ("Measuring citizens preferences in a globalized world") was entirely focused on multi-country research projects, the techniques for the design, implementation and analysis of data from multi-country projects. Consequently this report presents three examples of the different aspects which were part of the experienced researcher's work during his scholarship and afterwards.

The three examples presented in this report touch on several aspects of the ELECDEM objectives.

First of all, electoral research has a cross-national focus and needs to link and reflect experiences of different countries. In order to do this, researchers in the field need to understand the technical aspects of survey research in order to ensure the comparability of data sets to be analysed.

Comparability in survey research is a multi-layered concept, comprising very different things, such as survey design, questionnaire design, case selection, fieldwork implementation, mode effects, and analytic techniques, just to name some.

Comparability is the basis for the evaluation of the quality of electoral democracy. The workpackage and this report has a slight technical bias. Nevertheless it has relevance for the study and analysis of electoral systems.

As mentioned above, and also in the ELECDEM technical proposal, most scholars base their projects on cross-national data sets. Therefore it also seems important to reflect on data quality.

In the course of the scholarship related to workpackage 4 the experienced researcher applied several techniques which were also part of the training programme offered in ELECDEM. The report mainly focuses on electoral survey design, comparative methods and cross-national survey research, as well as multi-level approaches and experimentation. In other words, most of the key methodologies of ELECDEM have

been used to prepare this report and also on several other occasions since the end of the scholarship period.

Based on these experiences, it was very helpful for early stage and experienced researchers to have the chance to participate in the ELECDEM training programme. It was also helpful to participate in several other training programmes outside ELECDEM.

In addition to the courses offered by ELECDEM I had the opportunity to participate in courses on structural equation modelling and sampling design. Both aspects can be placed in the workpackage 4. Structural equation modelling in particular is an interesting technique for the evaluation of the cross-cultural equivalence of measurements. Furthermore, it can be used to analyse latent structures in data, which is quite a useful addition to the dominant approaches of multi-level regression models.

Good sampling and random sampling techniques form the basis for every survey research project.

4 Cross-national electoral research in practice

Theoretical approaches and concepts are central for the social sciences since they address social phenomena through generalisation and categorisation of social reality. Yet concepts may differ in the variety, number and dimensions they encompass from country to country, from one cultural context to another. Countries' distinct historical developments and variability in current political, societal and economic structures may lead to divergent interpretations of theoretical concepts when applied within cross-national or cross-cultural research.

"Questions should have a broadly equivalent meaning to all respondents to ensure that variations in the data derive from differences in their answers rather than in their interpretation of the questions." (Jowell 2007: 6)

Thus, theoretical models may be falsified for one country and verified for another. It is therefore important to be aware that middle range theories are bounded not only in time but also in space. Cultural, social and political conditions shape the settings in which people live and influence individuals' behaviour, attitudes and values. The design and implementation of cross-national research pose specific problems that go beyond the methodological effort that we encounter when conducting mono-cultural or within-country research. Heterogeneity within and between countries seems to challenge not only the analytical framework used to explain social phenomena but also the operationalisation of theoretical concepts.

"[...] we should be extremely cautious about treating respondents from different states as representing homogeneous 'values' or 'cultures' that are assumed to typify these states" (MacInnes 2005, p. 113).

Most of the problems related to heterogeneity originate from the process of gathering empirical data and could therefore be detected by studying the design or the sampling strategy of a specific survey. Bearing in mind the significant effect of heterogeneity on the predictive power of scientific explanations of social reality, and the necessity to secure comparability, standards have been established for the evaluation of cross-national surveys which involve inter alia: a) consistency of concepts, b) transparency in operationalisation, c) documentation of pre-testing and field work, d) sampling procedures, e) sources of error and bias, and f) modes of questionnaire translation.

Technically speaking, heterogeneity within and between countries may lead to non-equivalence of concepts and measures. Non-equivalence and thus non-comparability have been recognised as major problems in cross-national/cross-cultural survey research (Harkness 2007, Van de Vijver 1998, 2003).

Different criteria are therefore necessary which help to ensure that different voices in different populations are appropriately represented and taken equally into consideration when designing a comparative cross-national study.

Lack of equivalence in concepts and measures can undermine the central objective of comparative research. On the one hand, the research design already calls for control when setting up the study design itself. "Encounters with diversity on comparing countries heightened awareness of the problem of equivalence across systems. In order to compare something across systems it is necessary to have confidence that the components and their properties that are being compared are the 'same' or indicate something equivalent" (Teune 1990: 53 - 4)

The most appropriate way to secure control on the relationship between dependent and independent variables can be reached in experimental designs. In quasi- and non-experimental designs we rely on statistical control (applying statistical techniques that allow controlling for third variable effects). However, problems of incomparability that originate from inappropriate study design or inappropriate operationalisation of concepts may be detected through statistical control but not corrected for.

What other possibilities do we have for avoiding or controlling for the methodological problems that cross-national survey research faces?

This report discusses the problem of multi-country research from three different angles: the operationalisation of theoretical concepts, the implementation of surveys in different settings and the analysis of data coming from multi-country research projects.

This is of course just a snapshot of the different layers of multi-country projects and does not cover all the aspects that might occur in this kind of work. But the report focuses on aspects that are of importance for current research and sometimes even goes beyond current discussions.

The three examples are taken from the work of the former ELECDEM Experienced Researcher placed at TNS opinion. The projects focus on different aspect of empirical research, but at the same time cover the whole range of possibilities for getting hold of or controlling country-specific aspects.

The report aims at highlighting the very different dimensions which evolve in cross-cultural research. Each of the projects introduced above addresses a given methodological aspect in a multi-country research project, but they also have different implications for the preparatory stages and analytical phase of survey research.

Consequently, the report is structured in eight sections. It starts with a discussion of some fundamental principles of comparative survey research. In particular, differentiation between the unit of analysis and the unit of observation seems to be central. Since the selection of cases (countries) is closely related to both unit of analysis and unit of observation, and since it may be one source of non-equivalence of concepts and measures, it is of enormous importance for small and medium-scale research projects to take this into account.

5 Example 1 - Cross-national electoral research and equivalence¹

Information on voter turnout is one of the central units of analysis when studying electoral behaviour in liberal democracies. Most research on turnout is based on data stemming from survey research. However, the problem with survey data is that they might not reflect the actual behaviour of the respondents: respondents 'over-report' on turnout, meaning that they report turnout but did actually not vote. Thus, researchers are confronted with the phenomenon that the proportion of respondents who say that they voted is higher than the actual turnout in the election (e.g. Traugott & Katosh, 1979). Taking this further, analyses based on this data might subsequently produce biased results and conclusions.

There is broad evidence that these too-high proportions of reported turnout result from different sources (see Holbrook & Krosnick, 2011 for an overview) ranging from address sampling or coverage errors (Granberg & Holmberg, 1991; Traugott & Katosh, 1979) to measurement error ('over-reporting'). However, a literature review of the topic indicates that no comparative study which attempts to analyse whether new ways of wording the questions and response options work equally well in reducing over-reporting in different countries has been conducted.

In other words, most studies are limited to evaluating different forms of asking a question in one country. But, if attempts to reduce over-reporting are context-sensitive, it is important for comparative research projects to know what differences have to be taken into account and which questions therefore should be used.

5.1 *Theoretical basis and assumptions*

Attempts to reduce intentional over-reporting have to take the cognitive dissonance of respondents into account. If cognitive dissonance can be reduced, the respondent is more likely to admit to not voting. As pointed out, several attempts to reduce over-reporting have been initiated. Most prominently, election studies include an introductory sentence that is meant to reduce the social stigma of not voting. For instance, the European Social Survey (round 5) uses the statement "Some people don't vote nowadays for one reason or another", while the American National Election Study (ANES) uses "In talking to people about elections, we often find that a lot of people were not able to vote because they were not registered, they were sick, or they just didn't have the time". In both cases, the introductory sentence is followed by a simple and direct question on whether the respondent turned out or not ("yes" versus "no"). Although there are findings that suggest that these attempts do not reduce over-reporting (Abelson et al., 1992), they are nevertheless very widely used.

¹ The work for this example was carried out in collaboration with the Department of Methods in the Social Sciences, University of Vienna – Prof. Sylvia Kritzinger and Dr. Eva Zeglovits

Belli et al. (1999, 2006) also addressed the problem of over-reporting and tried to capture both the problem of memory failures and intentional misreporting. To circumvent the first problem, they included a long explanatory statement to assist respondents in remembering the election of interest (Belli et al., 1999)². For the second problem, they presented more options for reporting non-voting by including three different ways of doing so (Belli et al., 2006; Belli et al., 1999). In addition to simply reporting that they did not vote, respondents were also given the possibility to answer (1) that they usually vote, but this time did not, and (2) that they thought about voting, but then did not vote. Instead of only one non-voting option, three were presented to the respondents. The additional two measures reduced reported turnout substantially compared to the standard question in the ANES.

The idea of diversifying response options to make it easier for respondents to admit non-voting was taken up in a single country study (Austria). Zeglovits and Kritzinger (2011) tested a new form of response options. They adapted the idea of the propensity to turn out - used in pre-election studies - to post-election scenarios and self-reported turnout. Respondents received four possible response options, ranging from being sure that they did not vote to being sure that they did vote, with two increments in between, where they could say that they were not sure, but presumably did or presumably did not vote. This approach was developed firstly to capture valid turnout responses in surveys which are conducted not immediately after an election, and secondly for use in pre-election surveys asking about turnout in the last election. Again, it successfully reduced reported turnout compared to the standard question, with results approaching the official turnout rate.

So far, we therefore have some evidence that changes in question wording and wider response options help to reduce over-reporting in survey questions³. However, what we do not yet know is whether these results hold across different contexts. This paper therefore explores whether they can be applied to different settings, and if so, how.

5.1.1 Context and individual-specific causes of over-reporting

Analysis of the explanations of over-reporting has identified a number of individual level characteristics that are positively related to a respondent's tendency to over-report turnout.⁴

² In the beginning of their research, they used an even longer version of the introductory sentence (Belli et al., 1999), but reduced it to a shorter version in the course of the research (Belli, Moore, & VanHoewyk 2006).

³ Other recent attempts avoid the direct turnout question but take an indirect approach: the Item Count Techniques (Holbrook & Krosnick, 2011b; Zeglovits & Kritzinger, 2011) and the Randomized Response Techniques (Holbrook & Krosnick, 2011a). These indirect approaches have been tested with mixed success. Therefore, we will not test these techniques in a cross-country design.

⁴ In general, the characteristics that usually affect turnout also affect over-reporting.

Why do people over-report turnout intentionally? Here, social desirability is regarded as the main driver. Non-voters would like to reduce the cognitive dissonance caused by their behaviour: they are either anxious to please the interviewers or themselves (Bernstein et al., 2001).

Education is known to increase over-reporting, as do political engagement, political interest, frequency of following the news, and civic duty (Bernstein et al., 2001; Cassel, 2003; Granberg & Holmberg, 1991; Hill & Hurley, 1984; Karp & Brockington, 2005; Presser & Traugott, 1992; Silver et al., 1986). Results for age are puzzling, however. Some scholars observe that young non-voters are more likely to acknowledge abstention than older non-voters. This means that over-reporting increases with age (e.g. Granberg & Holmberg, 1991). However, other scholars report the reverse, namely that young voters are more likely to over-report than older voters (e.g. Hill & Hurley, 1984). In contrast, Zeglovits and Kritzinger (2011) did not find any age effect at all of question wordings reducing over-reporting. Finally, Silver et al. (1986) describe age as having a curvilinear effect on over-reporting, similar to the curvilinear effect of age on turnout. As a general conclusion, correlations between these characteristics and turnout might be inflated.

Unlike research at the individual level, there is little or no research into the contextual level. With the exception of Karp and Brockington's study (2005) no other study describes and analyzes cross-national differences in over-reporting in surveys. Moreover, to our knowledge no study deals with cross-national differences in question wording and diversifications in response options, and their effects on over-reporting. Due to this gap in research, we also lack knowledge of theoretical frameworks: more precisely, we cannot deduce any particular assumption on how contextual circumstances might influence alternative question formats. For instance, alternative question formats might be more or less successful in reducing over-reporting in circumstances where over-reporting is high. Either impact could be possible: on the one hand, if the probability of over-reporting is high due to certain contextual circumstances⁵, successfully tested attempts to reduce over-reporting will work better than the standard question; on the other hand, it might also be that new question formats have no impact at all as social desirability-led behaviour is enhanced by particular contextual factors. Thus, cognitive dissonances might be dissolved differently under different contextual circumstances.

This gap in the literature leads us to take an exploratory approach: we will simply explore whether new question formats produce a) positive differences at all, and b) whether these differences are equal across countries. From there we will draw descriptive inferences on contextual factors. As contextual factors we consider overall

⁵ Saliency of the election or high overall turnout (see Karp & Brockington, 2005).

turnout, election saliency (e.g. closeness of the race) and the time passed since the last election. Different question formats might perform differently under these contextual circumstances. For instance, in contexts where elections took place a long time ago, respondents might be able to dissolve their cognitive dissonance better if the alternative question format reflects likelihood. In the context where elections took place only recently, this might not be the case.

5.2 Experiment

The multi-country experiment was conducted in an online survey to compare three different question formats with varying response options. These different formats are referred to as treatments. The participants were randomized into one of the three treatment groups⁶.

Treatment A used the standard question format used in most election studies, and thus formed our reference group. Here, the question wording was already set up in such a way as to make it easier for respondents to report non-voting, in the usual way.⁷

Treatment B, based on the approach developed by Belli et al. (1999, 2006), was translated and adapted with a response scale including four possible answers. Although minor changes were necessary in the question wording⁸, the translation keeps the response options identical⁹.

Treatment C is the question format developed by Zeglovits and Kritzinger (2011). It combines the stimulus of Treatment A with four response options pointing to the likelihood of past behaviour¹⁰.

The survey experiment was conducted as a separate part in a survey on Internet service providers, which was coordinated by TNS opinion.¹¹ The experiment was

⁶ However, respondent were offered an escape option for each question: for all three treatment groups the "don't know" option was displayed when respondents tried to skip the question.

⁷ Treatment A: "The following question refers to the previous [federal election] in [month year]. In this [federal election] a lot of people could not vote or chose not to vote for some reasons. What about you? Did you vote or not?"

⁸ E.g. with regard to voters' registration mentioned in the original stimulus which does not apply to our setting in Europe.

⁹ The following question refers to the previous [federal election] in [month year]. In talking to people about elections, we often find that a lot of people were not able to vote because they were sick, did not have the time, or just were not interested. Which of the following statements best describes you?

I did not vote in the [federal election] in [month year]. I thought about voting this time but didn't. I usually vote but didn't this time I am sure I voted in the [federal election] in [month year]."

¹⁰ Treatment C: "The following question refers to the previous [federal election] in [month year]. In this election, a lot of people could not vote or chose not to vote for some reasons. Also, some time has passed since. Which of the following statements best describes you? I am sure I did not vote in the [federal election] in [month year]. I am not sure if I voted but I presumably did not. I am not sure if I voted but I presumably did. I am sure that I voted in the [federal election] in [month year]."

¹¹ TNS opinion is a Brussels-based research firm specializing in conducting multi-country studies in all parts of the world, working with and for clients such as the European Commission, the World Bank, the European Central Bank, and the

implemented in seven European democracies: Austria, Belgium, Denmark, France, Germany, the UK and Ireland.

Given the theoretical discussion of multi-country research and case selection, it must be observed that the selection of countries proved to fit the contextual needs, as substantial variation in some of the contextual factors mentioned above can be found in these countries: The sample included countries with variation on overall turnout rates. For instance, Belgium has very high levels of turnout, whereas in the UK turnout is rather low. Furthermore, there are also differences in the length of time since the last election (Saris & Gallhofer, 2007): Denmark is the country with the most recent election, while France represents a country where the election took place a long time ago.

On average 1,000 respondents per country completed the questionnaire using the technique of Computer Assisted Web Interviewing (CAWI). Data collection took place between 1 and 21 December 2011. For each country samples were built on multiple recruitment strategies, via telephone and online recruitment. Panelists were invited to participate based on soft quotas on gender, age and urbanization. The samples show an acceptable distribution in terms of gender, age and urbanization.

5.3 Results

First, the aim was to analyse whether the different question wordings do actually reduce reported turnout in the country in question. As the validation of turnout is hardly possible, the experiment relies on the comparison of reported levels of turnout for each treatment group. If reported levels of turnout decrease, one can assume that over-reporting is decreased.

The table below shows the official results and also the results for treatment A, B and C.

Table 1: Countries and results

	Election	Election day(s)	Official turnout	Treatment A	Treatment B	Treatment C
AUT	Nationalratswahl	28.09.2008	78.8	77	71	73
BE (Fr.)	Élections législatives fédérales belges de 2010	13.06.2010	87.7 ¹²	87	83	78
DK	Folketingsvalg	15.09.2011	87.2	89	83	78
FR	Élections législatives	22.04.2007 06.05.2007	83,77 (1st round) 83,97 (2nd round)	76	76	74

European Bank for Regional for Reconstruction and Development. TNS opinion also participates in the EU's framework programmes and Marie Curie Actions.

¹² Regions considered: Hainaut, Namur, Walloon Brabant, Liège, Luxembourg (Lussimbork).

GER	Bundestagswahl	27.09.2009	70.8	82	65	68
UK	British General Election	06.05.2010	65.1	79	73	77
IRE	Irish General elections	25.02.2011	70.1	76	68	70

First the answers in treatments B and C are compared to official turnout and the answers in the reference group treatment A to check if the new question formats lead to different levels of reported turnout from the standard question. This approach has been used repeatedly to evaluate new question formats (Belli et al., 2006; Holbrook & Krosnick, 2010).

The results indicate that the assumed over-reporting for treatment A is not always the case. For Austria, Belgium, Denmark and France treatment A does not result in over-reporting. But, except in case of France, the experiment demonstrates that the change in question format leads to a reduction of “self-defined” voters. Additionally, the reduction is not consistent in the different countries. In some, treatment B reduces “self-defined” voter numbers more; in other cases treatment C has more effect. However, treatment B seems to reduce over-reporting the most.

So, even if the different question formats work in the expected direction, the results also indicate that there is a random element observable, which might be related to country-specific conditions.

In order to test this, several logistic regression models were used to check for individual and context specific results¹³. Three models test whether the treatments work at all, and if so whether they work the same in all countries. The dependent variable is self-reported turnout, 1 indicating that a person reported that he or she had voted, 0 indicating self-reported non-voting (as in table above). The treatments B and C act as independent variables, with the standard question in treatment A forming the reference group (Model 1 – treatment only). Next, interaction terms were added between countries and treatments, meaning that we allow treatments to work differently in the countries (Model 2 – treatment in countries). Here, Ireland is the country of reference. To check whether the effects observed hold true the third model controls for individual characteristics that are known to affect turnout and over-reporting (Model 3 – with controls on the individual level).

¹³ A multilevel modeling cannot be used as only 7 countries have been included in this study. Random effects cannot successfully be modeled with that small n on the second level (Rabe-Hesketh & Skrondal, 2008, p.124). The far too simple logistic models allow taking account of the clustered structure of the data when calculating standards errors, with the option `vce(clustered)` in the STATA command.

Table 2: Logit models; Ireland as the reference group

Variable	treatment only	treatment in countries	with controls on the individual level
treatment B	-0.643***	-1.044***	-0.995***
treatment C	-0.416***	-0.569***	-0.498**
interaction B * AUT		0.344***	0.288***
interaction B * BEL		0.802***	0.733***
interaction B * DEN		1.029***	1.198***
interaction B * FRA		0.290***	0.035
interaction B * GER		0.142***	-0.02
interaction B * UK		0.512***	0.105*
interaction C * AUT		0.092***	0.122**
interaction C * BEL		0.349***	0.204***
interaction C * DEN		0.460***	0.628***
interaction C * FRA		0.119***	-0.143***
interaction C * GER		-0.094***	-0.317***
interaction C * UK		0.287***	-0.041
freq following news			0.175***
efficacy			0.246***
age			0.050*
age squared			0
female			0.091
education medium			0.302**
education high			0.525***
migrant			-0.841***
income			-0.054
currently working			0.269*
area big city			-0.117
area small city			-0.132*
_cons	1.421***	1.421***	-1.811***
N	6779	6779	6775
ll	-3.833.716	-3.802.520	-3.487.807

Note: * p<0.05; ** p<0.01; *** p<0.001

The results of the logistic regression models suggest that the alternative question wordings are successful in reducing respondents' tendency to report positive turnout. There are country-specific differences in how treatments B and C reduce reported turnout. In Ireland, treatment B works "best" in the sense that reported turnout is reduced most severely. Treatment C has the greatest effect in Germany with reported turnout being reduced even more than in Ireland. Running a linear test to

check if the treatment reduces self-reported turnout within each country¹⁴, we can observe that both treatments significantly reduce reported turnout in five of the seven countries, namely Austria, France, Germany, Ireland and the UK, where treatment effects plus country-specific correction terms are significantly lower than zero (both in models 2 and 3). In French-speaking Belgium the reduction is not significant.¹⁵ In Denmark no significant effect of the treatments remains when controlling for the individual characteristics. Thus, self-reported turnout could be reduced in five out of seven countries when introducing new question formats.

5.4 Discussion of the results in the light of this report

This experiment has shown that question wordings do indeed work differently in different contexts. Depending on contextual circumstances, question formats perform differently.

The experiment only addresses measurement errors. Here, we have to distinguish between unintentional misreporting (e.g. memory errors) (Belli, Moore & VanHoewyk, 2006; Belli, Traugott, Young & McGonagle, 1999; Stockè, 2007; Stockè & Stark, 2007) and intentional misreporting (Bernstein, Chadha & Montjoy, 2001; Silver, Anderson & Abramson, 1986; Stockè & Stark, 2007). In the latter case, respondents report one behaviour although they are conscious that they performed another. Thus, intentional over-reporting means that people say that they cast their ballot, but are aware that they abstained¹⁶.

So far we have not tackled the problem of language differences in comparative research. For future research, we need to raise the very general question of whether questions translated carefully from one language to another always capture the same phenomenon in the different countries.

The small-scale experiment shows considerable variations between countries in the effects of the different question wordings. It might be that one question form is simply more suitable in one country from a language perspective, whereas another question format captures turnout linguistically in a more accurate form in another country.

This, of course, has consequences for the use of cross-national data sets, as the analysis works on the assumption that the measurements are invariant across countries.

¹⁴ Thus, we tested if the sum of the coefficient of a treatment and the interaction term between the treatment and the variable indicating a specific country are significantly different from zero.

¹⁵ This might be due the small n for Belgium. The p-value for the sum of the effects for treatment C is smaller than 0.10, but from our perspective still not significant.

¹⁶ In contrast, under-reporting has been identified as a minor problem (e.g. Traugott & Katosh 1979; Abelson, Loftus & Greenwald 1992; Belli, Traugott, Young, & McGonagle, 1999).

The experiment above demonstrates that even a rather simple factual question might produce different results across countries.

Furthermore, the particular context seems to be important for cross-national research. Using a question like the one above and focusing on elections that took place a year ago in one country, and two or three years ago in other countries, might also influence results.

The experiment offers preliminary ideas for further investigation of standardized questions, that might work in a particular national environment, but might work differently or not at all in another political environment. Respondents act within their environment. Thus, over-reporting could also be context-dependent. Indeed, there is some general evidence of national differences in over-reporting: Karp and Brockington (2005) conducted a comparative analysis of over-reporting in countries where vote choice can be validated. They found two important contextual characteristics that increase intentional over-reporting: overall turnout and saliency of the elections. Granberg and Holmberg (1991) compared reported and validated turnout in Sweden and in the US and found that the percentage of non-voters reporting that they voted was nearly the same in these two countries.

6 Example 2 - Opinion polls in former authoritarian regimes

The second example can be situated in the comparative electoral studies framework, although it only relates to a single country, Armenia. The overall purpose of presenting the case of Armenia is twofold. First the example offers an insight into election conditions in a country that is still finding on its way to a stabilized political system. Secondly, in order to conduct an electoral study in Armenia the political context needs to be taken into account. Standard concepts used in all the western national election studies cannot be transferred easily and need to be adapted to the local conditions.

Example II thus shifts the perspective from a theoretical perspective, thinking about respondent perceptions and behaviour when answering questions, to the implementation of an electoral study programme in a democratic context which is not yet stabilized.

By way of background, it should be explained that the ER of the ElecDEM project was working together with an international NGO – European Friends of Armenia¹⁷ – on two projects dealing with electoral polls before the elections. The first study was conducted in March 2012, the second in late April 2012.

The role of TNS and the ER was to provide advice and guidance and also to observe the implementation of these studies. He was heavily involved at every stage in of the projects, from questionnaire design to disseminating the results in the local media.

The chapter here will first provide some theoretical and empirical background information on the situation of Armenia. In a second step, the critical points for implementing an electoral survey programme will be discussed. The chapter will end with some conclusions in the broader frame of this final report.

6.1 *Theoretical and empirical basis – Armenia, an emerging democracy*

On 6 May 2012 Armenia held parliamentary elections. After the previous flawed parliamentary elections in 2007 and similarly doubtful presidential elections in 2008, this election was crucial to the country's reputation in the western world. The election campaign started officially in April, a month before the elections, but unofficially, party campaigns began at the end of February. International observers were invited by different national and international political and non-political organizations to

¹⁷ The European Friends of Armenia were the client of TNS opinion for this project. The Experienced Research acted as project leader and spent some time in Armenia in order to implement the two studies.

assess the whether the elections were free and fair and to ensure domestic and international legitimacy.¹⁸

All observers declared the elections as generally legitimate, which of course was important for national politics, but also for international actors involved in the Armenian scene.

During the campaign some topics were hotly debated, including the role of powerful oligarchs in parliament and, of course, the legitimacy of the elections. There were also important geopolitical aspects, but these did not really influence the election campaign.

In order to understand Armenia, we need to consider the recent history of the country and also to reflect on its socio-political conditions.

“One of the most striking outcomes of the third wave of democratisation has been the proliferation of competitive authoritarian regimes” (Bunce & Wolchik 2010). But can this description really be used to describe the situation in Armenia?

The literature highlights some peculiarities of the transition from communist rule and the current context of post-communist politics in Eastern Europe (for an overview see Tiemann 2012). There are vast differences in the contextual factors which significantly affect the options to build representative democratic structures in post-communist countries in Eastern Europe. Tiemann refers to Kitschelt et al. (1999, 21–42) who distinguish between three different types of socialist *anciens régimes*:

(1) Bureaucratic-authoritarian communism corresponds to the ideal type of totalitarian systems (Linz and Stepan, 1996, 40–51; Linz, 2000). The party state is organized by an all-powerful, rule-guided bureaucratic machine. The system can only be overthrown, which will end in a presumably stable, programmatic party competition.

(2) In national-accommodative communism, the state apparatus and governance structures are less developed and professionalized. The legacies of national accommodative communism are expected to lead toward a mixture of programmatic (as opposed to personalistic, charismatic, or clientelist) linkages of parties and voters and thus to moderate levels of party and party system stability and nationalization.

(3) Finally, patrimonial communism relies on vertical chains of personal dependence between leaders in the state apparatus and their entourage. The state apparatus remains in a very low state of bureaucratic professionalism, and politics is dominated by patronage and clientelist networks (Eisenstadt and Roniger, 1984; Kitschelt and

¹⁸ OSCE / ODIHR: <http://www.osce.org/odihr/elections/88247>

Wilkinson, 2007). In extreme cases, these regimes are dominated by “sultanistic” rule of an individual and his family (Linz and Stepan, 1996, 51–54). Legacies of patrimonial communism regularly undermine the chances for the emergence of programmatic linkage structures; rather, they lead to personalistic, or clientelist linkages, to weakly institutionalized political parties, and to low degrees of party and party system nationalization. This is consistent with Bunce & Wolchik, who base their work on the definition of competitive authoritarian regimes as showing both authoritarian and democratic elements, as these regimes tolerate competition for political offices (2010). Due to unequal starting conditions and the use of political and economic power, political leaders reinsure their re-election. Especially in more distant Eastern European countries, the combination of political and economic power is more than obvious. This has the consequence that elections in these political settings usually lead to continuity of the economic pattern, rather than to change in ideological leadership or governing coalitions¹⁹.

Not only western European observers, but also politicians in Armenia very often criticize the domestic political economy and the dollarization of Armenian politics. “In fact, a clique of oligarchs retains considerable economic and political power and common wisdom is that parliament has been turned into an arena where their interests are negotiated.” (Larusso 2012, 4).

Both political parties and the government reflected on this issue and since January 2012 all political forces have tried to exclude or replace oligarchs in political positions in the national parliament.

But this is, of course, a primarily rhetorical act to recapture public support. The role of oligarchs is not just a matter of sitting in parliament. It is strongest in the different constituencies which are economically and politically dependent on these oligarchs.

The question is, therefore, whether this move really reinforced public confidence in the electoral process itself. This is also related to the fact that it is rather common in Armenia for oligarchs to use of their private funds to solve local problems instead of forcing the public administration to deal with them. These aspects are symptomatic of a competitive authoritarian system, as state institutions are weak, economic performance is poor and the level of corruption is rather high (Epstein, Goldstone, Kristensen, OHolloran 2006). We do not want to be drawn into a discussion of whether Armenia in the year 2012 can be described as a competitive authoritarian political system. But it underlines the fact that Armenia is still on its way to become a democratic country.

¹⁹ Even if the results in Georgia are more than satisfactory from the perspective of a lively democracy. The economic power of the political competitors should not be overlooked.

6.2 *Conducting an opinion poll in Armenia*

The parliamentary and presidential elections in Armenia in 2003 and 2008 failed, as authoritarian leaders and their parliamentary allies maintained their hold in political power (See Hyde 2007, OSCE/ODHIR 2003, Whitmore 2007, Tavernise 2008). The common understanding is that the results of the 2008 presidential elections were manipulated. So, the Sargsyan presidency started with the burden of the repression of a long and peaceful demonstration, which resulted in several casualties and injuries.

This, of course, had long-term consequences for these regimes. Another instance occurred during the administrative elections in February 2012, when possible frauds were denounced. Observers reported that direct vote-buying, the promise of administrative resources and positions, were used to put pressure on voters.

Therefore the elections in 2012 were rather important for the future of Armenia. The quality of elections is often seen as a key aspect of the democratic process in the western world. But outside the western world, elections are understood by citizens as a defining feature of democracy.

In the case of Armenia, several sources report unanimously that the United States has linked infrastructure support to “improved” electoral procedures rather than to free and fair elections. This approach was adopted as the country had had a long history of corrupt electoral practices (cited after Bunce & Wolchik 2010, 67).

This does not provide an answer, whether an election fulfils its function and serves the democratic political process such that it opens the way to political and regime change. One argument can be found with Hale, who argues that the promise of continuity in patronage is widely used to “stabilize” political systems (Hale 2005, Hale 2006).

Armenia in 2011 / 2012 can probably no longer be considered as an authoritarian regime. It seems to be more a hybrid system that allows the opposition to win elections. But this would require a level of mobilization, unity and skill, “which goes far beyond what would be normally required in a democracy” (Diamond 2002).

Beside the collaboration of opposition forces with civil society groups that could increase the pressure on the regime, the use of public opinion polls and anything in the nature of campaign and election monitoring could also help parties in opposition. However, this does not only benefit the opposition forces, but works to the advantage of political discourse as a whole in the country, as it introduces elements that increase competition and transparency for voters.

This last has motivated the European Friends of Armenia to organize a series of electoral studies (or opinion polls) to contribute more effectively to political discourse.

But how can a survey be conducted under the conditions described above? This raises a couple of important questions that are in the end valid for any kind of survey. First, the survey must meet the most basic requirements for any survey research: transparency. How can the sample be derived? How to sample respondents? What percentage of respondents needs to be re-surveyed to ensure interviewer accuracy?

If the method by which the survey was conducted is not transparent, it is not possible to estimate its accuracy, and any such survey should be used with caution.

In political regimes that are still not settled democratic systems, politically sensitive surveys are heavily dependent on the respondents' perception of whether or not they are allowed to give a true answer.

A Brussels based international NGO conducted a survey into political attitudes and perceptions in Armenia in February / March 2012. In order to justify the methodological approach used to conduct this survey, TNS was contracted with a local partner to monitor and evaluate the projects. As an organization expert in multi-country research projects, TNS opinion was involved in the project to supervise and contribute at each stage of the survey: from questionnaire design, sampling and field work to data processing and data analysis.

There is no single standard for conducting surveys as each project has unique attributes that need to be incorporated in the design and process. Survey research needs to be adapted to country-specific conditions and an appropriate research design has to be found and defined.

The question remains, however, as to whether these opinions are measurable by conventional methods. In places where dissenting opinions may be punished harshly, the reliability and validity of politically sensitive public surveys is far from self-evident. Yet, as discussed below, political opinion surveys have been used in a variety of settings for many years. As the current situation continues to unfold across new democracies, and as survey research becomes increasingly common under these circumstances, the accuracy of survey results as indicators of domestic opinion, democratic quality and political culture will become increasingly important.

In order to better understand the conduct of an electoral survey in Armenia, it is necessary to understand the background and specific conditions of carrying out surveys in an unfamiliar context. Therefore fieldtrips were organized to discuss the quality criteria with the local experts. The objective of the fieldtrips was to monitor

the quality of survey fieldwork undertaken by the local institute in order to provide external quality control and advice.

During these fieldtrips all relevant aspects for implementing a survey were problematized. These included the recruitment and training of interviewers and supervisors, the proposed sampling approach, data checks and controls, questionnaire design as well as data analysis and dissemination.

However, assuming the transparency of a survey's method, the method should be one that generates a survey with basic integrity: The sample should be randomly drawn from the population of interest and large enough to mirror that population with reasonable accuracy. A sample of inadequate size will lack precision and limit the analyst's ability to make inferences about the opinions of the population. However, quota-sampling techniques—which correct for the problem of undersized sampling by approaching more and more potential respondents until the survey's 'quota' is filled—may be inappropriate in repressive settings, where a particular type of respondent is willing to participate and another type is not. Quota sampling in repressive contexts may inadvertently oversample regime supporters and more outspoken regime opponents while under-representing those who oppose the regime or its policies but are fearful of voicing that opinion.

Of course, the reliability of a survey is only one part of its accuracy. Survey results also must be judged in terms of validity. Survey validity can be broken down into at least two parts: demographic validity and substantive validity. Demographic validity indicates the extent to which the sample maps onto the population it is intended to represent. Substantive validity is the degree to which responses to politically sensitive questions reflect actual opinions and preferences. Even if criteria for survey design, execution, reliability and demographic validity are met, the absence of substantive validity in politically sensitive surveys means that respondents are failing to tell the truth, presumably out of fear of the governing regime. In democratic countries, the honesty of individuals responding to politically sensitive questions, *ceteris paribus*, is the premise underlying all political and policy-oriented survey research. Such honesty cannot be assumed in circumstances of political repression, as respondents may fear to answer frankly.

Here we will only focus on questionnaire design and interviewing, as this covers several interesting aspects for comparative electoral research. However, it should be noted that the involvement of TNS and the ER has led to changes in the sampling design (from a quota to a stratified random sampling approach, with random walk and revisits). Additionally, the instruments for data quality were adapted to international standards. The evaluator also participated in several interviewer training courses, and accompanied interviews in Yerevan and rural areas of Armenia.

The most interesting part of electoral studies is of course the design of a questionnaire. This becomes even more crucial when designing a questionnaire for an unfamiliar political system.

To start with, questionnaires of the British Election Study, the American National Election Study, the Canadian Election Study and other electoral survey projects were used to create a general theme for the questionnaire. In cooperation with the local partners, certain questions needed to be "adapted or localized".

Some highly sensitive questions (on corruption, vote-buying, and dependency) have not been asked, as this could lead to major consequences for the whole study. Other questions were used in the questionnaire without knowing whether they will provide the researchers with valid responses. One of the questions that were excluded is the widely discussed concept of "who did you vote for in the last elections" (example 1). As explained above, the validity and reliability of this question cannot be assessed due to fact that the 2007 and 2008 elections were manipulated. So, what is gained by asking this question? On the other hand, not asking the question would lead to a gap in cross-checking results by asking similar concepts in different forms in the questionnaire²⁰.

This conceptual work was of course done in English, but due to the involvement of the local team, the translation into Armenian was checked directly. This brought up additional issues, as some terms, such as likeable, cannot be translated from English to Armenian without changing the meaning. This had consequences for several questions and answering options, because it was not appropriate to ask these questions in Armenia.

Furthermore, and this underlines the difficult situation in Armenia, in order not to lose credibility the questions related to political parties and leaders were modified so that all major politicians were listed and others could be named. This is of course not unusual, but the fragmented political party system in Armenia, with several popular politicians and former ministers, resulted in a long list of names. This was a trade-off between an efficient and an exhaustive question design, but was absolutely necessary for political reasons.

Interviews were only carried out by young interviewers, mainly female, as these are not perceived as dangerous by the respondents. This is because personal links, economic dependency etc. play a major role, especially in the rural areas of Armenia. This marks another departure from common practice in western Europe, where

²⁰ In the end, the question was used in the traditional form (example 1, treatment A) and newly developed form of Zeglovits / Kritzinger (example 2, treatment C). The results underline the argument in example 1 that the validity of treatment A can be challenged.

interviewers are sought with different backgrounds but very often with a certain level of education and maturity.

This also played a role when the external quality control (call-backs) was conducted. This is a standard procedure to ensure the data quality and is performed regularly on a subsample of the survey. It turned out to be very difficult to validate more than very factual and abstract questions. Respondents may be less likely to express criticism of a political party or leader to a stranger over the phone than to a face-to-face interviewer better positioned to gain the respondent's trust.

Again, the cultural and political context, as well as the personal situation of respondents, reduce the opportunity to validate survey results. But with a good team of interviewers it was possible to validate 200 interviews and in less than 1% of the cases, a single item of information had to be corrected in the final data base; all the other values have been verified for these interviews.

Having established the transparency and integrity of the survey method and execution, the next point covers the discussion and perception of survey results. Of course, several surveys were conducted in this period and therefore it was easy to compare results from different surveys conducted by different organizations. It turned out the results of opinion polls were widely discussed among the broader public and were also used to discredit other organisations. This is again a very interesting phenomenon, which brings to mind Welsh's observation (1981) that in Eastern European countries before 1990 state sponsored surveys were widely used for propaganda purposes. Something similar also happened in pre-electoral Armenia, reflecting the major involvement of oligarchs in party politics. As in other countries, "polling institutes" emerged during the campaign period. All of these institutes published results, which are in 80% of cases biased towards the organization sponsoring the survey. In combination with certain media coverage, these poll results are important aspects of campaigning. All this inspires a degree of distrust in opinion polls in Armenia.

The international NGO, together with the international survey organization TNS and a well-respected local institute (with no political leanings) had to defend themselves against criticism and allegations. Nevertheless, and this is a reminder of what was said before, the only objective of the NGO was to stimulate public discourse by providing valid and reliable measures of public opinion and voting behaviour. This was achieved by providing the public and interested experts with all the technical details for both surveys.

6.3 Discussion in the light of the report

To summarize, operating in full transparency and providing the wider public with all the materials are necessary to contribute to the quality of democratic processes when publishing survey results in countries with developing democratic political systems and an underdeveloped independent media structure. The overarching lesson for using opinion surveys to access public opinion in still weakly democratic societies is to take great care.

Even if past surveys were found to possess a considerable degree of reliability and validity, it cannot be assumed that future surveys will possess this same accuracy. Intervening events, such as election frauds, economic and personal independencies may influence the response patterns of many respondents. And, as discussed above, the choice of survey method may have exaggerated effects in difficult political environments.

To bring the discussion back to what is the overall objective of this final report, election studies cannot be easily transferred from one setting to another. Of course, it is rather unlikely that Armenia would be included in a research design based on opinion data from European countries. The arguments for not doing so would mainly focus on the overall political system and rather than on possible divergence in survey results.

This has also implications for European election studies, as the Armenian example highlights the fact that the transfer of concepts from one country to another might cause measurements that are no longer equivalent. It is interesting to see that this variance is probably not caused by systemic features of different political systems, but by different response patterns and the experiences of respondents in a certain political system.

This example mainly focused on implementing standard electoral research tools in an unfamiliar setting. As set out in the results for example 1, it is necessary to evaluate carefully whether concepts can be operationalized in the same manner in a very different context. Bearing in mind the definition of measurement equivalence, questions might be adapted in a way that leads to question variance, which will cause major problems for the analysis of the data.

That said, this project also underlines the importance of cross-national election studies, as these results can trigger conceptual questions for other election studies, which will ultimately (and on a positivist assumption) lead to more appropriate, valid and reliable measures in election studies.

7 Example 3 - Tackling national variations in the analytical phase²¹

The logistic regression in example 1 was used to analyse data from different countries. Due to the number of countries included in the study above, it was not possible to use multi-level analysis. Therefore another project will be used to show the specific power of multi-level modelling.

In comparative (electoral) research, data are measured at the individual level but also analysed to compare the results from different countries. Many theoretical concepts and hypotheses also rely on the causal power of different levels, such as institutional or cultural contexts. Nowadays the appropriate method for the analysis of hierarchical data is widely used in our discipline.

Multilevel data structures involve data that are ordered hierarchically. For example, in the first example of the question of turnout", data were collected from individuals in different countries, but also in different regions in the countries. This means that voters are nested within regions and countries.

In general, multilevel data structures exist whenever some units of analysis can be considered a subset of other units, and there are data for both subsets.

The goal of multilevel analysis is to account for variance in a dependent variable that is measured at the lowest level of analysis by considering information from all levels of analysis. There are good substantive and methodological reasons for using information from multiple levels of analysis. For the purposes of this paper only the substantive reasons are of importance.

The combination in one comprehensive model allows researchers to minimize misspecification and to establish a model that accounts for the theoretical assumptions.

Furthermore, **causal heterogeneity** can be explored when using multi-level modelling approaches. Cross-level interaction affects the causal effects of the predictors situated at different levels of analysis. This also means that it is possible to evaluate whether relationships vary across higher levels of analysis. Contextual behaviour arises through social interaction within a certain context. This context can be wider, when speaking about countries, or narrow, when analysing pupils in classes or citizens in a certain neighbourhood.

²¹ Based on a forthcoming book chapter: Schwarzer, S. and D. Connor (forthcoming 2012). Political engagement among the youth - effects of political socialization across Europe. Democracy in Transition. Political Participation in the European Union. Kyriakos N. Demetriou, Springer

Last but not least, multi-level analysis can be used to test the generalizability of findings, because the approach allows us to test whether one model works in different contexts at the same time. The question of generalizability arises rather often in comparative research. The classical literature indicates that case studies may work best to compare different cases and to generalize from these several cases (Lijphart 1971, Prezeworski and Teune 1970). But multi-level analysis can also contribute to these tests, because it allows us to explore causal heterogeneity. Assuming that the cases at the higher level are sampled randomly, multi-level analysis might help to overcome the case selection problem normally arising when comparing results of different cases (King et. Al. 1994; Geddes 1990).

From a statistical point of view, the multi-level analysis helps to minimize the danger of incorrect standard errors and also type I errors, in cases for which predictors show significant effect when they do not have an effect under the control for other variables (Steenbergen and Jones 2002).

For the following example a multilevel regression analysis using MLWin 2.25 was used to determine these relationships and their variability between European countries, within countries and within schools; the associated methodology was consistent with that employed by researchers involved with that particular software package (Rasbash, Browne et al. 2000; Jones and Subramanian 2012).

The research in this example is focused on the role of several agents contributing to the political socialisation and political engagement of youngsters. Without going into great detail of the theoretical literature on political socialization and political engagement, the necessary conditions under which young people become involved with politics still remains a puzzle. Aside from questioning what is implicit in creating good citizens, research has increasingly focused on the agents of political socialization; is it all about family or are other factors, such as schools and real and virtual friends also of importance?

7.1 *Theoretical basis of this example*

This chapter sheds further light on the political engagement of school pupils across Europe, while further consolidating important drivers of political engagement, mainly socialization agents and context effects. The relationship between engagement and the current political participation of young students will subsequently be addressed.

So far, political socialization studies have not compared the effect of different agents of socialization across countries. When assessing the improvements suggested in the field, we must address the subject of the agents that play a role in the political socialization process vis-à-vis each other and not strictly separately.

The data were taken from the ICC study on civic and citizenship education. The ICCS is a multi-country study which was conducted in 38 countries (mainly European and South American countries with a small number of Asian countries). The data set contains both questions to build, and pre-built latent constructs which can measure students in terms of important socialisation agents. However, we were not satisfied that these prebuilt indicators suited our needs and we instead rebuilt our own. As a concept, political engagement combines a set of psychological orientations and has been defined as the interrelation of political interest and internal political efficacy (Verba, Schlozman et al. 1995).

To become politically active people require some level of self-belief and self-confidence; in turn the construct of political efficacy reflects an individual's level of belief that political and social change is possible. However, since the ICCS was conducted within schools, this measurement may over-emphasise experiences in that environment. Thus, we note that students' self-reported confidence in their civic participation recognises both the out-of-school and internal school environments.

As a result, we distinguished and segmented the different aspects of political socialization from several perspectives; parental influences, school influences, media influences and more objective measures of the students' background. Home and family have long been considered the most influential factors shaping civic learning outcomes. The family domain is directly and indirectly influenced by parents; direct influences evolve through political discussions, which are regarded as a key element of democratic society. Students were asked about their parent's level of political interest, how often they discussed political and social issues generally, with parents and friends, and how often with regard to other countries.

The media are recognized as important source of secondary socialization (discussed previously), yet research on the effects of media on participation in a democratic society is inconclusive. However, research usually depicts a positive link between media use (information seeking in particular) and political participation. To establish this further, from an extensive literature review and a large-scale study, Norris (2000) found no conclusive evidence for a negative relationship between media use and political participation. The ICCS survey included questions pertaining to the frequency of watching television and reading newspapers and these were included in our media factor. We should note that TV and newspaper consumption has a direct parental influence, as parents typically control which programmes and newspapers are available within a family household.

The role of school cannot be limited to providing information on political institutions, systems and other aspects of a society. Schools are smaller units of society and should establish the values which dominate the wider society. Thus, the role of schools is to establish an environment which helps foster the learning of political

principles more generally. Discussions in schools differ from those conducted within families in both substance and context.

For adolescents, opportunities for active participation are rather limited. However, previous studies (Verba, Schlozman et al. 1995) have emphasized the link between adolescent participation and later involvement as adult citizens. Aside from formal participation (e.g. voting), relatively large numbers of students did participate in voluntary activities such as collecting money or volunteering within an organization dedicated to helping people in the community (Torney-Purta 2002; Torney-Purta 2002). Students may also experiment to determine what power they have to effectively influence how their schools are run, and in doing so may develop a sense of efficacy (Bandura 1997). Democratic practices in schools can provide students with a means of ascertaining the usefulness of political action and may foster strong values and norms for their adult political participation. Participation in civic activities reflects student involvement in collective civic engagement but is not part of the formal learning context and is only weakly related to education.

7.2 Modelling

Students responded to a vast array of questions on different aspects of socio-political life. When appropriately clustered, combinations of these responses can indicate a student's level of activity and opinion on a range of theoretically important facets.

Methods of Multiple Correspondence Analysis/Optimal Scaling were employed to extract these measures (Benzécri 1992; Meulman, van der Kooij et al. 2004). Optimal scaling provides a way of obtaining quantitative scale values for categorical or ordinal data (Greenacre 2007: 49 - 57) and in this instance was used to quantify students' original ordinal responses into single 'components'. The technique is believed not to be a population-generalizable tool in its own right but rather a method of reproducing the structure of categorical variables (Greenacre 1984; Meulman, van der Kooij et al. 2004; Panagiotakos and Pitsavos 2004) and is thus suitable here.

The variable selection was theory-driven and our scales are not intended to be broadly generalizable measures outside of this setting. The Political Engagement (dependent variable) ranking was constructed using two separate methods based on three broad latent attributes that characterize the student politically; citizenship-self efficacy, self-concept in politics and interest in politics (table 2). In the first instance we constructed these three broad variables and then merged them into a single construct using principal components analysis. In the second case, a single optimal scaling procedure yielded the dependent variable from the original 18 ordinal response questions; both solutions were tested in the model and yielded almost identical results. The variable produced from the latter method was chosen as it is both more transparent and less prone to distortion.

Table 1 Level structure

Hierarchy	Nested in higher level				
Levels	N	Min	Max	Mean	Standard Deviation
Countries (v)	23	1	23	-	-
Schools (u)	3114	1	199	137	47
Students (e)	67909	1	379	23	18

Six models were estimated and fitted using Iterative Generalized Least Squares Estimation (IGLS), Wald tests were used to evaluate individual parameters while the log likelihood deviance statistics were used to compare model fit. The best fitting model included both random intercepts and slopes across three levels.

7.3 Results

The predictors of 'Political Engagement' among adolescents in this age group appear to be quite diverse; our hierarchical regression coefficients appear reasonably balanced across home, school and media coefficients (table 4). Despite this, a dichotomy did emerge between family/parental indicators and all others. It is clear that in terms of model fit, parental discussion and parental interest should be modelled as both random slopes and fixed effects; the implication of this will be discussed in the following section.

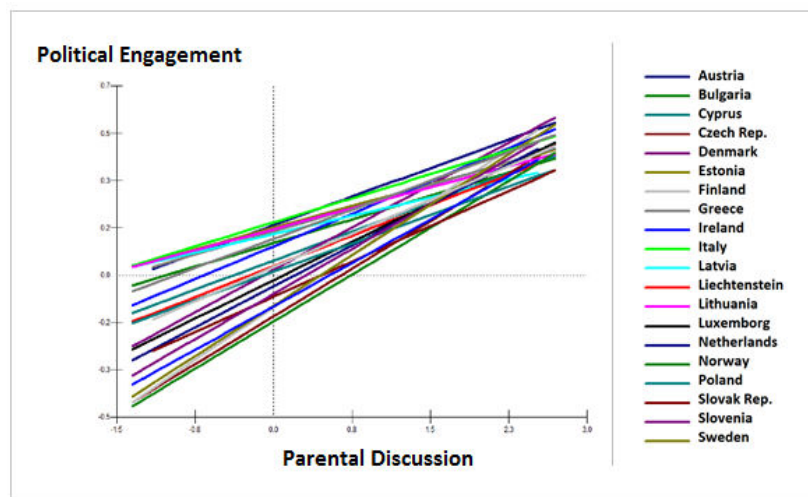
Table 2 Multilevel Model

	Effect Size	P-Values	+/- 95% CI
Random			
Level 3 – Europe			
European Constant (Var)	0.014	0.000**	0.008
Parental Discussion (Var)	0.002	0.000**	0.001
European/Parental Disc (Cov)	-0.005	0.000**	0.003
Level 2 – Schools			
School Constant (Var)	0.010	0.000**	0.008
Parental Discussion (Var)	0.003	0.000**	0.001
School/Parental Disc (Cov)	-0.003	0.000**	0.001
Level 1 - Student			
Student Constant (Var)	0.489	0.000**	0.008
Parental Discussion (Var)	0.08	0.000**	0.005
Parental Interest (Var)	0.059	0.000**	0.007
Student/Par Int (Cov)	-0.058	0.000**	0.004
Student/Par Disc (Cov)	-0.133	0.000**	0.004
Par Disc/Par Int (Cov)	-0.007	0.000**	0.004
Difference (Constants)			
Level 3 Cons – Level 2	-	0.35	-
Level 3 – Level 1	-	0.000**	-
Level 2 – Level 1	-	0.000**	-
Fixed			
Constant	0.039	0.12	
Individual			
Female	-0.058	0.024*	0.014
Media Use	0.178	0.000**	0.006
Age	0.012	0.045*	0.01
Family			
Parental Interest	0.186	0.000**	0.007
Political Discussion (Parents)	0.169	0.000**	0.022
Single Parent Family	0.054	0.000**	0.014
Nuclear Family	0.022	0.000**	0.049
School			
School Engagement	0.170	0.000**	0.006
School Participation	0.126	0.000**	0.006
Political Discussion (School)	0.058	0.000**	0.006

The model’s fixed effect parameters represent the resulting change in political engagement due to one unit increase in that parameter – all ‘factors’ (continuous parameters) were centred at their grand mean of zero (with standard deviations of one). A one unit change in one of these ‘factors’ is equivalent to a one standard deviation increase/decrease. In the case of our categorical or dummy parameters (e.g. Girl), a one unit increase implies that you do exhibit the parameter attribute.

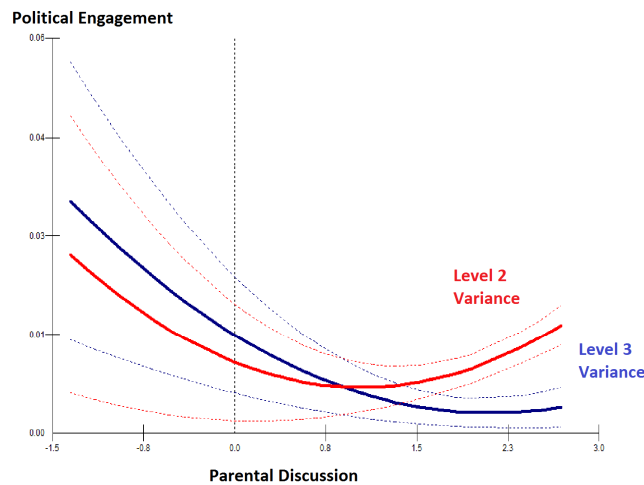
The ‘random’ parameters of the model illustrate differential intercepts (constants) between schools, countries and pupils and random slopes for the differential effect of parental interest and discussion. Put simply, our highly significant level two and three variance intercepts allows us, with a very high degree of confidence, to state that the average level of political engagement differs between schools within the same country and between schools in different countries, and that this is not the result of random chance. Similarly, there is quite a large degree of difference in political

engagement between students within the same school (mean of 0.7 standard deviations from the within class mean). The empty model was partitioned with 3% of the variance in the model being due to differences between countries, 4.5% due to differences between schools in the same country and 92.5% due to differences within schools. Students in the same school had a very weak positive correlation +0.07. The proportion of higher level variance due to difference between schools within countries is 43% - a value very close to half while differences in schools between countries is the reverse and would be 57%. This indicates that the degree of heterogeneity between schools within the same country and schools across different countries is similar. This was further corroborated through finding no significant difference in the intercepts of schools-within and schools-across countries. However, the three-level model was a significant improvement on the two level model and we wanted to model as much dependency as possible in our data

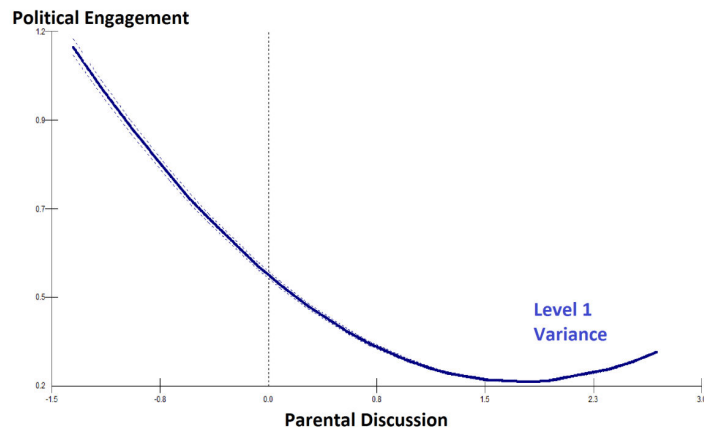


Even outside the political spectrum, parent-child relationships are complex and this remains true in this context. The effect that increasing political discussions with parents has on the child's level of political engagement was allowed to have a fixed effect for all students but also random slopes for individuals in within the same school, country and across countries. Similarly, parental interest had a fixed effect and a random slope at level one. These random slopes not only imply that cohorts in some hierarchies have higher or lower returns in political engagement due to parental discussion or interest (random intercepts) but the overall relationship or slope is context-dependent. This is depicted visually in figure x as we view the random slopes across different European countries. Our intention is not to engage in a discussion of differences throughout Europe but simply highlight that a) across-country differences are multidimensional and b) the 'fanning-in' of the country lines suggests that when discussions with parents are plentiful (a high value), the result in political engagement

will be similar despite the country of residence. However, when parental discussion is low, the effect on political engagement is far more variable. The school and within-country levels appear similar to this.



If we view this relationship through the higher level variance functions we are presented with a similar image. For low levels of parental discussion we observe high variance in political engagement (higher for schools across countries) decreasing until approximately $+1$ standard deviation from the mean in parental discussion when the variances diverge again. Countries become continually less variable in political engagement as parental discussion increases - schools tend to become more variable in the same country as parental discussion increases. The 'tram lines' encompassing the variable functions represent the 95% confidence intervals. It is clear that when parental discussion is low, we cannot distinguish between schools within the same country and between countries. However, as parental discussion increases, the distinction between the two is clearer. Given that the confidence intervals overlap at low levels of parental discussion, we cannot confidently state that schools vary differently across and within countries at these levels. We can, however, confidently state that differences do exist between and within countries for higher levels of parental discussion. We can note that within countries (level 2), political engagement reaches a minimal level of variance and then begins to become more variable again, giving the quadratic function a 'bowed' shape. In contrast, we observe that as discussion increases, variance between countries continually decreases toward zero. For pupils in the same school (figure x), the relationship holds and is more pronounced. Although the function does begin to increase again toward the tails, this is more of an outlier effect as we start to include students that are very far from the mean.



School-related factors clearly play a role in the student’s political engagement but these do not appear to be actually operationalized between schools in any consistent way; schools are still too heterogeneous. The effects of schools appear to be more ‘fixed’ than parental-related factors; indicating that the outcomes of school-related factors are less variable across contexts than parental factors. While we can tentatively state that with few exceptions, each student has a wholly individual set of parents and family relationships we cannot assume the same in the case of schools. Single schools not only contain many more students than the typical family, but their environments are more easily reproduced geographically; school-settings are in many senses contrived environments that teach from centrally designed curriculums employing teachers trained in similar environments. After controlling for school typologies, schools in different places may be quite similar; if we exclude differences in schools based on ‘school type’ (which include the economic, social, cultural and environmental indicators of their students), we are typically left with something closer to a centrally designed system that is reproduced regionally or nationally to educate citizens. Although parents with similar political dispositions may engage or transfer information in the same way to their children, schoolchildren do not typically cluster in schools based on this parental interaction. The relationship from parent to child is one-to-one, while school or teacher to child is one-to-many. Further, our hierarchy is designed as school, and not family-orientated. Thus, it is predictable that the impact of schools upon political socialization is considerably more ‘fixed’ than the effect of parents. Of course students perceive school information and experiences individually; nonetheless, the ‘school’ element in that relationship is certainly less dynamic than the one experienced at home. We can see that the social-environmental factors affecting individual political engagement, which are not related to schools, stand on strong roots embedded in individual family environments. That is, the ‘random’ element appearing within schools, has deeper roots outside school, than in.

One issue which hampered our ability to distinguish between schools was that school types are confounded with schools as a whole in the data; as a result, we could not model these differences. We have shown that differences do exist between schools and students at each level; however this might be accentuated if we had the means to classify schools based on multiple criteria (academic streams, geographic location, financial attributes). For example, it is plausible that some schools in Austria may be more similar to counterparts in Ireland or Norway than they actually are with other schools in Austria. This implies that the school (system) matters as we suspect schools vary differently within different countries. Certainly, further research is required to add validity to this claim.

The effects of gender appear to be heavily mediated by factors relating to school. After controlling for school-related parameters, the relationship between being a female and political engagement, goes from being positive to negative. Thus, it appears that school is an immediate source of gender division in political engagement at this age. Further, age appears to matter, but only slightly. Despite students being heavily clustered in age, each extra year does increase political engagement to a small degree.

Approximately half of the base higher level variance in political engagement was explained by school-based, political indicators. This indicates that political interaction differs across schools and this has a direct impact on students' political engagement. This implies that providing opportunities to participate in political decision-making, and opening up opportunity structures, help to develop a level of political engagement at this early age. As increasing age also shows an effect, we can hypothesize that engagement goes with experience and that the crystallization of attitudes is related to school-based factors.

We have already given much attention to the relationship between politically interested parents and the student; we have concluded that interested parents or higher political activity between child and parent is far less variable than politically disinterested parents or lower levels of political activity. Based on this result, we more confidently conclude that interested parents are more likely to have politically engaged children, than disinterested parents are to have disengaged children. This further implies that the family context matters significantly, as it lays the foundation for the political engagement of young people. Once again, this should emphasize the importance of schools in forming political engagement; as they provide children in the 'disinterested spectrum' a platform to encounter political socialization at the second stage.

Briefly turning to the individual-context parameters in the model, it is interesting to note that the socio-economic background of parents almost entirely lost its predictive

power for political engagement, whereas the political interest of parents remains influential. Educational background, which is normally a good predictor for school achievement, is less important when it comes to the political engagement of young pupils. Bearing in mind the strong link between political interest of parents and political discussions at home, both quite strong predictors of political engagement for this age group, we should be careful about drawing a further conclusion that higher education leads to a higher interaction between children and parents.

7.4 Discussion of the results in the light of this report

This analysis sought to evaluate the role of different agents of political socialization in political engagement on the one hand and political participation in a multi-level setting on the other.

Theoretical discussions indicate that agents such as media and schools should be considered as a second important source of political participation, as adolescents spend more time in school than they actually spend with their families.

Political socialization and political engagement are related factors that also determine the quality of electoral democracy.

Additionally, political engagement should, to a certain extent, function as a mediation variable between political socialisation on the one hand and political participation on the other. Of course, the relationship or the mediation cannot be established in a causal relationship as political participation depends on the complex interplay of individual factors and opportunity structures within a given political setting.

It appears that even after controlling for the multilevel structure of the data, all the agents of political socialization are important in the development of the political engagement of students. It can be safely argued that the agents of primary socialization are more influential than agents of the second phase of socialization, such as media and schools. The influence of parents and the family domain appear to be quite strong determinants of socialization.

Furthermore, based on the data structure we expected to find different effects within countries (schools) and between countries.

For this study, it can be assumed that the data comply with the requirement of measurement invariance across countries. Furthermore, at least by design, a wide range of countries was included in the study, which was not limited to a specific set of countries. Due to external factors, two-thirds of the countries are European, but countries from all over the world participated in the study and can be included in the analysis.

Earlier on the report put emphasis on equivalence of measurements in cross-country research projects. It has been shown that this is a key factor for the comparison of results from different countries, and also for applying a multilevel analysis.

For the multilevel analysis researchers have to assume that the measurements are equivalent in all the countries and that the contextual units reflect a proper sample. In reality, cross-national surveys have quite different design and measurement features.

This is related to the fact that meaningful cross-national comparisons require measurement invariance. Measurement invariance (MI) is defined in many different ways. It is also referred to as measurement equivalence, but in either case it refers to the consistency of measurement across some specified groups.

A lack of invariance can produce misleading results in a multilevel analysis as random variation in intercepts and slopes are directly affected. In the absence of measurement invariance the intercept and slopes for the manifest variables will vary across the second hierarchical level (Steenbergen and Jones 2002). In other words, random slopes and intercepts have two different sources: the relationship between the underlying latent variable and a covariate varies across countries; the measure of the latent variable is not invariant across the countries. Without further analysis it is unclear whether the effects are related to the nested structure of the data or to problems in the measurement of data. In fact, the dilemma in interpreting the results can only be resolved through measurement equivalence. In order to establish measurement invariance multiple indicators are necessary, which are rarely available in social science surveys.

Besides measurement equivalence, the number of cases at the second level is the second source of error. In many cross-national surveys the number of cases is limited and the countries are not drawn randomly. This has consequences for the estimates as variance components are downwardly biased (Mok 1995) or negative (Bian 2002).

To demonstrate the use of multilevel analysis the report uses the data from the International Civic and Citizenship Education study (ICCS). This study focuses on the achievement of students at 13 and 14 years of age in the field of citizenship education, political engagement and political socialization in 38 countries.

The particular characteristic of this study is its tripartite structure, with pilot study, field trial and main study. This rather complex study design specifically addresses the problem of measurement equivalence across all the countries. The pilot study in a subset of countries and the field trial in all participating countries are based on a multiple indicator approach and advanced techniques to analyse the dimensionality of the data.

8 Conclusion

The three examples presented are quite different but cover important aspects of cross-cultural (electoral) research. Furthermore, all three examples demonstrate the context-dependent nature of survey research. "Measurement is a crucial first stage in any comparative research project" (Peters 1998: 107)

The conceptualisation of the relationship between the macro and micro levels when trying to explain social reality has been an endeavour for social science theory and empirical research. In particular, the linkage between the context and the individual level involves difficulties when trying to deduce hypotheses that explain individual behaviour and attitudes by e.g. country characteristics. The development of the so-called multilevel models (hierarchical linear models – see e.g. Hox 2003) allow us to test for the interrelation between individuals and contexts, so that we can assess to what extent structural variables influence individual behaviour and attitudes.

However, a test for significant interrelations between structure and agency in cross-national research implies equivalence of concepts, constructs and measures.

Therefore, major efforts need to be made during the whole research process in order to secure comparability across contexts, including the development and selection of the underlying concepts for research and their adequate operationalisation (i.e. translation of concepts into measures) within the units of observation (countries)²². 'Adequate' here means able to secure validity, reliability and comparability of measurement within and between countries.

Especially in cross-cultural/cross-national research, we presuppose that concepts and measures can "travel" easily from one country or cultural context to another (see Satori 1970, also 1991). Yet the "travelling" of concepts can only be achieved if the indicators developed to measure them provide high measurement quality within and between contexts. Both the experiment in example 1 and the Armenian example demonstrate that concepts and questions do not travel easily.

Usually concepts - as ideas that unite phenomena under single terms (Bollen 1989: 180) - cannot be measured directly. A measurement model has to be specified which includes the latent constructs (factors) and the indicators (questionnaire items). This measurement or structural model is then tested for equivalence within and across groups.

As demonstrated in our examples, the quality of measurement in cross-cultural and cross-national studies is even more difficult to achieve. Bias and measurement error

²² See also Uskul et. al. 2010; Pan et.al 2010

produce non-equivalence, for which we have to control. Non-equivalence is not an intrinsic problem of items within a questionnaire, but appears in cross-national/cross-cultural research.

Bias refers to the presence of spurious factors that challenge the comparability of scores across social categories, groups or across countries. Bias can be produced when differences in institutional arrangements (time passed since the last elections) or historical pathways at national level (Armenia) are not explicitly accounted for when developing a comparative research instrument: "the psychological meaning is culture/group dependent and group differences in assessment outcome are to be accounted for, at least to some extent, by auxiliary psychological constructs or measurement artefacts" (Van de Vijver 2003: 144). This means that a concept is not measured in a way which recognises the specific culture of a group or does not reflect systemic arrangements in a country. By recognising the specificity of different cultures when establishing measurement instruments, researchers can minimize the danger of concept bias. One strategy to avoid or to minimize concept bias is decentralization (the simultaneous development of instruments); another is the convergence method (within-culture development and cross-cultural administration of questionnaires and surveys). An item is biased "if respondents with the same standing on the underlying construct (e.g., they are equally intelligent), but who come from different cultures, do not have the same mean score on the item" (Van de Vijver 2003: 148). Item bias can be avoided and minimized by using psychometric and judgemental methods or (cognitive) pre-testing of instruments in the different participating countries. Cross-cultural differences may also be identified in the modes and rates of respondents' predisposition to offer socially desirable answers.

Following this argument, we may conclude that different forms of bias result in considerable limitation of the comparability of concepts and measures, i.e. bias that originates from an inappropriate consideration of institutional differences between countries may lead to non-equivalence of constructs and measures. "Equivalence refers to the comparability of test scores obtained in different cultural groups" (Van de Vijver 2003: 144). Equivalence is a key concept and core requirement in comparative research. Bias and equivalence are two sides of the same coin. If bias occurs, non-equivalence is detectable and vice versa. But it seems important to disentangle these concepts in a cross-cultural analysis in order to show their specific effects on the scores measured. The identification of bias and the verification of equivalence are core methodological aims in cross-cultural and cross-national survey research.

Scholars have developed many definitions of equivalence and it occurs in a variety of forms, e.g. procedural vs. interpretative equivalence: "Although equivalence has

multiple dimensions, there seems to be a natural distinction between interpretive and procedural equivalence. While interpretive equivalence is primarily concerned with the subjective cross-cultural comparability of meaning, procedural equivalence, broadly speaking refers to the objective development of comparable survey measures across cultural groups.” (Johnson 1998: 29; see also Mohler et. al. 2010)

Another definition was developed by Fons van de Vijver who distinguishes three types of equivalence: “construct (identity of construct across cultures), measurement unit (identity of measurement unit), and scalar equivalence (identity of measurement unit and scale origin)” (Van de Vijver 1998: 41).

These three types constitute a hierarchical scheme with construct equivalence (also known and related to as functional or structural equivalence) as the basic form of comparability. It means that similar constructs are measured in each cultural group and describes the identity of constructs across cultures.

Construct equivalence indicates that the concepts compared are the same across social categories, groups or countries. This is the precondition for any comparative analysis. The aim is to examine constructs that are structurally or functionally equivalent. Group differences in mean scores could be related to a mixture of valid cross-cultural differences and measurement artefacts. To avoid this kind of problems, we should explicitly control for both validity and method variance.

This implies that comparisons of different groups (categories) with different cultural origins can only be made if the items show scalar or full score equivalence. “Scalar equivalence assumes the identical interval or ratio scales across cultural groups” (Van de Vijver 2003: 154). It is easier to achieve construct equivalence than to give evidence for full score equivalence.

The core of the measurement instrument in electoral research projects consists of behavioural, attitudinal or motivational variables measuring the political engagement and behaviour of citizens. The research objectives in these projects are of course very different, but the main focus is to analysis the data assuming functional equivalence of the applied concepts and measures. Functional equivalence refers to similarity rather than to identity of constructs. But in testing functional equivalence it is necessary to examine whether the structure of political engagement is comparable across the countries studied. Thus, functional equivalence could also be conceptualised as construct equivalence.

Although the specific context is often recognised as important for interpreting results, few research designs reflect the importance of case selection. This is to some extent related to the uncertainty regarding the definition of the unit of observation and the unit of analysis. Are we going to compare individuals or countries?

Individuals and social categories (groups, countries, cultures etc.) are often conceptualized in a hierarchical system, yet these separate levels of analysis are not implicit in the research design. In the case of Armenia, this has consequences for the way concepts have to be operationalized, whereas in our example addressing political socialization the differentiation of hierarchies is of importance.

The variance of any dependent variable can be explained at least by two types of variables: a) variables on the individual level, and b) variables on the meso- or macro level. Usually, the characteristics of countries are seen as independent variables representing the macro-level of comparison and thus must be explicitly accounted for in the explanatory model.

The unit of observation in all current cross-national / cross-cultural research is a culture or society (Denton 2007, p. 3). The unit of analysis denotes the object regarding which behavioural relations among theoretical constructs are conceptualized. It is 'what' or 'whom' that is being studied. This is not to be confused with the unit of observation, which is the unit on which one collects data (compare to Denton 2007). This means that the unit of observation is a group of units of analysis. For example, we might conceptualize political behaviour as that of individuals (the unit of analysis), but measure the countries' modal, median or mean individual political participation.

This means that a study design has to differentiate between the unit of analysis at the individual level and the unit of observation at the country level.

Conclusions based on project results are made at the national level, but the analysis is based on the individual level. The example of political socialisation demonstrates that data might have several levels and that differences between countries might be influenced by another level within countries.

In almost all multi-country projects individuals are the unit of analysis, but the general objective is to compare countries, which are in this case the units of observation. However, if we are interested in comparing countries, then the selection of cases marks a crucial point for the research design. By controlling for differences between countries when designing the research project, we can minimize but not totally avoid non-equivalence. Social sciences provide well-developed approaches to the comparison of similar and different countries (Lijphart 1971, Przeworski and Teune 1970, Teune 1975, Teune 1990).

Even if the unit of analysis is individuals, it is of importance to keep in mind that cross-national and cross-cultural research projects are faced with problems which go beyond adequate translation of questions within a questionnaire.

Furthermore, the case (country) selection is very often not made on the basis of reasonable arguments about the diversity of institutional arrangements and the historical evolutions of political systems and contexts. In most cases the countries covered reflect a pragmatic approach in international research projects - existing networks and availability of partners may determine case selection. This fact has significant implications for the explanatory power of the models – the smaller the number of cases as the units of observation, the more important is it to have a systematic and reasoned selection of cases (countries).

Case selection is thus related to equivalence of concepts and measures and should be implicit in the evaluation analysis of the measurement quality and comparability of items.

Aspects such as equivalence, comparability and the hierarchical structure of phenomena and data have been discussed in this report, but also during the course of ELECDEM.

Since I joined TNS opinion on the basis of the ELECDEM scholarship, these issues have become part of my day to day work.

The ELECDEM training programme thus enabled me to develop (or 'acquire') a wide range of insights into theoretical aspects.

9 References

- Abelson, R. P., Loftus, E. F., & Greenwald, A. G. (1992). Attempt to Improve the Accuracy of Self-Reports in Voting. In J. M. Tanur (Ed.), *Questions about Questions. Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Almond, G. A. and S. Verba (1963). *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, Princeton Univ. Press.
- Almond, G. A. and S. Verba (1989). *The civic culture: Political attitudes and democracy in five nations*. Newbury Park, SAGE Publications.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, Freeman and Company.
- Belli, R. F., Moore, S. E., & VanHoewyk, J. (2006). An experimental comparison of question forms used to reduce vote overreporting. *Electoral Studies*, 25(4), 751-759.
- Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63(1), 90-108.
- Berger, P. and T. Luckmann (1967). *Die Gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. Frankfurt a. M., Fischer Verlag.
- Benzécri, J. P. (1992). *Correspondence analysis handbook*. New York, NY, Marcel Dekker.
- Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65(1), 22-44.
- Bian, G. (2002). "Bayesian Estimates in an One-Way Anova Random Effects Model." *Australian & New Zealand Journal of Statistics* 44(1). 99 - 108.
- Billiet, J., Koch, A. and Philippens, M. (2007). Understanding and improving response rates. In: *Measuring Attitudes Cross-Nationally* (Eds, Jowell, R., Roberts, C., Fitzgerald, R. and Gillian, E.) SAGE Publications, London, pp. 113 - 139.
- Bollen, K.A. (1998). *Structural Equations with Latent Variables*. New York, John Wiley & Sons.
- Brownlee, J. (2007). *Authoritarianism in an Age of Democratization*. Cambridge, Cambridge University Press.
- Brownlee, J. (2009). *Harbinger of Democracy, Competitive Elections before the End of Authoritarianism. Democratization by Elections*. S. Lindberg. Baltimore, Johns Hopkins University Press.
- Bryk, A. S. and Raudenbush, S. (1992). *Hierarchical linear models. Applications and data analysis methods*, Sage Publications, Thousand Oaks.

- Bunce, V. J. and S. L. Wolchik (2010). Defeating Dictators. Electoral Change and Stability in Competitive Authoritarian Regimes. *World Politics* 62(1). 43 - 86.
- Campbell, D. E. (2008). Voice in the Classroom: How an Open Classroom Climate Fosters Political Engagement Among Adolescents. *Political Behaviour* 30: 437-454.
- Cassel, C. A. (2003). Overreporting and Electoral Participation Research. *American Journal of Politics*, 31(1), 81-92.
- Chang, & Krosnick, J. A. (2009). National Surveys Via Rdd Telephone Interviewing Versus the Internet Comparing Sample Representativeness and Response Quality *Public Opinion Quarterly*.
- Couper, M. P. and de Leeuw, E. D. (2003). Nonresponse in Cross-Cultural and Cross-National Surveys. In: *Cross-Cultural Survey Methods* (Eds, Harkness, J. A., Van de Vijver, F. J. R. and Mohler, P. P.) John Wiley & Sons, Hoboken, New Jersey.
- Dalton, R. (2006). *Citizen Politics: Public opinion and political parties in advanced industrial democracies*. Washington, Chatham House.
- Dalton, R. (2008). Citizenship Norms and the Expansion of Political Participation. *Political Studies* 56: 76 - 98.
- Dawson, R. E. and K. Prewitt (1969). *Political Socialization*. Boston, MA:, Little Brown and Co.
- Delli Carpini, M. X. and S. Keeter (1996). *What Americans know about politics and why it matters*. New Haven, CT, Yale University Press.
- De Leeuw, E., & Collins, M. (1997). Data Collection Methods and Survey Quality: an Overview. In L. Lyberg, P. P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 199-220). New York: Wiley.
- Denton, T. (2007). Unit of Observation in Cross-Cultural Research: Implications for Sampling and Aggregated Data Analysis. *Cross-Cultural Research*, 41, 3-31.
- Diamond, L. (2002). Thinking about Hybrid Regimes. *Journal of Democracy* 13(2). 21 - 35.
- Eisenstadt, S. and L. Roninger (1984). *Patrons, Clients and Friends. Interpersonal Relations and the Structure of Trust in Societies*. Cambridge, Cambridge University Press.
- Epstein, D., R. Bates, et al. (2006). "Democratic Transition." *American Journal of Political Science* 50(3). 551 - 569.
- EUYOUPART (2005). EUYOUPART Deliverable 17: Final Comparative Report. Political Participation of Young People in Europe - Development of Indicators for Comparative Research in the European Union (EUYOUPART), Forschungsprojekt im Rahmen des 5. Rahmenprogramms, Vertragsnummer HPSE-CT-2002-00123. Vienna, SORA.
- Franklin, M. N. (2004). *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge, Cambridge University Press.

- Galston, W. A. (2003). Civic education and political participation. *Phi Delta Kappan* 85(1). 29-33.
- Geddes, B. (1990). How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2: 131 - 150.
- Granberg, D., & Holmberg, S. (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35(2), 448-459.
- Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London, Academic Press.
- Greenacre, M. J. (2007). *Correspondence analysis in practice*. Boca Raton, Fla. ; London, Chapman & Hall/CRC.
- Hale, H. E. (2005). Regime Cycles: Democracy, Autocracy, and Revolution in Post-Soviet Eurasia. *World Politics* 58(1). 133 - 165.
- Hale, H. E. (2006). Democracy or Autocracy on the March? The Colored Revolution as Normal Dynamics of Patronal Presidentialism. *Communist and Post-Communist Studies* 39(3). 305 - 329.
- Harkness, J. A. (2007). Improving the comparability of translations. In: *Measuring Attitudes Cross-Nationally* (Eds, Jowell, R., Roberts, C., Fitzgerald, R. and Gillian, E.) SAGE Publications, London, pp. 79 - 95.
- Held, D. (1996). *Models of Democracy*. Standford, Standford University Press.
- Hill, K. Q., & Hurley, P. (1984). Nonvoters in Voters' Clothing: The Impact of Voting Behaviour Misreporting on Voting Behaviour Research. *Social Science Quarterly*, 65(1), 199-206.
- Hooghe, M. (2004). "Political socialization and the future of politics." *Acta Politica. International Journal of Political Science* 39(4). 331 - 341.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social Desirability Bias in Voter Turnout Reports. *Public Opinion Quarterly*, 74(1), 37-67.
- Hox, J. (2003). *Multilevel Analysis. Techniques and Applications*, Lawrence Erlbaum Associates, Mahwah, NJ.
- Horne, C. D. (2011). *Opinion and Policy beyond the Democratic West: The Case of Iran*.
- Hyde, S. (2007). The Observer Effect in International politics: Evidence from a Natural Experiment. *World Politics* 60(3). 37 - 64.
- Hyman, H. (1959). *Political Socialization*. Glencoe, Illinois, Free Press.
- International Election Observation (2012). *Statement of Preliminary Findings and Conclusions*, 7 May 2012. Yerevan.
- James, A. and A. L. James (2004). *Constructing Childhood*. New York, Palgrave Macmillan.

- Johnson, T. P. (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. In: Cross-Cultural Survey Equivalence. ZUMA-Nachrichten Spezial, Nr. 3. (Ed, Harkness, J. A.) ZUMA - Zentrum für Umfragen, Methoden und Analysen, Mannheim.
- Jones, K. V. and S. Subramanian (2012). Developing multilevel models for analysing contextuality, heterogeneity and change using MLwiN 2.2.
- Jowell, R. (1998). How Comparative is Comparative Research. CREST Working Paper, No. 66, Center for Research into Elections and Social Trends, Oxford
- Jowell, R. (2007). The European Social Survey as a measurement model. In: Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey (Eds, Jowell, R., Roberts, C., Fitzgerald, R. and Gillian, E.) SAGE Publications, London, pp. 1 - 33.
- Karp, J. A., & Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries. *Journal of Politics*, 67(3), 825-840.
- Khachatryan, H. (2005). Armenia: A Guinness-Fuilled Vote? Transition on Line, December 1.
- Khachatryan, H. (2005). Armenia: Not Sizzling, but Fizzling Out. Transition on Line, December 19.
- King, G., R. O. Keohane, et al. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, Princeton University Press.
- Kitschelt, H., Z. Mansfeldova, et al. (1999). *Post-Communist Party Systems. Competition, Representation and Interparty Cooperation*. Cambridge, Cambridge University Press.
- Kitschelt, H. and S. I. Wilkinson (2007). *Patrons, Clients and Policies*. Cambridge, Cambridge University Press.
- Larusso, M. (2012). The 2012 Armenian Parliamentary Elections: Implications for Armenias Foreign Policy. Rome, Istituto Affari Internazionali IAI Working Paper 12/ 14 May 2012.
- Latcheva, R. (2008, forthcoming). Cognitive Interviewing and Factor-Analytic Techniques: A Mixed Method Approach to Validity of Survey Items Measuring National Identity. Vienna.
- Langton, K. and K. M. Jennings (1968). Political socialization and the high school civics curriculum in the United States. *American Political Science Review* 62: 862-867.
- Lijphart, A. (1971). Comparative Politics and the Comparative Method. *American Political Science Review*, 65, 682-693.
- Lijphart, A. (1975). The Comparable-Cases Strategy in Comparative Research. *Comparative Politics Studies*, 8, 158-178.
- Linz, S. L. L. (1996). *Problems of Democratic Transition and Consolidation*. Baltimore, London, John Hopkins University Press.
- Linz, S. L. L. (2000). *Totalitarian and Authoritarian Regimes*. Lynne Rienner Publishing, Boulder.

- MacInnes, J. (2005). Category and Comparison Across what kind of Frontiers? ZUMA-Nachrichten, Spezial Band, nr. 12, Papers from the CSDI 2005
- Markus, G. B. (1979). The political environment and the dynamics of public attitudes. A panel study. *American Journal of Political Science* 42(2). 461 - 487.
- McIntosh, H. and J. Youniss (2010). Toward a Political Theory of Political Socialization of Youth. *Handbook of Research on Civic Engagement in Youth*. . L. R. Sherrod, J. Torney-Purta and C. A. Flanagan. Hoboken (NJ), Wiley.
- Meulman, J. J., A. van der Kooij, et al. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. *The Sage handbook of quantitative methodology for the social sciences*. D. Kaplan and S. P. inc. Thousand Oaks, Calif., Sage: xiii, 511 p.
- Mohler, P. P., Johnson, T. (2010). Equivalence, Comparability, and Methodological Progress. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Harkness, J., Braun, M., Edwards, B, Johnson, T., Lyberg, L., Mohler, P., Pennel, B-E., Smith, T. Wiley 17 - 32
- Mok, M. (1995). Sample Size Requirements for 2-Level Designs in Educational Research. *Multilevel Modeling Newsletter* 7(2). 11 - 15.
- Mondak, J. J. and K. Halperin (2008). A framework for the study of personality and political behaviour. *British Journal of Political Science*(38). 335 - 363.
- Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanations*, Harcourt, Brace & World., New York.
- Nie, N., J. Junn, et al. (1996). *Education and Democratic Citizenship in America*. Chicago, University of Chicago Press.
- Niemi, R. G. and J. Junn (1998). *Civic education: What makes students learn?* New Haven, Yale University Press.
- Norris, P. (2000). *A virtuous circle: Political communication in postindustrial societies*. Cambridge, UK, Cambridge University Press.
- OSCE / ODIHR (2003). *Republic of Armenia: Parliamentary Elections 25 May 2003. Final Report*. Warsaw, OSCE.
- Panagiotakos, D. B. and C. Pitsavos (2004). Interpretation of epidemiological data using multiple correspondence analysis and log-linear models. *Journal of Data Science*(2). 75-86.
- Peters, B. G. (1998). *Comparative politics: theory and methods*, New York University Press, New York.
- Presser, S., & Traugott, M. W. (1992). Little White Lies and Social-Science Models - Correlated Response Errors in a Panel Study of Voting. *Public Opinion Quarterly*, 56(1), 77-86.

Przeworski, A. and Teune, H. (1970). *The logic of comparative social inquiry*, Wiley-Interscience, New York.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal Modeling Using Stata* (2nd ed.). College Station (Texas). Stata Press.

Rasbash, J., W. Browne, et al. (2000). *A user's guide to MLwiN* London, UK, Center for Multilevel Modelling - University of London Institute of Education

Riepl, B. and Wintersberger, H. (1999). *Political Participation of youth below voting age*. Eurosocial Reports Vol.66, Vienna: European Center for Social Welfare Policy and Research.

Sapiro, V. (2004). "Not Your Parents' Political Socialization: Introduction for a New Generation." *Annual Reviews*(7). 24.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken (NJ). Wiley.

Satori, G. (1970). *Concept Misinformation and Comparative Politics*. *American Political Science Review*, 64, 1033 - 41.

Satori, G. (1991). *Comparing and Miscomparing*. *Journal of Theoretical Politics*, 3, 243 - 57.

Scherr, A. (2008). *Sozialisation, Person, Individuum. Einführung in Hauptbegriffe der Soziologie*. B. Schäfers. Wiesbaden, VS Verlag für Sozialwissenschaft: 45-68.

Schwarzer, S. and D. Connor (forthcoming 2012). *Political engagement among the youth - effects of political socialization across Europe*. *Democracy in Transition*. Political Participation in the European Union. Kyriakos N. D., Springer.

Schwarzer, S. and E. Zeglovits (forthcoming 2012). *If you don't come across it in your family or in school, where else? - How 16-and 17-year old Austrian first time voters perceive the role of schools in preparing them for the election*. *New Directions of Political Socialization Research: Old and New Challenges*. S. Abendschoen. Colchester, ECPR Press.

Searing, D., J. Schwartz, et al. (1973). *The structuring principle: Political socialisation and belief systems*. *American Political Science Review* 67(2). 415 - 432.

Silver, B. D., Anderson, B. A., & Abramson, P. R. (1986). *Who overreports voting?* *American Political Science Review*, 80(2), 613-624.

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*, Sage Publishers, London.

Steenbergen, M. and B. S. Jones (2002). *Modeling Multilevel Data Structures*. *American Journal of Political Science* 46(1). 218 - 237.

Stockè, V. (2007). *Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting*. *International Journal of Public Opinion Research*, 19(2), 237-246

- Stockè, V., & Stark, T. (2007). Political involvement and memory failure as interdependent determinants of vote overreporting. *Applied Cognitive Psychology*, 21(2), 239-257.
- Tavernise, S. (2008). Protester and Policy Clash as Armenian Unrest Grows. *New York Times*, March 2.
- Tedin, K. L. (1980). Assessing Peer and Parental Influences on Political Attitudes. *American Journal of Political Science*(24). 13 - 54.
- Teune, H. (1975). Comparative Research, Experimental Design, and the Comparative Method. *Comparative Political Studies*, 8, 195-200.
- Teune, H. (1990). Comparing Countries: Lessons Learned. In: *Comparative Methodology: Theory and Practice in International Social Research* (Ed, Oyen, E.) SAGE Publications, London, pp. 38 - 62.
- Tiemann, G. (2012). The Nationalization of Political Parties and Party Systems in post-communist Eastern Europe. *Communist and Post-Communist Studies* 45(1). 77 - 89.
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Development Science* 6(4). 203-212.
- Torney-Purta, J. (2002). What adolescents know about citizenship and democracy. *Educational Leadership* 59(4). 45-50.
- Torney-Purta, J., B. Wilkenfeld, et al. (2008). How adolescents in twenty-seven countries understand, support, and practice human rights. *Journal of Social Issues* 64: 857 - 880.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2009). *The psychology of survey response* (10th ed.). Cambridge: Cambridge University Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking Sensitive Questions: The impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60(2), 275-304.
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859-883.
- Traugott, M. W., & Katosh, J. P. (1979). Response Validity in Surveys of Voting Behavior. *The Public Opinion Quarterly*, 43(3), 359-377.
- Uskul, L., Oyserman, D., Schwarz, N. (2010). Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey Response Process. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Harkness, J., Braun, M., Edwards, B, Johnson, T., Lyberg, L., Mohler, P., Pennel, B-E., Smith, T. Wiley 203 - 226
- Van de Vijver, F. J. R. (1998a). Towards a Theory of Bias and Equivalence. In: *Cross-Cultural Survey Equivalence*. ZUMA-Nachrichten. Spezial, Nr. 3 .(Ed, Harkness, J. A.) ZUMA - Zentrum für Umfragen, Methoden und Analysen, Mannheim, pp. 41 - 65.

Van de Vijver, F. J. R. (2003). Bias and Equivalence: Cross - Cultural Perspectives. In: Cross-Cultural Survey Methods (Eds, Harkness, J. A., Van de Vijver, F. J. R. and Mohler, P. P.) John Wiley & Sons, Hoboken, New Jersey, pp. 143 - 157.

Van Deth, J. W. (1998). Equivalence in comparative political research. In: Comparative Politics. The problem of equivalence. (Ed, Van Deth, J. W.) Routledge, London, New York, pp. 1 - 19.

Verba, S., K. L. Schlozman, et al. (1995). Voice and equality: Civic voluntarism in American politics. Cambridge, MA, Harvard University Press. Wagner, M., D. Johann, et al. (forthcoming 2012). "Voting at 16: Turnout and the Quality of Vote Choice." Electoral Studies.

Welsh, W. A. (1981). Survey Research and Public Attitudes in Eastern Europe and the Soviet Union. New York, Pergamon Press.

Whitmore, B. (2007). Armenia: Parliamentary Vote sets Scene for 2008 Presidential Race. Radio Free Europe / Radio Liberty May 11.

Zeglovits, E., & Kritzing, S. (2011). Reducing Overreporting in the Voter Turnout Question. Paper presented at the Fourth Conference of the European Survey Research Association (ESRA).