PRZEGLĄD SOCJOLOGICZNY
2021 70(1): 131–147
ISSN 0033-2356; e-ISSN 2450-9361
https://doi.org/10.26485/PS/2021/70.1/7

OLENA OLEKSIYENKO
Polish Academy of Sciences

THE APPLICATION OF DATA HARMONIZATION IN MINORITY STUDIES. THE CASE OF THE POLITICAL PARTICIPATION OF THE RUSSIAN SPEAKING POPULATION IN FORMER SOVIET STATES

Abstract

This article discusses the theoretical possibilities and practical implications of survey data recycling and survey data harmonization. Using the example of political participation (participation in demonstration) of the Russian-speaking population in former Soviet states, the article presents the procedure of key variable harmonization (minority status), the rules, and the procedures of creating a harmonization control variable, and the possibilities of using harmonized variables in substantive statistical analysis. The harmonization procedures described in this article can be used to study other rare events and other minority groups – studies that often struggle with small and insufficient samples.

Keywords: data harmonization, survey data recycling, secondary data, survey methodology minority studies, political participation

INTRODUCTION

In this article, I discuss the methodological possibilities of analyzing patterns of political participation, exemplified by participating in demonstrations across former Soviet countries, using harmonized data for both minority status and political participation. The suggested approach, with some modifications, can be used to analyze other minority groups, as well as other phenomena not limited

Dr, Institute of Philosophy and Sociology; e-mail: olena.oleksiyenko@gmail.com; https://orcid.org/0000-0003-3215-4816

to political participation. Data harmonization in the proposed case study helps to overcome the problem of small samples for both minority populations and studying informal political participation, considered a rare event.

The first part of the article discusses the methodological aspects of minority status harmonization, while the second part presents an example of substantive analysis using harmonized data.

The topic of the Russian or Russian-speaking population in former Soviet space became extremely political in 2014 due to the Russian Federation's annexation of Crimea under the premise of protecting "their" people from the Ukrainian nationalists. Participation in the ostensible referendum and almost unanimous vote to join the Russian Federation was a triumph of "Russkiy mir" [Laruelle 2015][1] and a sign of political loyalty to the external motherland [Brubacker 1996] rather than nationalizing the state of Ukraine, which ignored the request for cultural and linguistic autonomy of the Russian speaking citizens. Nevertheless, the question emerges: was the political participation of Russian speakers in Ukraine and other former Soviet states any different (lower engagement) since the collapse of the USSR in the early 1990s, and did this outcome really came as such a surprise?

To answer that question, and more generally, the question on how the Russian/ Russian-speaking population participates in politics of the former Soviet States, it is essential to identify the distinct features of the group under analysis, specify types of political participation available to this group (formal or electoral/informal or non-electoral political participation), and investigate the available data, which is often limited to one country and a specific topic (e.g., electoral participation). The vast majority of the research conducted in the former Soviet Union (FSU) utilized qualitative methods (interviews or focus groups), and therefore, the results cannot be generalized or extended to other populations.

A NEW METHODOLOGICAL APPROACH TO STUDYING THE POLITICAL PARTICIPATION OF MINORITY GROUPS: SURVEY DATA HARMONIZATION

Survey Data Recycling (SDR) is an approach that makes it possible to overcome the problem of the underrepresentation of some countries or periods [Tomescu-Dubrow, Słomczyński 2016; Słomczyński, Tomescu-Dubrow 2018]. This

[1] According to Laruelle [2015: 1] the concept "serves as a justification for what Russia considers to be its right to oversee the evolution of its neighbors, and sometimes for an interventionist policy. Secondly, its reasoning is for Russia to reconnect with its pre-Soviet and Soviet past through reconciliation with Russian diasporas abroad".

approach makes it possible to combine survey and non-survey data (e.g., country-level indicators) and build tailored datasets that ensure sufficient geographical and time coverage of the data. The key stage of survey data recycling is the ex-post data harmonization – the technique that ensures the comparability of surveys and collected measures [Granda, Blaszczyk 2016]. Harmonization refers to efforts to standardize either the input or output of the multinational, multiregional, or multicultural survey. Harmonization aims to combine data from different sources into one dataset and make it as comparable as possible [Granda, Wolf, Hadorn 2010; Granda, Blaszczyk 2016; Doiron et al. 2012; Winters, Netscher 2016].

Harmonization makes it possible to obtain new target variables from source items that were not meant to be comparable. In the ex-post data harmonization process, comparability is achieved through the conversion process during which items from different sources (e.g., different countries or surveys) are assessed and edited (recoded, rescaled) to be merged into new data items. As a result of this process, source variables from the original datasets are transformed into target variables, which are then combined into the harmonized dataset.

Data harmonization and survey data recycling are still new approaches in different fields, and they often lack standardized guidelines and coherent methodology, but certain guidelines can be learned from existing harmonization projects [Dubrow, Tomescu-Dubrow 2015; Wysmułek et al. 2015].

I will start discussing the data harmonization process conducted for this paper by identifying the crucial aspects of the required data. Wang and Strong [1996] cite the key features of data to be accessibility, interpretability, relevance, and accuracy. To be relevant for research purposes, two indicators – native language or the language spoken at home – as well as indicators of informal political engagement (participation in demonstrations) must be available. Additionally, the data must cover as many countries/country-years as possible from the geographical area and period of interest, i.e., the FSU after the dissolution of the USSR. From the methodological perspective, I aimed for accurate data, which was at least partially ensured by controlling the data quality at different stages of the survey life cycle. It is acknowledged in the academic survey research; hence, the data collected and processed by academic institutions rather than the commercial organizations were preferred. This also leads to the last aspect: the data must be interpretable and accessible for free, meaning it must have sufficient documentation in English or Russian.

The *SDR Masterbox* harmonized dataset, which is the outcome of the project "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective"

[Słomczyński et al. 2016], contains already harmonized data on political participation[2] in the FSU and covers almost all former Soviet republics with the exception of Turkmenistan. This dataset contains harmonized variables originally collected as part of academic surveys to ensure the representativeness of the samples. In addition to the quality of the original data, the reliability and validity of the data are ensured before the harmonization by a quality assessment of the source data and detailed documentation of the harmonization procedures [Kołczyńska, Schoene 2018; Oleksiyenko, Wysmułek, Vangeli 2018; Kołczyńska, Słomczynski 2018; Zieliński, Powałko, Kołczyńska 2018].

The final selection of the survey project is presented in Table 1. Based on the available documentation for harmonization procedures, additional waves of the World Values Survey and European Social Survey were added to the original SDR Masterbox dataset. It should be noted, however, that some data, namely the European Social Survey indicator of participation in demonstrations, is not comparable to other survey projects, which use a protest potential [Marsh 1974] indicator and therefore were excluded from the analysis. The final dataset contains data from 6 international surveys, 19 survey waves, and 72 national samples (survey-country-year level), with 103,399 respondents in total. This dataset covers 13 non-Russian former Soviet republics except for Turkmenistan within the timespan from 1993 to 2015.

Even though this dataset is the most suitable and complete for the intended analysis, there was a need to harmonize additional variables unavailable in the original dataset, e.g., the indicators of the language spoken by the respondents at home/their native language. In line with the methodology proposed by the SDR project, the indicators of minority status (linguistic identity) and other indicators this dataset lacked were harmonized and merged with the existing dataset.

The first step in any harmonization project is formulating a clear definition of the target harmonization concept. Cross-national comparisons of minority groups are rarely straightforward since the constitutive concept of "minority group" can be different in each state. The methodological literature proposes different approaches to increase the concept's comparability. The "absolutist" approach suggests that only one marker of minority status should be considered, e.g., citizenship or language. The obvious advantage of such a solution is conceptual

[2] A detailed report summarizing the harmonization of two protest potential variables, i.e., participation in demonstrations and signing petitions, can be found at Słomczyński et at. 2016. The author worked as a research assistant in the Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective project in 2014-2016.

clarity, but on the other hand, one can argue that the complexity of minority status cannot be precisely studied using only one indicator. An alternative approach is the "relativist" approach to comparing the minority groups across countries. This involves cross-classification of different identity markers to obtain a single, cross-nationally equivalent concept of "ethnic minority status" [Lambert 2005]. The difficulty with the "relativist" approach is the low availability of the same markers across all surveys.

TABLE 1. Time and country coverage of the international survey projects utilized in the final dataset

Survey project	Time span	Country coverage
Consolidation of Democracy in Central and Eastern Europe (wave 2) – CDCEE/2	1998–2001	BY, EE, LV, LT, UA
European Social Survey (round 2) ESS/2	2004–2005	EE, UA
European Social Survey (round 4) ESS/4	2009	EE, UA
European Social Survey (round 6) ESS/6	2012	EE
European Social Survey (round 7) ESS/7	2014–2015	EE
European Values Study (wave 3) EVS/3	1999	EE, LV
European Values Study (wave 4) EVS/4	2008	AM, AZ, BY, EE, GE, LT, LV, MD, UA
Life in Transition (wave 2) LITS/2	2010	AM, AZ, BY, EE, GE, KG, KZ, LT, LV, MD, TJ, UA, UZ
New Baltics Barometer (wave 1) NBB/1	1993	EE, LV, LT
New Baltics Barometer (wave 2) NBB/2	1995	EE, LV, LT
New Baltics Barometer (wave 3) NBB/3	1997	EE, LV, LT
New Baltics Barometer (wave 4) NBB/4	2000	EE, LV, LT
New Baltics Barometer (wave 5) NBB/5	2001	EE, LV, LT
New Baltics Barometer (wave 6) NBB/6	2004	EE, LV, LT
World Values Survey (wave 2) WVS/2	1996–1997	AM, AZ, BY, EE, GE, LT, LV, MD, UA
World Values Survey (wave 4) WVS/4	2002–2003	KG, MD
World Values Survey (wave 5) WVS/5	2008–2009	GE, MD, UA
World Values Survey (wave 6) WVS/6	2011–2014	AM, AZ, BY, EE, GE, KG, KZ, UA, UZ

Source: Own elaboration based on SDR Masterbox 2016.

When examining data to identify Russian minorities, there is no clear and simple solution. In the former Soviet republics, the Russian-speaking minority consists not of ethnic Russians *per se*, but specifically, people who speak Russian and who also have a social identity that is opposed to full integration [for a review of the Russian speaking minority formation in the former USSR see: Laitin 1995; Laitin 1998; Hagendoorn et al. 2001; Barrington et al. 2003; Kosmarskaya 2006; Pavlenko 2008; Hansen 2009]. Exposure to the Russian language can be considered a key aspect of ethnic socialization. As Hansen [2009] argues, individuals who grew up in the "ethnic" environment, i.e., those who spoke the minority language distinct from the titular language at home, feel a stronger attachment to their minority identity. Therefore, I concentrate on this marker as a key identifier of the Russian-speaking minority, in line with the "absolutist" approach.

Table 2 presents the various survey items concerning ethnic identity available in the survey projects selected for harmonization. All questions can be divided into three main categories: native language (Consolidation of Democracy in Central and Eastern Europe, New Baltics Barometer, and Life in Transition), the language of the interview (European Values Study), and the language used at home (the remaining surveys). In the *European Values Survey,* the only available information about the language the respondent preferred to use is the language of the interview, which could be the native language, the language used at home, or neither of these. In another project, the World Values Survey, both questions on the language of the interview and the language respondent speaks at home are available; hence, I checked the correlation between these two, and it proved to be high (r = 0.78). Based on this finding, I decided to include this variable into harmonization and mark it with the control additional control variable.

With the majority of international survey projects keeping a consistent methodological approach in defining indicators, this research can be extended to the new and forthcoming waves of the same projects, or new projects using the same indicators.

TABLE 2. Items concerning language: wording and control variables

	Item Wording	Type of question/control variable		
		At home	Native language	Language of the interview
Consolidation of Democracy in Central and Eastern Europe (CDCEE)	*In what language did/do you communicate with your mother?*		x	
EuropeanW Social Survey (ESS)	*What language or languages do you speak most often at home?*	x		
European Values Study (EVS)	*Language of the interview*			x
Life in Transition Survey (LITS)	*What is your mother tongue?*		x	
World Values Survey (WVS)	*What language do you normally speak at home?*	x		
New Baltics Barometer (NBB)	*What language did you speak at home when you were a child?*		x	

Source: Harmonized dataset based on SDR Masterbox 2016.

The key variable for the research is the indicator of the language the respondent used at home or declared to be his or her native language. Those respondents who mentioned Russian as their native or most frequently used language of everyday communication, but who also selected the language of the interview in the European Values Survey, were coded as 1, and if the respondent that they spoke a language other than Russian, they were coded as 0.

I do not distinguish between the titular population, Russian-speaking minority, and other minority groups. This topic can be seen as an avenue for future research. I argue that the Russian-speaking minority is a specific linguistic minority, unlike other minority groups which historically settled in the former Soviet Union. Therefore, the opposition of the Russian speakers vs. non-Russian speakers seems to be theoretically justified. An additional reason for not introducing other minority groups in the FSU was that, unlike Russian speakers, these groups are not exclusively language-based minorities, and comparisons across countries are not methodologically justified [Lambert 2005].

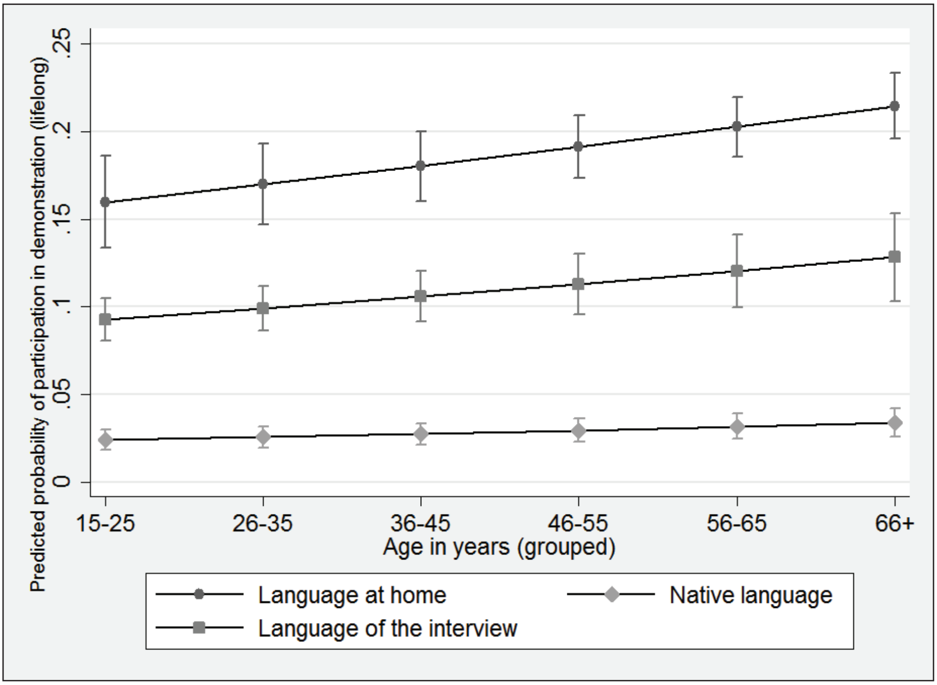
As there is no universal approach to constructing the indicator of belonging to a minority group, even choosing one distinctive feature does not solve the problem of data availability and conceptual equivalence of the measures used.

However, even this “simple” solution does not completely resolve the problem of conceptual equivalence between three concepts – the main or native language, the language spoken at home, and the language of the interview. Andreenkova [2019] analyzed the specifics of language use in the former Soviet republics and noted that native language, language used at home, and language used most often are not always the same for people living in the post-Soviet space. In many cases, native language, the researcher claims, is used more as an indicator of belonging to a group rather than proof of actual language proficiency, while the language used most often and the language used at home might indicate a certain level of language proficiency rather than an emotional attachment to the ancestry of the group. Understanding this additional dimension might also lead to a deeper understanding of both the identity of the Russian-speaking population and an understanding of their political choices. Moreover, as they are often bilingual or even multilingual, people of the former Soviet states might use different code-switching strategies as a political strategy [Heller 1992]. Bilingual respondents also tend to interpret survey questions and answer them differently, depending on the language of the interview [Peytcheva 2008].

Harmonizing survey data from different sources will often involve a series of trade-offs between increasing conceptual equivalence and increasing country and time coverage. To lessen such compromises as much as possible, I constructed a control variable that accounts for inter-project differences in question wording [Słomczyński et al. 2016]. The control variable indicating the type of question used to harmonize the Russian speaking indicator is a categorical variable, where Russian as the language used at home was coded as 0, Russian as the native language was coded as 1, and Russian as the language of the interview was coded as 2.

To emphasize the need for methodological control variables for harmonized data, I graphically predicted the probability of reporting participation in demonstrations at least once in the lifetime for the Russian speakers. Figure 1 shows that the predicted probability is lowest in the group of Russian speakers reporting Russian as their native language (symbolic attachment) and highest in the group reporting that Russian is the language they use in everyday communication (proficiency and symbolic attachment). On the other hand, the control variables also indicate the differences in the data quality of each survey project included in the harmonization procedure [Kołczyńska, Schoene 2018; Oleksiyenko, Wysmułek, Vangeli 2018; Kołczyńska, Słomczyński 2018; Zieliński, Powałko, Kołczyńska 2018].

FIGURE 1. Predicted probability of participation in demonstrations by age group depending on the type of harmonization control variable (N=13,570, sample restricted only to Russian speakers)



Source: Harmonized dataset, own elaboration based on the SDR Masterbox.

PRACTICAL IMPLICATIONS: USING HARMONIZED DATA IN SUBSTANTIVE ANALYSIS

Survey data recycling and survey data harmonization procedures described in this article made it possible to create a dataset suitable for the substantive analysis of participation in demonstrations among the Russian and non-Russian-speaking populations of the former Soviet states. The results of multilevel logistic regression using the harmonized data are presented in this section of the article.

Participation in informal politics can be explained by three broad types of factors: ideological, resources (e.g., education, occupation), and individual (biographical) availability, including marital status, employment status, and age [e.g., Corrigall-Brown 2012]. Demonstrations are also believed to be an urban phenomenon, so the place of residence is another variable included in the model explaining involvement in informal politics [e.g., Rüdig, Karyotis 2014].

There is an ongoing discussion on whether attitudinal variables, e.g., life satisfaction or political trust, can be treated as predictors of behaviors such as participating in demonstrations, signing petitions, or voting. Some studies suggest that there is an intent to behave in a certain way rather than the immediate behavior, which can be explained by the attitudinal variables [Fishbein, Ajzen 1974]. However, the two explanatory variables I use as control variables in the models of formal and informal political engagement proved to be significant predictors of such behaviors.

Life satisfaction in the research on political participation served both as dependent and independent variables [Pacheco, Lange 2010; Weitz-Shapiro, Winters 2008] since theoretical concepts can be of two kinds: either being politically active (efficient) causes life satisfaction, or the scale of life satisfaction influences the propensity to engage in political activity.

Political trust is believed to be associated both with the engagement in informal (protest) politics and electoral participation; however, the existing literature does not provide a coherent answer to the question of whether this relationship is positive or negative. Some studies claim that the relationship is negative, since citizens who are dissatisfied with the traditional representative channels of democracy are more likely to engage in additional informal political activities [e.g., Dalton 2006; Hooghe, Marien 2012], while other studies prove that this relationship is no longer relevant [Dubrow, Słomczyński, Tomescu-Dubrow 2008].

Minority participation gap [Leightley, Vedlitz 1999] theory provides an explanation for the lower level of political involvement of minority groups. In this article, belonging to the Russian-speaking minority is considered a key explanatory variable.

One must also consider Russian speakers and non-Russian speakers in the former Soviet states as nested in the specific opportunity structures captured (but not limited to) by the two indicators – level of democracy (Freedom House) and economic development (GDP per capita). Stockemer and Carbonetti [2010] note that democratic development is often associated with the changing mode of political participation or moving from elite-oriented (electoral) to elite-challenging (informal political participation). At the same time, economic well-being is generally associated with a higher interest in formal political participation (rewarding the government for financial stability), but it can also be rooted in economic grievances and the need to express dissatisfaction with participation in demonstrations.

Accounting for the nested structure (respondents nested in country-years), the hypothesis about lower odds for participation in demonstrations among the Russian-speaking population, controlling for individual and country-level characteristics, are tested using the multilevel mixed effect logistic regression models – model 1 (without the harmonization control variables) and model 2 (with the harmonization control variables). The dependent variable is operationalized as life-long protest potential, as all survey projects in the SDR project and the final harmonized variable are based on the question about life-long participation in demonstrations or the willingness to participate in one.

Model 1

Participation in demonstrations-EXP-logij = γ00 + γ1*Russian_speakerij + γ2*age1ij + γ3*age2ij + γ4*age3ij + γ5*age4ij + γ6*age5ij + γ7*femaleij + γ8*marriedij + γ9*tertiaryij + γ10*employed1ij + γ11*employed2ij + γ12*employed3ij + γ13*employed4ij + γ14*metropolitanij+ γ15*life satisfactionij+ γ16*trust in parliamentij u0j+γ01*FH + γ02*GDP

Model 2

Participation in demonstrations-EXP-logij = γ00 + {γ1*Russian language control_type of questionij}+ γ2*Russian_speakerij + γ3*age1ij + γ4*age2ij + γ5*age3ij + γ6*age4ij + γ7*age5ij + γ8*femaleij + γ9*marriedij + γ10*tertiaryij + γ11*employed1ij + γ12*employed2ij + γ13*employed3ij + γ14*employed4ij + γ15*metropolitanij+ γ16*life satisfactionij+ γ17*trust in parliamentij u0j+γ01*FH + γ02*GDP

{} – harmonization control variables

Table 3 presents the result of the mixed effect logistic regressions for Model 1 and Model 2. Based on the results of Model 2, being a member of a Russian-speaking community decreases the odds of declared life-long protest potential (the potential to participate in demonstrations) by 10%. The other explanatory variables, apart from marital status, such as age, tertiary education, being in employment, residing in a metropolitan area, but also trust in parliament and life satisfaction, proved to be significant predictors of participation in demonstrations.

The results also reveal the significant negative effect of both GDP per capita and the Freedom House civil liberties indicator, which means that countries with lower GDP and higher civil liberties scores (the lower the score, the higher the reported level of civil liberties) create the opportunity for higher potential participation in demonstrations. These results, although not directly related to political participation of the Russian-speaking minority, show that in countries in transition, such as former Soviet states, participation in demonstrations might be motivated by economic hardship or people flourishing in a more democratic environment (freedom of gathering, freedom of speech, etc.).

TABLE 3. Results of the mixed effect logistic regression for participation in demonstrations: Model 1 (no harmonization control variables) and Model 2 (with harmonization control variables)

Mixed-effects logistic regression: Lifelong participation in demonstrations	Model 1			Model 2 (with harmonization control variables)		
	Coefficient	SE	Odds ratio	Coefficient	SE	Odds ratio
Harmonization control variables ref. Russian language as language at home						
Russian language as the native language	x	x	x	–1.052***	0.225	0.349
Russian language as the interview language	x	x	x	0.035	0.232	1.036
Background characteristics (individual level)						
Russian speaker (dichotomous)	–0.104**	0.040	0.901	–0.106**	0.040	0.899
Age in years *ref. 19–25*						
26–35	0.139*	0.059	1.149	0.142*	0.059	1.152
36–45	0.391***	0.059	1.479	0.394***	0.059	1.482
46–55	0.556***	0.060	1.744	0.558***	0.060	1.748
56–65	0.690***	0.068	1.994	0.693***	0.068	2.000
66+	0.648***	0.080	1.912	0.652***	0.080	1.919
Male	–0.202***	0.031	0.817	–0.202***	0.031	0.817
Married (dichotomous)	0.017	0.034	1.017	0.010	0.034	1.010
Tertiary education (dichotomous)	0.414***	0.035	1.513	0.415***	0.035	1.514
Employment status *ref. employed*						
Retired	–0.253***	0.059	0.777	–0.253***	0.059	0.776
Inactive	–0.346***	0.061	0.707	–0.344***	0.061	0.709
Student	–0.056	0.084	0.946	–0.057	0.084	0.944
Unemployed	–0.208***	0.050	0.813	–0.203***	0.050	0.816
Metropolitan area	0.283***	0.034	1.327	0.284***	0.034	1.328
Life satisfaction (11 points scale)	0.015*	0.007	1.016	0.015*	0.007	1.015
Trust in parliament (11 points scale)	–0.031***	0.007	0.969	–0.031***	0.007	0.969
Country-level predictor						
Freedom House Civil liberties	–0.288***	0.081	0.750	–0.227***	0.062	0.797
Log GDP per capita	–0.615***	0.162	0.541	–0.462**	0.135	0.630
Cons.	3.811	1.570		2.580	1.247	
Random effect parameters estimate	0.709	0.085		0.526	0.065	
Log likelihood	–15684.282			–15673.072		
N=Level1	46.220			46.220		
N=Level2	39			39		
LR test vs. logistic model	chibar2(01) = 1287.89; Prob >= chibar2 = 0.0000			chibar2(01) = 808.66; Prob >= chibar2 = 0.0000		

*** Significant at P ≤ 0.001; ** Significant at P ≤ 0.01; * Significant at P ≤ 0.05

Source: Harmonized dataset, own elaboration based on the SDR Masterbox.

Adding the harmonization control variable to the model (categorical variable identifying source question type) does not change the significance of any explanatory variables; however, the significance of the harmonization control variable cannot be ignored. The coefficient of the Russian as the native language control variable is statistically significant and negative compared to Russian as the language used at home. On the one hand, methodological control variable, in this case, adds an additional layer to the analysis, stressing that despite being closely related, different types of linguistic identity (native vs. the language of the everyday communication) might have different implications for informal political activity when respondents from the broadly defined category of the Russian-speakers are considered. On the other hand, as cited before, the methodological control variable might be a reflection of differences in the quality (e.g., sampling, quality of data processing) of the harmonized survey projects and point to those differences rather than substantive ones in the analyzed population. If this approach is taken, one can view the results as an indication of certain differences between a survey that uses the indicator of language used at home vs. a survey that uses the indicator of native language. The differences, in this case, are not limited to this question or to the substantive focus of the analysis, but the overall methodological accuracy of the survey procedures.

DISCUSSION

This article explores the possibilities of using quantitative data from international survey projects to analyze rare events and small groups. This study can be extrapolated to a variety of research questions.

Despite presenting certain challenges, the proposed methodology, i.e., survey data recycling and survey data harmonization, opens new possibilities for conducting statistical analysis of topics that were previously rarely studied using these techniques due to limited data availability. The topic of political participation of the Russian-speaking minority in the former USSR was rarely studied using quantitative and comparative methodologies due to limited data availability, among other reasons. By introducing data harmonization, more similar topics can be looked at through a new lens.

Before starting a harmonization project, it is important to formulate clear definitions for each variable to be used in the analysis. As in the case study presented in this article, a clear definition of the minority group under analysis was the key to creating an indicator and conducting the subsequent harmonization process. The same is true for the dependent variable – despite being available in the European Social Survey, it was excluded from further analysis due to the different underlying concept it measures – participating in a demonstration in the last 12 months vs. protest potential used in the remaining projects.

The example of the multilevel model using harmonized data and harmonization control variables shows how certain limitations in data availability can be solved, preserving the conceptual equivalence of the target variable despite differences in the source variables.

The harmonization of minority status, even in a simple one-component indicator, as mentioned in the article, can be long and involve certain trade-offs, but it is no reason to be discouraged since the advantages of this approach outweigh the efforts.

The key element of each endeavor of this kind should be detailed documentation of every step of the process. All variable harmonization reports for the data presented in this paper are available upon request.

BIBLIOGRAPHY

Andreenkova Anna. 2018. How to choose interview language in different countries. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 293–324. Hoboken: John Wiley & Sons.

Barrington Lowell W., **Erik S. Herron**, **Brian D. Silver**. 2003. "The Motherland is calling. Views of Homeland among Russians in the Near Abroad". *World Politics* 55: 290–313.

Corrigall-Brown Catherine. 2012. "From the balconies to the barricades, and back? Trajectories of participation in contentious politics". *Journal of Civil Society.* 8: 17–38.

Dalton Russel J. 2006. *Citizen politics: Public opinion and political parties in advanced industrial democracies, 4th ed.* Washington, DC: CQ Press.

Doiron Dany, **Parminder Raina**, **Vincent Ferretti**, **François L'Heureux**, **Isabel Fortier**. 2012. "Facilitating collaborative research: Implementing a platform supporting data harmonization and pooling". *Norsk Epidemiologi* 21: 221–224.

Dubrow Joshua K., **Irina Tomescu-Dubrow**. 2016. "The rise of cross-national survey data harmonization in the social sciences: Emergence of an interdisciplinary methodological field". *Qual Quant* 50:1449–1467.

Dubrow Joshua K., **Kazimierz M. Słomczyński**, **Irina Tomescu-Dubrow**. 2008. "Effects of democracy and inequality on soft political-protest in Europe: Exploring the European social survey data". *International Journal of Sociology* 38(3): 36–51.

Fishbein Martin, **Icek Ajzen**. 1974. "Attitudes towards objects as predictors of single and multiple behavioral criteria". *Psychological Review* 81(1): 59–74.

Granda Peter, **Emily Blaszczyk**. 2016. Data harmonization guidelines. Cross-cultural survey guidelines. https://ccsg.isr.umich.edu/chapters/data-harmonization [access: 14.02.2021].

Granda Peter, **Christof Wolf**, **Reto Hadorn**. 2010. Harmonizing survey data. In: *Survey methods in multinational, multiregional, and multicultural contexts*, J.A Harkness, M. Braun, B. Edwards, T.P. Johnson, L. Lyberg, P. Mohler, B-E. Pennell, T.W. Smith (eds.), 315–334. New York: Wiley.

Hagendoorn Louk, **Hub Linssen**, **Sergei Tumanov (eds.)**. 2001. *Intergroup relations in states of the former Soviet Union: The perception of Russians*. Abingdon-on-Thames: Taylor and Francis.

Hansen Holley E. 2009. Ethnic voting and representation: Minority Russians in post-Soviet states. Iowa Research Online. https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1560&context=etd [access: 14.02.2021].

Heller Monica. 1992. "The politics of code-switching and language choice". *Journal of multilingual and multicultural development* 13: 123–142.

Hooghe Marc, **Sofie Marien**. 2013. "A comparative analysis of the relation between political trust and forms of political participation in Europe". *European Societies* 15(1): 131–152.

Kołczyńska Marta, **Kazimierz M. Słomczyński**. 2018. Item metadata as controls for ex post harmonization of international survey projects. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 1011–1033. Hoboken: John Wiley & Sons.

Kołczyńska Marta, **Matthew Schoene**. 2018. Survey data harmonization and the quality of data documentation in cross-national surveys. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 963–984. Hoboken: John Wiley & Sons.

Kolstø Pal. 1995. *Russians in the former Soviet republics*. London/Bloomington: Christopher Hurst/Indiana University Press.

Kolstø Pal. 2011. "Beyond Russia, becoming local: Trajectories of adaption to the fall of the Soviet Union among ethnic Russians in the former Soviet Republics". *Journal of Eurasian Studies* 2(2): 153–163.

Kosmarskaya Natalia. 2006. *Дети Империи в постсоветской Центральной Азии. Адаптивные практики и ментальные сдвиги (русские в Киргизии, 1992–2002) [Children of the "Empire" in post-soviet Central Asia: Mental shift and practices of adaptation (Russians in Kirgizia, 1992–2002)]*. Moscow: Natalis Press.

Laitin David. **D.** 1995. "Identity in formation: The Russian-speaking nationality in the post-Soviet diaspora". *European Journal of Sociology/Archives Européennes De Sociologie/ Europäisches Archiv Für Soziologie* 36(2): 281–316.

Laitin David. **D.** 1998. *Identity in formation: The Russian-speaking populations in the near abroad*. Ithaca: Cornell University Press.

Lambert Paul. **S.** 2005. Ethnicity and the comparative analysis of contemporary survey data. In: *Methodological aspects in cross-national research*, Hoffmeyer-Zlotnik J.H.P., J. Harkness (eds.), 259–278. Mannheim: GESIS-ZUMA.

Laruelle Marlene. 2015. The "Russian World". Russian's soft power and geopolitical imagination. The Center on Global Interests (CGI). http://globalinterests.org/wp-content/uploads/2015/05/FINAL-CGI_Russian-World_Marlene-Laruelle.pdf [access: 14.02.2021].

Leighley J.E., **Vedlitz A.** 1999. "Race, ethnicity, and political participation: Competing models and contrasting explanations. *Journal of Politics* 61(4): 1092–1114.

Marsh Alan. 1974. "Explorations in unorthodox political behavior: A scale to measure protest potential. European". *Journal of Political Research* 2*:* 107–29.

Oleksiyenko Olena, **Wysmulek Ilona Vangeli Anastas**. 2018. Identification of processing errors in cross-national surveys. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 985–1010. Hoboken: John Wiley & Sons.

Pacheco Gail, **Thomas Lange**. 2010. “Political participation and life satisfaction: A cross-European analysis”. *International Journal of Social Economics* 37(9): 686–702.

Peytcheva Emilia. 2008. *Language of administration as a source of measurement error: Implications for surveys of immigrants and cross-cultural survey research*. Ann Arbor: University of Michigan.

Rüdig Wolfgang, **Georgios Karyotis**. 2014. “Who protests in Greece? Mass opposition to austerity”. *British Journal of Political Science* 44(3): 487–513.

Słomczyński Kazimierz M., **Irina Tomescu-Dubrow**. 2006. “Representation of European post--communist countries in cross-national public opinion surveys”. *Problems of Post-Communism* 53(4): 42–52.

Słomczyński Kazimierz M., **Irina Tomescu-Dubrow**, **Craig J. Jenkins**, **with Marta Kołczyńska**, **Przemek Powałko**, **Ilona Wysmułek**, **Olena Oleksiyenko**, **Marcin W. Zieliński**, **Joshua K. Dubrow**. 2016. *Democratic values and protest behavior. Harmonization of data from international survey projects.* Warsaw: IFiS Publishers.

Słomczyński Kazimierz M., **Irina Tomescu-Dubrow**. 2018. Basic principles of survey data recycling. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 937–962. Hoboken: John Wiley & Sons.

Stockemer Daniel, **Benjamin Carbonetti**. 2010. “Why do richer democracies survive? The non-effect of unconventional political participation”. *Social Science Journal* 47(2): 237–251.

Wang Richard Y., **Diane M. Strong**. 1996. “Beyond accuracy: What data quality means to data consumers”. *Journal of Management Information Systems* 12(4): 5–33.

Weitz-Shapiro Rebecca, **Matthew S. Winters**. 2008. “Political participation and quality of life”. *Working Paper* 638*, Inter-American Development Bank, Research Department, Washington, DC.*

Winters Kristi, **Sebastian Netscher**. 2016. Proposed standards for variable harmonization documentation and referencing: A case study using QuickCharmStats 1.1. PLoS ONE 11(2): e0147795. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147795 [access:14.02.2021].

Wysmułek Ilona, **Olena Oleksiyenko**, **Przemek Powałko**, **Marcin W. Zieliński**, **Kazimierz M. Słomczyński**. 2015. Towards standardization: Target variable report template in the harmonization project. In: *Harmonization: Newsletter on survey data harmonization in the social sciences*, I. Tomescu-Dubrow, J.K. Dubrow (eds.), 1(2): 13–17. https://www.asc.ohio-state.edu/dataharmonization/wp-content/uploads/2015/11/harmonization-newsletter-v1n2-fall-2015-with-issn-_final-3.pdf [access: 14.02.2021].

Zieliński Marcin W., **Przemek Powałko**, **Marta Kołczyńska**. 2018. The past, present, and future of statistical weights in international survey projects: Implications for survey data harmonization. In: *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, T.P. Johnson, B-E. Pennell, I. Stoop, B. Dorer (eds.), 1035–1052. Hoboken: John Wiley & Sons.

Database: **Słomczyński Kazimierz M.**, **Craig J. Jenkins**, **Irina Tomescu-Dubrow**, **Marta Kołczyńska**, **Ilona Wysmułek**, **Olena Oleksiyenko**, **Przemek Powałko**, **Marcin W. Zieliński. 2017**. SDR 1.0 Master Box, https://doi.org/10.7910/DVN/VWGF5Q, Harvard Dataverse, V1, UNF:6:HIWud4wueVRsU8wTN+lySg== [fileUNF]

Olena Oleksiyenko

MOŻLIWOŚCI ZASTOSOWANIA HARMONIZACJI DANYCH W BADANIACH MNIEJSZOŚCI. PRZYKŁAD PARTYCYPACJI POLITYCZNEJ MNIEJSZOŚCI ROSYJSKOJĘZYCZNEJ W KRAJACH BYŁEGO ZWIĄZKU RADZIECKIEGO

Streszczenie

Artykuł przedstawia teoretyczne możliwości i praktyczne zastosowania recyklingu danych sondażowych i ich harmonizacji. Na przykładzie partycypacji politycznej (udziału w demonstracji) ludności rosyjskojęzycznej w krajach byłego Związku Radzieckiego w artykule przedstawiono procedurę harmonizacji kluczowej zmiennej (status mniejszości), zasady i procedury tworzenia zmiennych kontrolnych (kontrola procedur harmonizacyjnych) oraz możliwości praktycznego wykorzystania zmiennych zharmonizowanych do analiz statystycznych. Procedury harmonizacji opisane w tym artykule mogą znaleźć zastosowanie w badaniach zarówno innych rzadkich zjawisk, jak i innych grup mniejszościowych, czyli w badaniach, które często borykają się z problemem małych prób.

Słowa kluczowe: harmonizacja danych, recykling danych sondażowych, dane wtórne, metodologia badań sondażowych, badania mniejszości, partycypacja polityczna