Survey Research Methods (2022)
Vol. 16, No. 1, pp. 79-96
doi:10.18148/srm/2022.v16i1.7909
European Survey Research Association

Accounting for cross-country-cross-time variations in measurement invariance testing. A case of political participation

Piotr Koc
Institute of Philosophy and Sociology
Polish Academy of Sciences
Warsaw, Poland

Artur Pokropek
Institute of Philosophy and Sociology
Polish Academy of Sciences
Warsaw, Poland

In many works involving measurement invariance testing, researchers concentrate on one type of grouping only, such as countries, even when the comparisons they make involve multiple types of grouping, such as countries and years. In this article, we propose a procedure allowing to incorporate more than one type of grouping into the invariance testing. For that, we use the example of political participation which is often studied in a comparative perspective where both countries and years are considered. The results show that the comparability of levels of political participation in Europe over the last 20 years is limited. With a simulation study, we show that one remedy for this could be alignment optimization which produces more accurate estimates of means and standard errors. Furthermore, we demonstrate that ignoring the non-invariance can change our substantive conclusions regarding the aggregated trends of participation.

Keywords: measurement invariance; political participation; measurement equivalence; alignment optimization; civic participation

1 Introduction

Researchers have become increasingly aware that measurement invariance should not be assumed but can and ought to be tested (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014; Poortinga & Malpass, 1986). Only after checking if measurement invariance holds, we can claim that "under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992, p. 117). As methodologists stress, lack of measurement invariance may preclude us from drawing valid inferences about the differences across groups (Stegmueller, 2011; van de Vijver & Tanzer, 2004). That is, we cannot be sure to what extent varying regression coefficients or scale means mirror "true" differences, and to what extent the observed differences are due to measurement biases (Boer, Hanke, & He, 2018; Cieciuch, Davidov, Schmidt, & Algesheimer, 2019).

Yet, while conducting measurement invariance testing, researchers usually concentrate on one type of grouping, such as countries, even when the comparisons they make involve multiple types of grouping, such as countries and years. Scholars do so because accounting for more than one type of grouping becomes more complicated and there is very little guidance as to how exactly one should proceed in such situations.

In this article, we propose a novel approach to investigate multiple types of grouping in measurement invariance testing, concentrating on two variables: country and year. The approach we present allows us to distinguish different sources of non-invariance and subsets for which we can make comparisons across both time and countries. Additionally, with a simulation study, we demonstrate what can be gained by using alignment optimization in the context of cross-country-cross-time comparisons when the classical method relying on the Multiple Group Confirmatory Factor Analysis (MG-CFA) fails to reach the scalar level of invariance.

As a substantive example, we have chosen political participation as measured in the European Social Survey (ESS). We have done so for two reasons. First, in plenty of works where researchers use the ESS set of participatory items in a comparative context (e. g. Bäck & Christensen, 2016; Hooghe & Marien, 2013; Kostelka, 2014; Marien, Hooghe, & Quintelier, 2010), we will not find any test of measurement invariance. Second, in the few works on the measurement invariance of political participation that exist (García-Albacete, 2014; Goroshit, 2016), only the cross-country invariance is tested. What is missing in the literature is a test, in which both spatial (cross-country) and temporal dimensions would be taken into consideration. This would enable checking not only whether we could compare levels of participation across countries in a given time point, but also if

Contact information: Piotr Koc, Nowy Świat 72, 00-330 Warsaw, Poland (E-mail: piotr.koc@yahoo.com)

we could compare the levels from the same country across time and from different countries across time.

The results show that the political participation scale, measured by the ESS battery, does not fulfill the classical measurement invariance assumptions that would allow us to compare levels of participation across countries, across time points, and to make cross-country-cross-time comparisons. These cannot be done if we use the classical MG-CFA model that assumes full invariance of item parameters (scalar model). With a simulation study, we demonstrate that whereas the MG-CFA model significantly underestimates the standard errors of country-year means, the estimates of means and standard errors produced by the alignment are more reliable and accurate. Consequently, conclusions about the levels of participation are less likely to be influenced by the imprecise measurement.

2 The Scale of Political Participation

Due to the rapid increase of the number of participatory forms, van Deth (2014) proposes a different strategy to the conceptualization of political participation. He argues that necessary and sufficient conditions for political participation should be worked out in a stepwise manner instead of trying to derive the concept from a theory (Hooghe, Hosch-Dayican, & van Deth, 2014). Following this proposition, we would start with the minimalist definition, where political participation is defined as an activity or action that is carried out voluntarily by citizens, and which is located within the realm of politics, government, or the state. Additionally, actions targeted at a subject from one of the realms and actions aiming at solving a community problem could be also included in the conceptualization. Lastly, when the target remains unclear, the researcher would investigate the political context and the motives of the participants (van Deth & Theocharis, 2018). In this framework, one could capture all forms of political participation, both traditional, like voting, and new ones, like online participation.

Quite often, though, researchers are interested in a broader picture of how citizens participate in politics, not only in a particular act of participation. The broader picture could be captured by treating political participation as a latent quantity that forms a scale composed of many participatory forms (e. g. García-Albacete, 2014; Marsh, 1974, 1977; Milbrath, 1965; Quaranta, 2013; van Deth, 1986; Verba, Nie, & Kim, 1978; Vráblíková, 2014). Any valid scale must reflect the changes that unfolded at the beginning of the 21st century in how citizens choose to participate. Nowadays, people have multiple participatory forms at their disposal, and they opt for acts that allow them to express themselves, which match their interests and resources (Bang, 2009; van Deth & Theocharis, 2018). Politically skilled citizens use whatever means appropriate to influence the policies they care about (Dalton, 2008). In doing so, they combine various forms of participation (Norris, Walgrave, & van Aelst, 2005; van Aelst & Walgrave, 2001). As a result, acts like taking part in demonstrations, buycotting and boycotting, which for a long time deemed to be unconventional, have gradually become mainstream today (Norris, 2007). Given all this, it makes sense to treat political participation as a complex phenomenon where different behavioral indicators constitute one scale. Such a view is also supported by psychometric studies in which the ESS battery of participatory items was tested (e. g. Koc, 2021).

3 Data

In ESS, respondents were asked to answer several questions concerning political participation. A full list of indicators can be found in Table 1. We have to leave out the first edition of the ESS (2002) because the list of items and their wording differ substantially from the items in the subsequent editions. Moreover, we concentrate here on non-electoral participation. Many studies (e. g. Marien et al., 2010; Parry, Moyser, & Day, 1992) have shown a special character of voting: It is a very frequent activity and therefore an exception on other indicators After removing voting, we end up with seven indicators of political participation which form a unidimensional scale.

The items which measure political participation were recoded into dichotomous variables, with values 1 “participation” and 0 “no participation”. Refusals, “No answer”, and “Don’t know” answers were recoded as missing values. Besides, all interviewees-minors were excluded from the analysis as they could not participate in some participatory forms for legal reasons. Also, we used a reduced number of countries, i. e. countries that took part in all rounds 2-9 of the ESS[1]. Sample sizes for each country-round can be found in Table A1 in the Appendix.

4 Methods

4.1 Classical Measurement Invariance Modeling

Measurement invariance modeling is embedded in the latent variable modeling framework and in most applications relies on Multiple Group Confirmatory Factor Analysis (MG-CFA) models and their extensions. MG-CFA model is defined as

$$y_{ipg} = \nu_{ig} + \lambda_{ig}\eta_{pg} + \epsilon_{ipg}, \quad \eta_{pg} \sim \mathcal{N}\left(\alpha_g, \psi_g^2\right), \tag{1}$$

[1]Belgium, Switzerland, Germany, Estonia, Spain, Finland, France, the United Kingdom, Hungary, Ireland, the Netherlands, Norway, Poland, Portugal, Sweden. We had to exclude Slovenia because of the missing values on the item “working in another organization or association” (this variable has been omitted from the waves 2-7 in the integrated file due to a translation error) (ESS, 2018).

Table 1
Item labels and items measuring non-electoral participation in ESS

Item label	Item
contplt	Contacted a politician or government official in the last 12 months
wrkprty	Worked in a political party or action group in the last 12 months
wrkorg	Worked in another organization or association in the last 12 months
badge	Worn or displayed a campaign badge/sticker in the last 12 months
sgnptit	Signed a petition in the last 12 months
pbldmn	Taken part in a lawful public demonstration in the last 12 months
bctprd	Boycotted certain products in the last 12 months

where $i = 1, \ldots, I$ denotes the item index, p the person index, $g = 1, \ldots, G$ the group index, y_{ipg} is a response to the item, ν_{ig} and λ_{ig} are the item parameters, factor intercept and loading, respectively, ϵ_{ipg} is a normally distributed error term with $\epsilon_{ipg} \sim \mathcal{N}(0, \theta^2_{ig})$, η_{pg} is a factor reflecting the latent trait which is assumed to be normally distributed in each group. This multiple group CFA model is not identified because not all item loadings and item intercepts can be simultaneously estimated along with group means $\alpha = (\alpha_g)$ and standard deviations $\psi = (\psi_g)$. Usually, the problem of identification is solved by either fixing one item loading in each group to a constant or by fixing latent mean and standard deviation for one group to 0 and 1, respectively.

MG-CFA models initially were formulated for continuous variables (Jöreskog, 1971). However, as discussed by B. Muthén and Asparouhov (2014), Equation 1 is relevant also for binary outcomes, where the dependent variable in (1) is a continuous latent response y^*_{ipg} underlying the observed binary variable y_{ipg}, and τ is a threshold parameter:

$$y_{ipg} = \begin{cases} 0, & if \quad y^*_{ipg} \leq \tau_{ig} \\ 1, & if \quad y^*_{ipg} > \tau_{ig} \end{cases} \tag{2}$$

The variance of the residual ϵ_{ipg} is standardized as $\pi^2/3$ in line with the logistic model. As such, the CFA model for binary data is equivalent to the Item Response Theory (IRT) two-parameter logistic (2PL) model (B. Muthén & Asparouhov, 2014). In the paper, we refer to the IRT model but the CFA for binary data is a different name for the same model. Methodologists distinguish several levels of measurement invariance defined by increasingly restrictive equality constraints on parameters of interest across groups in MG-CFA models (Byrne, Shavelson, & Muthén, 1989; Meredith, 1993). The most fundamental is the configural level of invariance which assumes the same dimensional structure but different item parameters. It confirms that measured concepts are similar but does not assure direct comparisons. The second level of invariance—metric (or week)—is defined by item loadings being constrained to be equal across groups. Achieving metric invariance allows for comparisons of parameters that do not rely on the levels of latent means (e. g. correlations and regression coefficients). For valid mean comparisons, scalar (strong) measurement invariance is required. Scalar invariance enables full comparability and requires that both factor loadings and the intercepts for each item used in the construction of a scale be the same across groups (in our case countries and survey years). Levels of measurement invariance are established by examining the change in the model fit indices between these three types of models in a step-wise manner, from the least to the most restrictive one (Cheung & Rensvold, 2002; Schaffer & Riordan, 2003).

Unfortunately, the assumption about scalar invariance rarely holds in practice making it impossible for an investigator to compare means of measured constructs in a valid way (Davidov et al., 2014). It is because in cross-national surveys we usually deal with many countries, that are heterogenous to various extents. Similarly, surveys that aim at monitoring cross-time differences, tend to have problems with scalar invariance due to global time-specific effects that affect all the units simultaneously. Due to these problems, in recent years, the concept of partial approximate measurement invariance (PAMI) has gained considerable attention (B. Muthén & Asparouhov, 2012, 2013). The PAMI postulates that the estimation of reliable and comparable parameters for groups in multiple-group models is possible as long as the differences between parameters are small (approximate invariance) and/or the large differences are limited to a small number of items (partial measurement invariance). One way of examining the PAMI is to employ the alignment approach, which we will pursue in this article.

4.2 Partial approximate measurement invariance with alignment

Asparouhov and Muthén (2014) describe the alignment approach as a procedure that aligns item parameters from group-specific configural CFA or IRT models (B. Muthén & Asparouhov, 2014) into the most optimal (according to the chosen loss function) invariance pattern that enables the estimation of group-specific factor means and variances without requiring the exact measurement invariance.

The alignment procedure solves the identification issue by determining α and ψ in a way that the amount of measurement non-invariance is minimized. This is done by the means of the alignment function that optimally aligns group-specific item parameters. The alignment procedure consists of two steps. In the first step, configural measurement models are estimated for each group. Those models might be CFA models for continuous indicators or IRT models for categorical indicators. The configural model for each group is identified by setting the means to zero and the standard deviations to one while all item parameters are estimated freely in each group. This results in group-wise estimated item loadings $\hat{\lambda}_{ig,0}$ and item intercepts $\hat{\nu}_{ig,0}$.

Asparouhov and Muthén (2014) show that the group-specific parameters (λ_{ig}, ν_{ig}) could be described as a function of item parameters from the configural model $(\lambda_{ig,0}, \nu_{ig,0})$ and the group-specific means α_g and standard deviations ψ_g in the following way:

$$\hat{\lambda}_{ig} = \frac{\hat{\lambda}_{ig,0}}{\psi_g} \quad \text{and} \quad \hat{\nu}_{ig} = \hat{\nu}_{ig,0} - \frac{\hat{\lambda}_{ig,0}}{\psi_g}\alpha_g \tag{3}$$

With these relations established, in the second step, the alignment algorithm searches for aligned means, α, and standard deviations, ψ, by minimizing the alignment optimization function, F.

The rationale for the definition of F is to minimize overall deviations between item parameters from different groups in a way that it results in a few large noninvariant parameters and many approximately invariant parameters. (B. Muthén & Asparouhov, 2014) proposed the same loss function for slopes and intercepts:

$$f_\lambda(x) = f_\nu(x) = \sqrt{|x|}, \tag{4}$$

where x is the difference between item parameters. Further studies showed that this function could be generalized (Pokropek, Lüdtke, & Robitzsch, 2020; Robitzsch, 2020) but in this study, we use the original loss function proposed by B. Muthén and Asparouhov (2014). The alignment procedure penalizes differences in item intercepts and item slopes between groups and, hence, minimizes the extent of measurement non-invariance according to the chosen loss function.

To sum up, the alignment procedure replaces the cross-group equality constraints with a technique similar to the rotation in exploratory factor analysis. An algorithm estimates a solution that minimizes overall differences between groups' parameters using a simplicity function. The simplicity function is optimized at a few large noninvariant parameters and many approximately invariant parameters. This, however, does not mean that alignment optimization will always provide correct inference for the group parameters. It will give the best possible solution given the selected loss function and the data under investigation. If indicators are not invariant or the pattern of non-invariance is different from that assumed by the particular form of the loss function, estimated group parameters might be biased (for examples, see Pokropek et al., 2020).

5 Analytical Strategy

The majority of measurement invariance analyses focus on single-type comparisons, such as countries. Researchers do so because the analysis becomes more complicated when multiple types of grouping are taken into account. However, for many cyclical cross-country surveys like the ESS or World Values Survey, where multiple countries are surveyed multiple times, having more than one type of grouping is a natural consequence. In contrast to simple between-country comparisons, such a situation generates not one but at least three legitimate questions related to measurement invariance:

1. Are the analyzed constructs comparable between different rounds in a given country?
2. Are the analyzed constructs comparable between different countries in a given round?
3. Are the analyzed constructs comparable between different countries from different rounds?

Each question has its importance but the last one is key when researchers conduct analyses on multiple rounds of a cross-national survey. Yet, despite its significance for investigating cross-country-cross-time trends, there is no widely accepted methodological procedure that would guide the steps of the measurement invariance analysis involving both spatial and temporal dimensions. Davidov, Schmidt, and Schwartz (2008), while analyzing human values with the ESS data, test the cross-country-cross-time measurement invariance by analyzing cross-classified groups defined by both country and wave variables. Raudenská (2020) analyzes the well-being construct in two rounds of the ESS by, first, checking the cross-country comparability in each wave and, then, performing a simultaneous test on cross-classified groups defined by country and wave variables. A similar approach is employed by Lee (2019) who uses multiple cycles of PISA and investigates the measurement invariance of a socioeconomic status index. A different strategy was proposed by Borgonovi and Pokropek (2019) who argue for a two-step procedure where in the first step they test for cross-time invariance in each country separately and in the second step for cross-country invariance separately for each round. Our analytical strategy is based on the last proposition but consists of three steps.

5.1 Step 1. Within-country-cross-time invariance testing

We start with the cross-time measurement invariance testing because it should hold in most cases. For each country, there should be no cross-time problems with translation. Differences due to changes in institutional and cultural settings within the countries are usually smaller than the differences between the countries. This type of measurement invariance may not be warranted, however, when some extraordinary events occur or when we deal with rapidly changing constructs (for instance, constructs based on the usage of technology, fashion etc.). Yet, because the within-country-cross-time differences are usually not very large, the cross-time analyses usually require more precision to conduct valid comparisons. Hence, even relatively small violations of the assumption of measurement invariance can harm conclusions. If any problems are detected in this step, they should be taken into account in the final analysis.

5.2 Step 2. Within-time-cross-country invariance testing

Regardless of the results in step one, in the second step, the cross-country invariance testing should be conducted for each time point separately. We advise including also the time points and countries that lack cross-time invariance because cross-country measurement invariance may hold in a particular wave even when cross-time invariance for a particular country does not. The literature shows that the measurement invariance between countries, especially the scalar level, is rarely warranted. If the within-time-cross-country invariance does not hold, the analysis should stop here and, depending on the results from step one, we could conclude lack of measurement invariance or within-country-cross-time measurement invariance only. If measurement invariance holds at least for some waves, one should proceed to step 3 or go back to step 1 and repeat the analysis starting with fewer waves, which could lead to more comparable results.

5.3 Step 3. Cross-country-cross-time invariance testing

In this step, cross-time and cross-country dimensions of invariance are tested simultaneously in the whole dataset. Now, we have $C \times T$ groups, where C denotes the number of countries and T – the number of time points. Depending on the results from step two, all waves or a subset should be taken. This step of the analysis should be supplemented with the results from step one. Even if the combined model fits satisfyingly, the results from step one should be taken into account and problematic countries flagged. If the measurement invariance is not warranted, countries where cross-time invariance does not hold should be removed and the simultaneous check of cross-time and cross-country invariance should be repeated.

Additionally, similar to Borgonovi and Pokropek, Davidov, and Schmidt (2019), we advocate for testing simultaneously for classical and partial approximate invariance. In the classical invariance testing, to compare the model fit of increasingly constrained models, we use the criteria described by Rutkowski and Svetina (2016), i. e. ΔRMSEA cutoff of 0.05 for tests of equal slopes and 0.01 for equal slopes and intercepts; ΔCFI threshold of -0.004 for the tests of equal slopes and -0.004 for equal slopes and intercepts. If these indices suggest different conclusions, we will rely on the CFI since it has been proved to be more robust to various misspecifications (Sokolov, 2019). To conduct the analyses, we utilize the mirt package (Chalmers 2012) in R.

For the inspection of partial approximate invariance, we use alignment optimization with the sequential method which was proposed by Asparouhov and Muthén (2014) to detect the non-invariance of parameters. According to the previous studies, alignment provides a good recovery of latent means if the percent of non-invariant parameters is lower than 20-30% (Flake & McCoach, 2018; B. Muthén & Asparouhov, 2013; Pokropek et al., 2019). Any larger number of non-invariant parameters may result in inaccurate estimates. We report the non-invariant parameters to explore which items and in which groups may cause the problem. Additionally, we use R^2 indices for each parameter. The R^2 measure indicates the level of non-invariance that can be absorbed by group-varying factor means and variances. Therefore, an R^2 value close to 1 indicates a high degree of measurement invariance and a value close to 0 a low degree of invariance after the alignment procedure (B. Muthén & Asparouhov, 2014). If R^2 is close to 1, one should not expect differences between the results obtained with the alignment approach and (strict) metric or scalar invariance approaches. If R^2 is considerably below 1, (strict) metric or scalar MI models will fit the data poorly and R^2 will mark the turning point from rather approximate to rather partial (approximate) measurement invariance conditions.

To determine whether the alignment can be used for the country-year mean estimation when the MG-CFA model fails to reach the scalar level, we perform a simulation study for the country-year groups according to the Asparouhov and Muthén (2014) recommendations. That is, we generate the data using the final parameters of the fixed alignment estimation as population parameters. Having these data as the input, we estimate two models: (1) IRT with alignment optimization and (2) scalar IRT model. Comparison of those models allows us to assess the gains we obtain by using alignment optimization instead of the scalar model which ignores the problem of invariance. Simulations are needed because the alignment procedure will always result in the fit equal to the fit from the configural model (the alignment procedure uses parameters from this model). Therefore, it is not possible to determine the level of invariance using fit indices. Simulation studies overcome this problem by checking whether, assuming partial approximate invariance, the recovery of the point estimates and standard errors in the alignment procedure would be better than using the scalar model. We could safely assume that the model that recovers parameters with higher precision and with more accurate standard errors fits the data better.

Due to the complexity of the estimation and the computation time, we use 100 replications. We investigate the correlations between true latent means and estimated means, as well as the mean squared error and the 95% coefficient interval coverage of means. The correlation measure was popularized by Asparouhov and Muthén (2014) as a simple measure of how reliably a given method estimates the latent means. Asparouhov and Muthén (2014) suggest that for reliable rankings of groups one would expect correlations larger than 0.98. This measure is, however, depended not only on the estimation precision but also on how the latent means are close to each other. If the differences between the means are small, it is harder to obtain stable rankings. If the differences are large—it is easier. To account for this problem, we use the mean squared error (MSE), which measures the overall accuracy of the parameter estimation. Finally, we assess the interval estimation of both models by inspecting the coverage of the true means with a 95% confidence interval (CI), generated using the standard errors of the estimated means. For the alignment optimization and simulations, we use Mplus (L. Muthén & Muthén, 1998–2017).

5.4 Comparison

In the end, we will compare the results under the assumption of invariance (scalar model) and approximate invariance (alignment model), and we will investigate the criterion validity by regressing political participation means from each model on three frequently used macro predictors: economic development—GDP per capita (The World Bank, 2020a), economic inequalities—GINI index (The World Bank, 2020b), and the democracy level—V-Dem polyarchy index (Coppedge et al., 2020).

To establish criterion validity, we estimate four Bayesian multivariate models, each with the two political participation variables (standardized country-round means from the scalar and alignment solutions) as regressands using Hamiltonian Monte Carlo as adopted in Stan (Carpenter et al., 2017). The economic variables are tested in bivariate regression models, and the polyarchy index, first, in a bivariate model, then, in a multivariable regression since we want to adjust for the confounding effect of the economic development (Kołczyńska, 2020). Because the data have several missing values on the GINI index, we impute the missings using predictive mean matching with five iterations, and with GDP, household disposable income (OECD, 2020), and the V-Dem egalitarian democracy index (Coppedge et al., 2020) as predictors. We use the mice package (van Buuren & Groothuis-Oudshoorn, 2011) for imputations. Pooled results from multiple imputed data sets are obtained by combining the posterior samples (five for each regression model).

Moreover, to reduce the influence of the extreme observations, we replace the Gaussian model with a thicker-tailed Student-t distribution. We standardize the variables and use weakly informative priors for regressions weights and intercepts, Normal(0,1) and Normal(0, 0.5), respectively, the default half Student-t(3, 0, 2.5) for scales of the marginal Student-t distributions, and we fix the parameter denoting the degrees of freedom of the Student-t distributions to two. We use 2000 iterations, out of which 500 are warmups, and four chains. We conduct the analysis in R with the brms package (Bürkner, 2017).

6 Results

6.1 Step 1. Within country cross-time invariance testing

Following the procedure described earlier, we start with the within-country-cross-time invariance testing. Table 2 shows the results from the classical measurement invariance testing approach using multiple group CFA models for categorical variables (or IRT models), where invariance was tested within each country.

Relying on the ΔCFI values, two countries out of 15 reach only configural invariance, 12 countries reach the metric level of measurement invariance, and one country reaches scalar invariance. Therefore, according to classical invariance testing, the construct of political participation, as measured by the ESS items, cannot be quantitively compared in Portugal and Ireland. For these two, the multiple-group analysis suggests that only the configural, the least restrictive model, can be retained. In 12 countries the measurement units are the same across time but cross-time anchoring of the scales could not be established. As a practical consequence, for these 12 countries, we can reliably compare the relationships between the scale of political participation and other variables between time points. In the case of the United Kingdom, the analysis does not indicate any violations up to the scalar level which suggests a possibility of reliable comparison of means across time.

The results of the within-country-cross-time approximate measurement invariance testing are presented in Table 3.

The results are consistent with the classical measurement invariance analysis but conclusions are more optimistic. Countries that reached only configural invariance in classical measurement invariance analysis are also the countries with the highest R^2s (Portugal: average $R^2 = 0.69$; Ireland: average $R^2 = 0.60$). Interestingly, the problem of non-invariance seems to have a different nature in both countries. In Ireland, nine parameters were flagged as fully non-invariant, hence, indicating that the non-invariance problem is most likely driven by large differences in these parameters. In Portugal, only one parameter was flagged as fully non-invariant suggesting that the non-invariance might be caused by small differences spread across all parameters. The results confirm also that the measurement of political participation is the least problematic in the UK, where the R^2 values are moderate but no parameters show full non-invariance.

Table 2
Cross-time analysis: classical measurement invariance

	Invariance level				
	Metric		Scalar		
Country	ΔRMSEA	ΔCFI	ΔRMSEA	ΔCFI	Level reached
Belgium (BE)	−0.0031	−0.0020	−0.0002	−0.0113	Metric
Estonia (EE)	−0.0023	0.0000	0.0003	−0.0082	Metric
Finland (FI)	−0.0033	−0.0033	0.0008	−0.0307	Metric
France (FR)	−0.0026	−0.0006	0.0008	−0.0090	Metric
Germany (DE)	−0.0037	−0.0016	−0.0005	−0.0115	Metric
Hungary (HU)	−0.0026	−0.0002	0.0031	−0.0183	Metric
Ireland (IE)	−0.0025	−0.0050	−0.0005	−0.0067	Configural
The Netherlands (NL)	−0.0026	−0.0017	0.0011	−0.0204	Metric
Norway (NO)	−0.0036	−0.0040	−0.0006	−0.0123	Metric
Poland (PL)	−0.0026	0.0000	0.0016	−0.0105	Metric
Portugal (PT)	0.0001	−0.0084	0.0013	−0.0118	Configural
Spain (ES)	−0.0020	−0.0029	0.0016	−0.0106	Metric
Sweden (SE)	−0.0027	−0.0036	0.0033	−0.0447	Metric
Switzerland (CH)	−0.0037	−0.0032	0.0003	−0.0197	Metric
United Kingdom (GB)	−0.0018	−0.0016	−0.0002	−0.0040	Scalar

6.2 Step 2. Within time cross-country invariance testing

In step 2, we move to the cross-country comparisons that are performed for each round separately. In Table 4, the results from the classical measurement invariance testing are reported.

Results presented in Table 4 suggest that any quantitative comparison of political participation across countries within each round of the ESS can be rather difficult if not impossible. In none of the rounds, metric or scalar level of measurement invariance was retained. This is not surprising given the number of country groups (15) and their heterogeneity (countries with different political cultures, political systems, and history). Based on this, we may say that the construct of political participation is manifested in different ways across countries for a specific round. When juxtaposed with Table 2, the results also show that it is much more difficult to derive meaningful conclusions about the differences regarding levels and patterns of political participation across countries for a given time point than it is for a country across time since the cross-country differences are often more pronounced (cf. Borgonovi & Pokropek, 2019).

The results from the alignment analysis, presented in Table 5, confirm those obtained in the classical measurement invariance analysis: Establishing cross-country comparability may be very difficult. Scalar and metric measurement invariance levels would be hard to retain as the R^2 values are significantly lower than 1 (around 0.6) and the number of non-invariant parameters is large. It is especially true for the thresholds where, on average, 34% of the parameters are flagged as non-invariant, which slightly exceeds the recommended levels.

6.3 Step 3. Cross-country-cross-time invariance testing

In step 3, alignment optimization was performed for 120 groups defined by countries and years to investigate the cross-country-cross-time invariance. As in the previous steps, measurement invariance was investigated by the means of the R^2 index and the percentage of non-comparable item parameters. Table 6 presents the results from the perspective of indicators. The alignment procedure suggests that 3% of factor loadings and 39% of thresholds are non-invariant which corresponds to R^2 of 0.53 and 0.64, respectively. R^2 values smaller than one imply that the problem of non-invariance exists but, to some extent, could be accommodated by the alignment method. The most problematic item, in terms of thresholds, is the "taking part in demonstrations", where a large number of non-comparable thresholds (44%) is accompanied by a rather low R^2 value (0.47). Two other items—"working for organizations" and "boycotting products"—are also problematic in terms of thresholds in almost half of the groups. Overall, Table 6 suggests that the comparability of groups might be problematic.

The results of simulations presented in Table 7 confirm the earlier findings. Correlations between the true and estimated means are higher for the alignment solution (0.97) than for the scalar solution (0.93). MSE is much lower for the alignment model (0.40) than for the scalar model (3.46). The 95% mean coverage is unacceptably low for the scalar model (0.28) and acceptable for the alignment model (0.82).

Table 3
Cross-time analysis: alignment optimization with R^2 indices and non-invariant parameters listed

	R^2			Non-invariant			
Country[a]	All	Loadings	Thresholds	parameters[b]	rounds	%L	%T
BE	0.27	0.19	0.35	T WRKORG	2	0	2
CH	0.18	0.22	0.15	T BCTPRD T WRKORG	2 8	0	4
DE	0.43	0.28	0.57	T BCTPRD	3 2	0	4
EE	0.30	0.20	0.41	T BCTPRD	5	0	2
ES	0.54	0.39	0.68	L PBLDMN T BCTPRD T PBLDMN	2 6 4 2	2	5
FI	0.26	0.19	0.33	T SGNPTIT	5 6 2	0	5
FR	0.26	0.22	0.30	T BADGE T BCTPRD T PBLDMN	9 3 4 2 5	0	9
GB	0.34	0.21	0.47	none	none	0	0
HU	0.32	0.31	0.34	T BADGE	8	2	2
IE	0.60	0.53	0.68	L BCTPRD L WRKPRTY T BCTPRD T PBLDMN	6 9 2 4 5 6 7 8	4	11
NL	0.42	0.31	0.53	T WRKORG	2	0	2
NO	0.37	0.33	0.41	T BADGE T WRKORG	2 2	0	4
PL	0.26	0.20	0.32	T PBLDMN	9	0	2
PT	0.69	0.60	0.78	T PBLDMN	6	0	2
SE	0.40	0.23	0.57	T BADGE T SGNPTIT	3 4 5 6 3 4 2	0	13

[a] See Table 2 for meaning of country codes [b] L = loadings; T = thresholds; %L = percentage of loadings; %T = percentage of thresholds.

6.4 Model comparisons and scale validation

How different are the results from scalar and alignment models in practice? This can be answered by comparing the dashed and solid lines in Figure 1, which shows standardized country-round means from the scalar and alignment solutions. Overall, there are no big differences in the mean estimates. There is an instance of systematic overestimation—the UK—and underestimation—Estonia. Also, we would find two big discrepancies in Hungary for rounds 2 and 7.

However, the key difference between the two models lies in the uncertainty estimates[2], i. e. the scalar model underestimates the uncertainty, with Hungary being the most vivid example of this.

By looking at Figure 1, we can also make substantive observations about the levels of political participation in 15 countries between 2004 and 2018. For this purpose, consider the results from the alignment model. First, we can distinguish two groups of countries with respect to the levels of participation. We have a group where the average level of participation is equal or greater than the global mean, with 11 Western European countries, and a group of four countries—Hungary, Poland, Estonia, and Portugal—where the levels,

[2]Standard errors for the standardized solution were calculated using the delta method.

Table 4
Cross-time analysis: classical measurement invariance

	Invariance level				
	Metric		Scalar		
Round	ΔRMSEA	ΔCFI	ΔRMSEA	ΔCFI	Level reached
2	−0.0003	−0.0170	0.0104	−0.1581	Configural
3	0.0001	−0.0166	0.0102	−0.1464	Configural
4	−0.0002	−0.0169	0.0095	−0.1415	Configural
5	0.0008	−0.0212	0.0087	−0.1144	Configural
6	0.0000	−0.0167	0.0114	−0.1714	Configural
7	−0.0005	−0.0148	0.0109	−0.1602	Configural
8	0.0001	−0.0217	0.0097	−0.1567	Configural
9	0.0005	−0.0254	0.0107	−0.1736	Configural

on average, are lower than the global mean. This grouping corroborates existing research. Lower levels of participation in the first three countries from the second group can be ascribed to the communist legacy, including the weak civil society, and low levels of interpersonal trust (Letki, 2004; Pop-Eleches & Tucker, 2013). In the case of Portugal, lower levels of political participation are explained by a mixture of the "post-honeymoon effect", i. e. the disillusionment after the initial enthusiasm with the democratic system (Inglehart & Catterberg, 2002), and low levels of educational attainment among the population (Magalhães, 2005).

As to the dynamic, in most countries, we cannot observe substantial changes. There are two exceptions, however, i. e. Spain and Ireland. In Spain, the fluctuations ranged from -0.5 to 1 *SD*s between 2008 and 2012 which might be attributed to the activity of the *Indignados* movement (Gerbaudo, 2016). In Ireland, we can observe a substantive drop in 2010, right after the economic crisis of 2008. Additionally, there are countries where we can observe a steady increase over the years: Sweden, Estonia, Germany, and Portugal. In the rest, the average level of political participation remained roughly the same between 2004 and 2018.

Apart from the differences in the descriptive statistics, we should also check the behavior of the two solutions when predicted by external variables. Figure 2 shows standardized regression coefficients and Bayesian R^2 for the four regression models described in the previous section. GDP per capita, which measures economic development, has a positive relationship with political participation. Similarly, higher levels of democracy, as measured by the polyarchy index, are associated with higher participation rates. Once adjusted for economic development, the coefficient for polyarchy is still positive but gets smaller. By contrast, the coefficient for inequalities, which is measured with the GINI index, has a negative sign. Out of the three predictors, GDP per capita has the greatest explanatory power, and the GINI index—most modest.

Regardless of whether we use the scalar or alignment solution, we obtain similar results in terms of the sign of the coefficients and the effect sizes. However, the exact effect sizes and the uncertainty around them differ between the scalar and alignment solutions. Similarly, there are no large discrepancies in terms of the amount of explained variance but the exact numbers differ. We need to bear in mind that the differences might be more pronounced while comparing the quantities of interest for particular countries or rounds.

7 Discussion

In this article, we presented a thorough investigation of measurement invariance of political participation using the ESS battery of items. To do so, we proposed a novel approach to measurement invariance testing that allows us to directly analyze different sources of measurement non-invariance—due to time and due to country differences. Moreover, we showed what we gain by using alignment optimization when the classical method based on the MG-CFA fails to reach the scalar level of invariance, which is often the case when multiple sources of variation are present.

In the previous research on political participation, we could rarely find an attempt to test the measurement invariance. If there were any attempts, they were usually confined to the cross-country measurement invariance testing. The temporal dimension was missing, and so the question of whether we could compare means or regression coefficients from different countries across time, say, country A in t_1 with country B in t_2. Additionally, the methodological approaches to the cross-country-cross-time measurement invariance testing either did not allow us to identify the source of measurement non-invariance (e. g. Davidov et al., 2008) or precluded the identification of subsets for which cross-country-cross-time invariance was possible (e. g. Raudenská, 2020). Also, while using alignment optimization, researchers rarely demonstrated the gains of using the alignment technique.

The results of our research show that if we relied only on

Table 5
Cross-country analysis: alignment optimization with R^2 indices and non-invariant parameters listed

	R^2			Loadings Non-invariant		Intercepts				Non-invariant	
Round	All	Load.	Thres.	item	country	item	country	item	country	%L	%T
2	0.57	0.52	0.62	contplt	FR HU IE	badge	DE NO	pbldmn	DE ES FI PT	5	37
2				pbldmn	ES	bctprd	CH DE FR GB SE	sgnptit	CH DE EE FI GB SE		
2				wrkprty	CH	contplt	CH DE ES FI FR NO SE BE	wrkorg	DE ES NL SE BE		
2						wrkprty	CH EE ES HU IE NL PL PT BE				
3	0.53	0.50	0.55	contplt	FR PT	badge	NO PT SE	pbldmn	DE ES FI FR HU PT	4	37
3				wrkorg	BE	bctprd	CH DE FR GB NO SE	sgnptit	CH EE GB HU PL PT SE		
3				wrkprty	CH	contplt	FI FR HU IE	wrkorg	DE GB HU NL NO SE BE		
3								wrkprty	CH EE HU PL PT BE		
4	0.62	0.56	0.67			badge	DE SE	pbldmn	CH DE ES FI FR IE NO PT BE	0	33
4						bctprd	CH DE FI FR GB SE	sgnptit	CH DE ES FR GB SE BE		
4						contplt	CH FR NO	wrkorg	DE FI GB NL SE BE		
4								wrkprty	CH EE		
5	0.72	0.73	0.71	badge	GB	badge	DE FI NO PL SE	pbldmn	DE ES FI FR IE PT BE	2	36
5				contplt	IE	bctprd	CH DE FI FR GB IE SE	sgnptit	EE FI GB HU NO		
5						contplt	ES FR HU NO SE BE	wrkorg	DE FI GB NL SE BE		
5								wrkprty	CH BE		
6	0.57	0.51	0.63	bctprd	ES	badge	FI NO SE	pbldmn	DE EE ES FI FR HU IE PT BE	3	31
6				pbldmn	PT	bctprd	CH DE FI FR GB SE	sgnptit	EE FI HU NO PL BE		
6				sgnptit	FI	contplt	ES FR NO SE	wrkorg	DE FI GB NL SE		
7	0.58	0.52	0.64	badge	GB	badge	DE FI GB NO	pbldmn	ES FI FR HU IE PT	2	36
7				sgnptit	FI	bctprd	CH DE FI FR GB NO PT SE	sgnptit	EE GB PT		
7						contplt	CH DE ES FI FR NO SE BE	wrkorg	DE ES FI NL NO SE BE		
7								wrkprty	CH SE		
8	0.65	0.63	0.67	contplt	IE	badge	CH DE FI NL NO	pbldmn	CH EE ES FI NL NO SE	5	29
8				sgnptit	FI	bctprd	CH DE FI FR SE	wrkorg	DE ES FI NL NO SE BE		
8				wrkorg	IE NO SE	contplt	CH EE FR PT SE	wrkprty	PT		
9	0.68	0.64	0.72	contplt	IE	badge	CH DE FI NO	sgnptit	GB	4	35
9				pbldmn	ES	bctprd	CH DE FI FR GB PL PT SE	wrkorg	EE FI FR GB IE NL PL SE		
9				wrkorg	IE	contplt	CH DE ES FI FR NO PL SE	wrkprty	CH EE		
9				wrkprty	IE	pbldmn	ES FI FR HU PL PT				

Table 6
Cross-country-cross-time analysis: alignment optimization with R^2 indices and non-invariant parameters listed

	R^2		Loadings Non-invariant		Intercepts		Non-invariant	
Item	Load.	Thres.	Round	Country	Round	Country	%L	%T
contplt	0.16	0.39	2	IE	2	CH DE ES FR NO SE	2	39
			3		3	CH DE ES FI FR NO SE		
			4		4	BE CH FR SE		
			5	IE	5	BE ES FI FR NO		
			6		6	CH DE ES FR NO SE		
			7		7	DE ES FI FR NO SE		
			8	IE	8	CH DE ES FR NO SE		
			9	IE	9	DE ES FI FR IE NO SE		
wrkprty	0.44	0.65	2	CH	2	FI FR GB SE	2	33
			3		3	FI FR GB NO		
			4		4	DE FI FR GB SE		
			5		5	FI FR GB NO SE		
			6		6	FI GB IE SE		
			7		7	DE EE FI FR GB SE		
			8		8	DE FI FR GB SE		
			9	IE	9	DE FI FR GB IE NO SE		
wrkorg	0.69	0.77	2		2	BE DE FI SE	5	49
			3	BE	3	BE DE ES FI GB HU NL NO SE		
			4	SE	4	BE DE ES FI GB NL NO SE		
			5		5	DE ES FI GB NL NO SE		
			6	NL	6	DE ES FI GB NL NO SE		
			7	SE	7	DE ES FI GB NL NO SE		
			8	NO	8	BE DE ES FI GB NL NO SE		
			9	IE	9	BE CH DE ES FI GB IE NL SE		
badge	0.57	0.68	2		2	DE NO	2	28
			3		3	DE FI NO SE		
			4		4	DE FI NO SE		
			5		5	CH DE FI NO SE		
			6		6	CH DE FI NO SE		
			7	GB	7	CH DE FI NO SE		
			8	GB	8	CH DE FI NO		
			9		9	CH DE FI NO PL		
sgnptit	0.60	0.82	2		2	CH DE FI GB PT SE	3	34
			3	FI	3	CH GB PT SE		
			4		4	CH FR GB SE		
			5		5	EE GB HU		
			6		6	EE ES FI FR GB IE SE		
			7	FI	7	DE EE ES FR GB IE		
			8	FI	8	CH DE ES FR GB SE		
			9		9	DE FR GB IE NO		

Continues on next page

Continued from last page

Item	R^2		Loadings Non-invariant		Intercepts		Non-invariant	
	Load.	Thres.	Round	Country	Round	Country	%L	%T
pbldmn	0.60	0.47	2	ES	2	BE CH DE ES FI FR	3	44
			3		3	BE DE ES FI FR HU		
			4		4	BE CH DE ES FI FR IE PT		
			5		5	BE DE ES FI FR IE		
			6		6	DE EE ES FI FR HU IE PT		
			7		7	BE DE ES FI FR HU IE		
			8		8	BE DE ES FI FR IE PL		
			9	ES, PL	9	ES FI FR HU PL		
bctprd	0.67	0.73	2		2	CH DE FI FR GB NO SE	2	48
			3		3	CH DE FI FR GB NO PT SE		
			4		4	CH DE FI FR GB NO SE		
			5		5	CH DE FI FR GB PT SE		
			6	DE, ES	6	CH DE FI FR GB NL NO SE		
			7		7	CH DE FI FR GB NO SE		
			8		8	CH DE FI FR GB NO SE		
			9		9	CH DE FI FR GB SE		
Total	0.53	0.64	-		-		3	39

Table 7
Simulation study for the alignment and scalar models: recovery of means and rankings based on the means

Correlations		MSE		95% mean coverage	
Alignment	Scalar	Alignment	Scalar	Alignment	Scalar
0.97	0.93	0.40	3.46	0.82	0.28

the MG-CFA approach, we would conclude that the comparability of the scale of political participation is very limited. The only possible quantitative comparison would involve unstandardized regression coefficients for each country across time, with a few exceptions. Any comparison of means would violate the measurement invariance assumptions and, as the simulation study shows, would lead to poor estimates of the standard errors, which would be significantly underestimated. The simulations also demonstrate that the alignment method provides the most accurate point estimates and standard errors, and better recovery of rankings of countries with respect to the levels of political participation. Hence, the substantive conclusions are less likely to be influenced by an imprecise measurement while using the alignment.

Substantively, the study identifies two groups of countries where the average levels of participation differ. One is comprised of Western European countries, with higher mean levels, and another one—composed of three post-communist democracies and Portugal. What is more, we found that most of the countries did not witness substantive changes in terms of levels of participation, with exception of Spain and Ireland. Also, we observed four cases, where the levels of participation steadily increased between 2004 and 2018: Sweden, Estonia, Germany, and Portugal. Lastly, we found that the economic development, inequalities, and the level of democracy are all associated with the aggregated levels of participation, with economic development being the most influential predictor.

Our study has, however, several limitations. Although the simulations suggest that we could expect reasonably reliable results, the level of non-invariant parameters was relatively high and above the thresholds proposed in the literature. More simulation studies are needed to understand how effective alignment optimization could be in such situations. Moreover, some studies (Finch, 2016; Lin, 2020) suggest that DIF detection as implemented in the Mplus alignment algorithm might not be optimal in all situations. Alignment optimization produces the best solution given the assumed loss function. But one should keep in mind that this best solution might still not be comparable enough to make valid

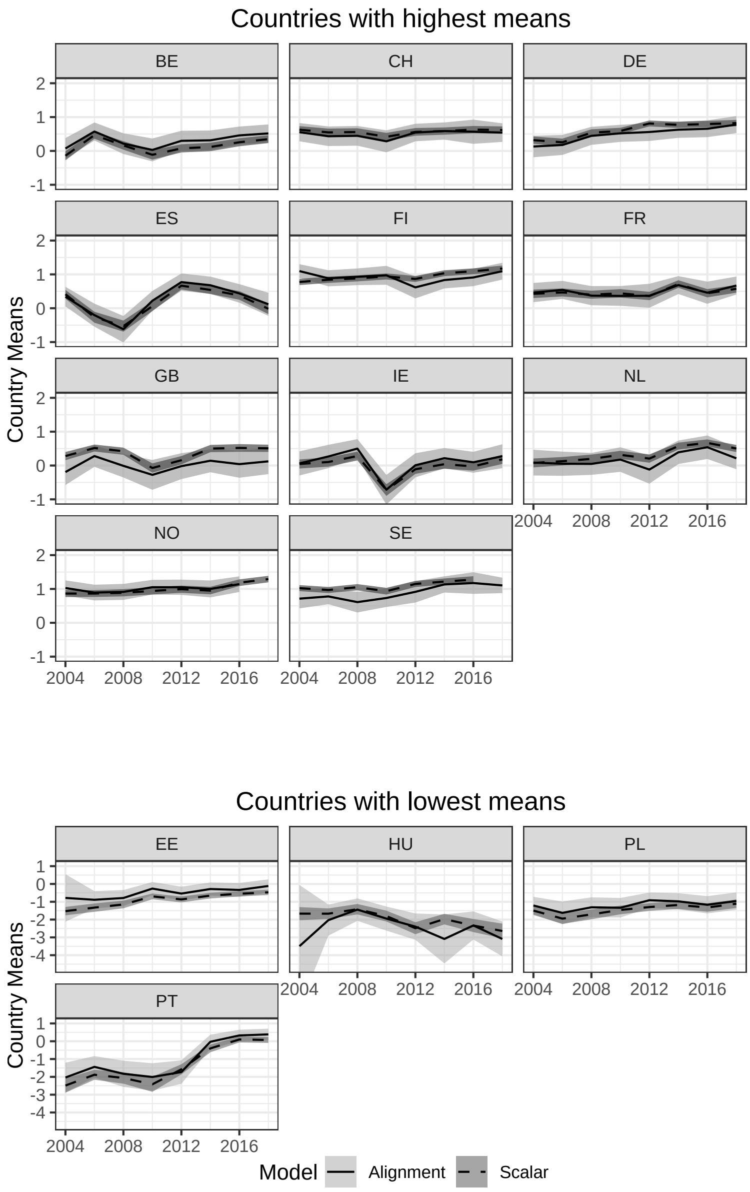

Figure 1. Standardized country-year means from scalar and alignment models, 2004-2018. Note: Ribbons represent 95% confidence intervals. There are no estimates for Norway, round 9, alignment solution, and for Sweden, round 9, scalar solution because these parameters were fixed for estimation.

claims about differences between countries and our analyses provide the most plausible but no definitive results.

Finally, further qualitative studies would be welcome to "explain" the non-invariance of items. In this study, we were able to show that some items are non-invariant but further studies could try to answer the question as to why particular items behave differently in different countries and propose some changes to make the measures more comparable.

Our analysis adopted an all-or-nothing strategy in the sense that we assumed that there is one cluster of countries in which countries could be successfully compared. This does not need to be necessarily true. One might assume that there are several clusters of countries in which countries are comparable without between-cluster comparability. There are no established methodologies for such an alternative approach although some works (de Roover, Vermunt, Timmerman, & Ceulemans, 2017; Kim, Joo, Lee, Wang, & Stark, 2016) suggest that mixture modeling for measurement invariance test-

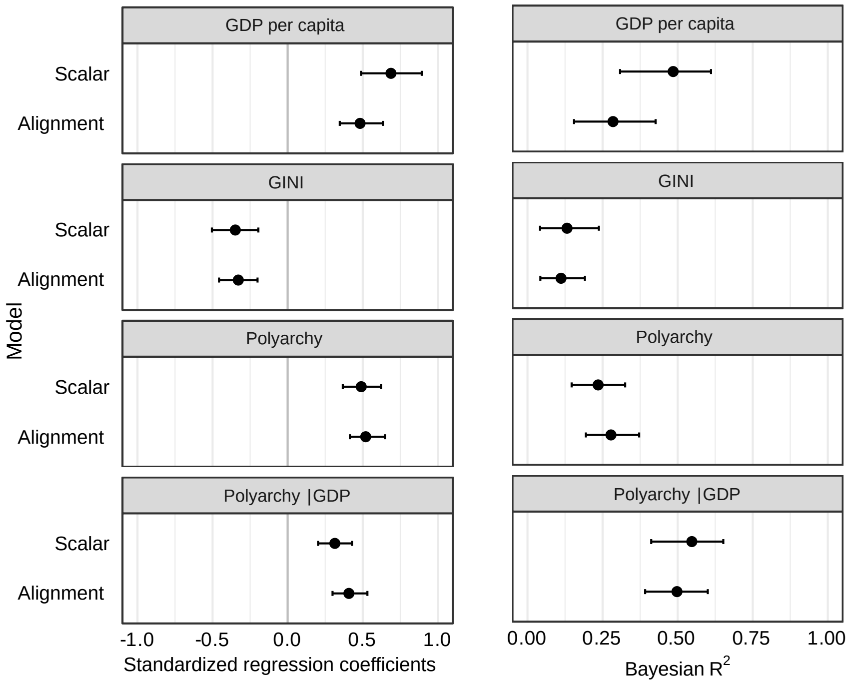

Figure 2. Regression coefficients and Bayesian R^2 for scalar and alignment models with 95% credible intervals

ing could be an option.

For future research, we would also recommend including more countries and time points to see how the levels of participation look like in other countries and years. Lastly, it would be beneficial to assume a more causal approach to the explanation of cross-country-cross-time variations, with clearly identified and singled-out causal mechanisms.

Acknowledgement

The work of Artur Pokropek has been prepared under the project *Scales Comparability in Large-Scale Cross-Country Surveys*, which is funded by the Polish National Science Centre, as a part of the grant competition Sonata 8 (UMO-2014/15/D/HS6/04934).

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508. doi:10.1080/10705511.2014.919210

Bäck, M., & Christensen, H. S. (2016). When trust matters—a multilevel analysis of the effect of generalized trust on political participation in 25 European democracies. *Journal of Civil Society*, *12*(2), 178–197. doi:10.1080/17448689.2016.1176730

Bang, H. P. (2009). 'yes we can': Identity politics and project politics for a late-modern world. *Urban Research & Practice*, *2*(2), 117–137. doi:10.1080/17535060902979022

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology*, *49*(5), 713–734. doi:10.1177/0022022117749042

Borgonovi, F., & Pokropek, A. (2019). Education and attitudes toward migration in a cross country perspective. *Frontiers in psychology*, *10*, 2224. doi:10.3389/fpsyg.2019.02224

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). doi:10.18637/jss.v080.i01

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. doi:10.1037/0033-2909.105.3.456

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan : A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. doi:10.18637/jss.v076.i01

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. doi:10.1207/S15328007SEM0902{\textunderscore}5

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *71*(S1), 157–186. doi:10.1007/s11577-019-00598-7

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., … Ziblatt, D. (2020). V-dem dataset v10: Varieties of democracy (V-Dem) project. Retrieved from https://doi.org/10.23696/vdemds20

Dalton, R. J. (2008). Citizenship norms and the expansion of political participation. *Political Studies*, *56*(1), 76–98. doi:10.1111/j.1467-9248.2007.00718.x

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. doi:10.1146/annurev-soc-071913-043137

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, *72*(3), 420–445. doi:10.1093/poq/nfn035

de Roover, K., Vermunt, J. K., Timmerman, M. E., & Ceulemans, E. (2017). Mixture simultaneous factor analysis for capturing differences in latent variables between higher level units of multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 506–523. doi:10.1080/10705511.2017.1278604

ESS. (2002). European Social Survey round 1 data data file edition 6.6: NSD—Norwegian Centre for Research Data. doi:10.21338/NSD-ESS1-2002

ESS. (2018). European Social Survey round 7 (2014) documentation report. doi:10.21338/NSD-ESS7-2014

Finch, W. H. (2016). Detection of differential item functioning for more than two groups: A monte carlo comparison of methods. *Applied Measurement in Education*, *29*(1), 30–45. doi:10.1080/08957347.2015.1102916

Flake, J. K., & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 56–70. doi:10.1080/10705511.2017.1374187

García-Albacete, G. M. (2014). *Young people's political participation in Western Europe*. doi:10.1057/9781137341310

Gerbaudo, P. (2016). The indignant citizen: Anti-austerity movements in southern europe and the anti-oligarchic reclaiming of citizenship. *Social Movement Studies*, *16*(1), 36–50. doi:10.1080/14742837.2016.1194749

Goroshit, M. (2016). Political participation: A latent variable approach testing measurement equivalence of political participation using ESS data. *Eurasian Journal of Social Sciences*, *4*(1), 26–38. doi:10.15604/ejss.2016.04.01.003

Hooghe, M., Hosch-Dayican, B., & van Deth, J. W. (2014). Conceptualizing political participation. *Acta Politica*, *49*(3), 337–348. doi:10.1057/ap.2014.7

Hooghe, M., & Marien, S. (2013). A comparative analysis of the relation between political trust and forms of political participation in Europe. *European Societies*, *15*(1), 131–152. doi:10.1080/14616696.2012.692807

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, *18*(3-4), 117–144. doi:10.1080/03610739208253916

Inglehart, R., & Catterberg, G. (2002). Trends in political action: The developmental trend and the post-honeymoon decline. *International Journal of Comparative Sociology*, *43*(3-5), 300–316. doi:10.1177/002071520204300305

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426.

Kim, E. S., Joo, S.-H., Lee, P., Wang, Y., & Stark, S. (2016). Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(6), 870–887. doi:10.1080/10705511.2016.1196108

Koc, P. (2021). Measuring non-electoral political participation: Bi-factor model as a tool to extract dimensions. *Social Indicators Research*, *156*(1), 271–287. doi:10.1007/s11205-021-02637-3

Kołczyńska, M. (2020). Micro- and macro-level determinants of participation in demonstrations: An analysis of cross-national survey data harmonized ex-post. *mda. methods, data, analyses*, *14*(1), 91–126. doi:10.12758/mda.2019.07

Kostelka, F. (2014). The state of political participation in post-communist democracies: Low but surprisingly little biased citizen engagement. *Europe-Asia Studies*, *66*(6), 945–968. doi:10.1080/09668136.2014.905386

Lee, S. S. (2019). Longitudinal and cross-country measurement invariance of the PISA home possessions scale. Retrieved January 26, 2021, from https://www.semanticscholar.org/paper/Longitudinal-And-Cross-Country-Measurement-Of-The-Lee/8fe54015ec384590b24337582b8bae4748f4248e

Letki, N. (2004). Socialization for participation? Trust, membership, and democratization in East-Central Europe. *Political Research Quarterly*, *57*(4), 665–679. doi:10.1177/106591290405700414

Lin, L. (2020). *Evaluate measurement invariance across multiple-groups: A comparison between the alignment optimization and the random item effects model* (Doctoral thesis, University of Pittsburgh).

Magalhães, P. C. (2005). Disaffected democrats: Political attitudes and political action in Portugal. *West European Politics*, *28*(5), 973–991. doi:10.1080/01402380500310626

Marien, S., Hooghe, M., & Quintelier, E. (2010). Inequalities in non-institutionalised forms of political participation: A multi-level analysis of 25 countries. *Political Studies*, *58*(1), 187–213. doi:10.1111/j.1467-9248.2009.00801.x

Marsh, A. (1974). Explorations in unorthodox political behaviour: A scale to measure "protest potential". *European Journal of Political Research*, *2*(2), 107–129. doi:10.1111/j.1475-6765.1974.tb01233.x

Marsh, A. (1977). *Protest and political consciousness*. Beverly Hills, CA: Sage Publications.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi:10.1007/BF02294825

Milbrath, L. W. (1965). *Political participation: How and why do people get involved in politics?* Chicago: Rand McNally College.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological methods*, *17*(3), 313–335. doi:10.1037/a0026802

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. *Mplus web notes*, *17*, 1–48.

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in psychology*, *5*(Article 978). doi:10.3389/fpsyg.2014.00978

Muthén, L., & Muthén, B. (1998–2017). *Mplus user's guide: Eighth edition* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

Norris, P. (2007). Political activism: New challenges, new opportunities. In C. Boix & S. C. Stokes (Eds.), *The oxford handbook of comparative politics* (pp. 628–649). New York: Oxford University Press.

Norris, P., Walgrave, S., & van Aelst, P. (2005). Who demonstrates? Antistate rebels, conventional participants, or everyone? *Comparative Politics*, *37*(2), 189. doi:10.2307/20072882

OECD. (2020). Household disposable income (indicator). doi:10.1787/dd50eddd-en

Parry, G., Moyser, G., & Day, N. (1992). *Political participation and democracy in Britain*. Cambridge University Press.

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 724–744. doi:10.1080/10705511.2018.1561293

Pokropek, A., Lüdtke, O., & Robitzsch, A. (2020). An extension of the invariance alignment method for scale linking. *Psychological Test and Assessment Modeling*, *62*(2), 305–334.

Poortinga, Y. H., & Malpass, R. S. (1986). Making inferences from cross-cultural data. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 17–46). Beverly Hills, Calif.: Sage.

Pop-Eleches, G., & Tucker, J. A. (2013). Associated with the past? *East European Politics and Societies: and Cultures*, *27*(1), 45–68. doi:10.1177/0888325412465087

Quaranta, M. (2013). Measuring political protest in Western Europe: Assessing cross-national equivalence. *European Political Science Review*, *5*(3), 457–482. doi:10.1017/S1755773912000203

Raudenská, P. (2020). The cross-country and cross-time measurement invariance of positive and negative affect scales: Evidence from European Social Survey. *Social science research*, *86*, 102369. doi:10.1016/j.ssresearch.2019.102369

Robitzsch, A. (2020). Lp loss functions in invariance alignment and haberman linking with few or many groups. *Stats*, *3*(3), 246–283. doi:10.3390/stats3030019

Rutkowski, L., & Svetina, D. (2016). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education*, *30*(1), 39–51. doi:10.1080/08957347.2016.1243540

Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, *6*(2), 169–215. doi:10.1177/1094428103251542

Sokolov, B. (2019). Sensitivity of goodness of fit indices to lack od measurement invariance with categorical indicators.

Stegmueller, D. (2011). Apples and oranges? The problem of equivalence in comparative research. *Political Analysis*, *19*(4), 471–487. doi:10.1093/pan/mpr028

The World Bank. (2020a). World development indicators: GDP per capita (current US$). Retrieved January 5, 2021, from https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD

The World Bank. (2020b). World development indicators: GINI index (world bank estimate). Retrieved January 5, 2021, from https://databank.worldbank.org/reports.aspx?source=2&series=SI.POV.GINI

van Aelst, P., & Walgrave, S. (2001). Who is that (wo)man in the street? From the normalisation of protest to the normalisation the protester. *European Journal of Po-*

litical Research, *39*(4), 461–486. doi:10 . 1023 / A : 1011030005789

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice : Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. doi:10 . 18637/jss.v045.i03

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. doi:10.1016/j.erap.2003.12.004

van Deth, J. W. (1986). A note on measuring political participation in comparative research. *Quality and Quantity*, *20*(2), 261–272. doi:10.1007/BF00227430

van Deth, J. W. (2014). A conceptual map of political participation. *Acta Politica*, *49*(3), 349–367. doi:10.1057/ap.2014.6

van Deth, J. W., & Theocharis, Y. (2018). *Political participation in a changing world: Conceptual and empirical challenges in the study of citizen engagement*. New York, N.Y.: Taylor & Francis Ltd.

Verba, S., Nie, N. H., & Kim, J.-o. (1978). *Participation and political equality: A seven-nation comparison*. Cambridge: Cambridge University Press.

Vráblíková, K. (2014). How context matters? Mobilization, political opportunity structures, and nonelectoral political participation in old and new democracies. *Comparative Political Studies*, *47*(2), 203–229. doi:10 . 1177/0010414013488538

Appendix

(*Appendix table follows on next page*)

Table A1
Sample sizes for each country-round

Country	ESS round	n	Country	ESS round	n	Country	ESS round	n	Country	ESS round	n
BE	2	1682	ES	2	1549	HU	2	1415	PL	2	1601
BE	3	1699	ES	3	1786	HU	3	1448	PL	3	1614
BE	4	1675	ES	4	2469	HU	4	1494	PL	4	1534
BE	5	1637	ES	5	1827	HU	5	1521	PL	5	1647
BE	6	1785	ES	6	1831	HU	6	1928	PL	6	1802
BE	7	1678	ES	7	1862	HU	7	1647	PL	7	1551
BE	8	1698	ES	8	1894	HU	8	1553	PL	8	1619
BE	9	1675	ES	9	1598	HU	9	1610	PL	9	1423
CH	2	2066	FI	2	1917	IE	2	2148	PT	2	1988
CH	3	1742	FI	3	1810	IE	3	1573	PT	3	2159
CH	4	1762	FI	4	2099	IE	4	1709	PT	4	2283
CH	5	1429	FI	5	1788	IE	5	2498	PT	5	2096
CH	6	1419	FI	6	2110	IE	6	2555	PT	6	2103
CH	7	1452	FI	7	2013	IE	7	2290	PT	7	1211
CH	8	1447	FI	8	1860	IE	8	2662	PT	8	1234
CH	9	1449	FI	9	1677	IE	9	2127	PT	9	1007
DE	2	2663	FR	2	1754	NL	2	1822	SE	2	1851
DE	3	2761	FR	3	1920	NL	3	1842	SE	3	1825
DE	4	2651	FR	4	1997	NL	4	1735	SE	4	1721
DE	5	2861	FR	5	1675	NL	5	1778	SE	5	1418
DE	6	2806	FR	6	1910	NL	6	1806	SE	6	1761
DE	7	2913	FR	7	1839	NL	7	1851	SE	7	1705
DE	8	2708	FR	8	2005	NL	8	1634	SE	8	1474
DE	9	2259	FR	9	1948	NL	9	1582	SE	9	1490
EE	2	1869	GB	2	1818	NO	2	1686			
EE	3	1431	GB	3	2301	NO	3	1662			
EE	4	1589	GB	4	2263	NO	4	1478			
EE	5	1722	GB	5	2336	NO	5	1466			
EE	6	2282	GB	6	2208	NO	6	1538			
EE	7	1975	GB	7	2197	NO	7	1356			
EE	8	1959	GB	8	1882	NO	8	1465			
EE	9	1846	GB	9	2146	NO	9	1289			