



## OPEN ACCESS

EDITED AND REVIEWED BY  
Matthijs Bogaards,  
Central European University, Hungary

\*CORRESPONDENCE  
Jose A. Aleman  
aleman@fordham.edu

†These authors have contributed  
equally to this work

SPECIALTY SECTION  
This article was submitted to  
Methods and Measurement,  
a section of the journal  
Frontiers in Political Science

RECEIVED 08 September 2022  
ACCEPTED 04 October 2022  
PUBLISHED 21 October 2022

CITATION  
Aleman JA, Schmidt P, Meitinger K and  
Meuleman B (2022) Editorial:  
Comparative political science and  
measurement invariance: Basic issues  
and current applications.  
*Front. Polit. Sci.* 4:1039744.  
doi: 10.3389/fpos.2022.1039744

COPYRIGHT  
© 2022 Aleman, Schmidt, Meitinger  
and Meuleman. This is an open-access  
article distributed under the terms of  
the Creative Commons Attribution  
License (CC BY). The use, distribution  
or reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Editorial: Comparative political science and measurement invariance: Basic issues and current applications

Jose A. Aleman<sup>1\*†</sup>, Peter Schmidt<sup>2,3†</sup>, Katharina Meitinger<sup>4†</sup> and Bart Meuleman<sup>5†</sup>

<sup>1</sup>Department of Political Science, Fordham University, New York, NY, United States, <sup>2</sup>Department of Political Science, International Centre for Development and Environment (ZEU), University of Giessen, Giessen, Germany, <sup>3</sup>Department of Psychosomatics, University of Mainz, Mainz, Germany, <sup>4</sup>Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands, <sup>5</sup>Center for Sociological Research, KU Leuven, Leuven, Belgium

## KEYWORDS

political science, measurement invariance, multigroup confirmatory factor analysis, item response theory, mixture models

## Editorial on the Research Topic

## Comparative political science and measurement invariance: Basic issues and current applications

In the last decade, a growing number of researchers have become interested in applying new tools to verify the equivalence of measurements in comparative political science using mass surveys (Davidov, 2009; Ariely and Davidov, 2012; Coromina and Davidov, 2013; Alemán and Woods, 2016; Welzel and Inglehart, 2016; Sokolov, 2018). The increasingly available cross-national datasets offer tremendous possibilities for comparative survey analysis, including cross-sectional comparative analyses, analysis of cross-national repeated cross-sections, and analysis of cross-national panels. Many of these datasets can be linked to information about contextual attributes of the different countries and important economic, social, or political information (such as GNP, social spending, migration flow data, or religious composition) that facilitates multi-level analyses. A similar comparative logic can be applied to a lower level of aggregation as well, for example when regions or even smaller units within countries are compared.

In all these types of comparative analysis using different kinds of data, comparability of the measurements is a necessary condition to obtain valid results. There is a steadily growing literature on measurement equivalence, specifying the statistical prerequisites for comparing unbiased covariances, regression coefficients, and latent means in regression analysis, structural equation models, Item Response Theory (IRT) approaches, multi-level models, and latent class and mixture models (Jöreskog, 1971; Meredith, 1993; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000; Davidov et al., 2014, 2018; Kim et al., 2016; Verhagen et al., 2016; van de Vijver et al., 2019; Roover, 2021; Pokropek and Pokropek, 2022). This rather technical literature—that often focuses on statistical details and pays less attention to theoretical validity—has more recently been complemented by new approaches investigating how respondents interpret particular

items, for example by probing questions concerning the content of the items. This framework has been expanded in recent years by implementing the probing technique in web surveys (web probing), which results in much larger sample sizes compared to traditional face-to-face cognitive interviewing (Behr et al., 2017, 2020; Meitinger, 2017).

Since establishing the necessity of testing for measurement invariance, confirmatory factor analysis with multiple groups (MGCFA) has arguably been the most widely used technical tool to evaluate various levels—configural, metric and scalar—of measurement invariance for continuous variables (Brown, 2015). In the case of ordered-categorical items with few categories and a high degree of skewness, the ordinal approach to MGCFA is more appropriate (Brown, 2015; Liu et al., 2017). However, more recently, these tests of exact equivalence have been criticized for being too restrictive, often leading to the conclusion that comparisons should not be made even when cross-cultural differences are negligible (Zercher et al., 2015). To tackle this criticism, more liberal approaches have been developed for continuous variables—called approximate invariance—that allow comparisons of many groups and countries which would not be possible with the traditional approaches. A notable step in the direction of approximate rather than exact invariance is the application of Bayesian estimation in measurement models (Muthén and Asparouhov, 2012; van de Schoot et al., 2013; Davidov et al., 2015). In the case of dichotomous items, Item Response Models (IRT) have predominantly been used for this purpose. Additionally, Exploratory and Confirmatory Latent Class Analyses for multiple groups have been applied for the purpose of testing measurement equivalence. Another promising development is the use of multilevel regression models and structural equation multilevel models by combining individual-level data and higher-order level data, to explain why there is no metric or scalar invariance (Davidov et al., 2012; Jak et al., 2013; Jak and Jorgensen, 2017). All these procedures are grounded in the latent variable approach and make specific assumptions concerning the direction of relationships between latent variables and items. The models just mentioned assume reflective indicators, that is, indicators conceptualized as consequences (reflections) of the underlying latent variable. While this is an appropriate assumption in many cases, the literature also discusses examples of formative constructs (i.e., assuming that items determine the latent variable) (Sokolov, 2018; Stadler et al., 2021). This issue of reflective vs. formative indicators and the necessity of testing measurement invariance has been a point of critical discussion among political scientists (Aleman and Woods, 2016; Welzel and Inglehart, 2016; Sokolov, 2018; Welzel et al., 2021; Meuleman et al., 2022).

This Research Topic wants to inform political scientists about the state-of-the-art in this very fast-developing branch of survey methodology and statistics, where a lot of basic research has been done outside political science (e.g., in the fields of

psychometrics and statistics). To this end, it presents a collection of studies that apply the different techniques to central political science concepts (see Table 1 for a summary).

In the article “*The Relationship Between (sub)national Identity, Citizenship Conceptions, and Perceived Ethnic Threat in Flanders and Wallonia for the Period 1995–2020: A Measurement Invariance Testing Strategy*”, Billiet et al. study the relationship between (sub)national identity and perceived ethnic threat in Belgium and relate it to ethnic and civic citizenship conceptions in Flanders and Wallonia. They assess measurement invariance over time (1995–2020) using data from the Belgian National Election Studies. They find that conceptualization and measurement of (sub)national identity had to be adjusted in Wallonia, and illustrate how deviations from measurement invariance can be useful sources of information on social reality.

Maskileyson et al. assess in their article the measurement equivalence of self-reported depressive symptoms among the elderly in 17 European countries and Israel (“*The EURO-D Measure of Depressive Symptoms in the Aging Population: Comparability Across European Countries and Israel*”). They test measurement invariance of the EURO-D scale in the sixth wave (2015) of the Survey on Health, Aging and Retirement in Europe (SHARE) using multigroup confirmatory factor analysis (MGCFA) as well as alignment, and conclude that partial equivalence is present.

Scotto et al. discuss the issue of measurement invariance testing of ordinal scales with the example of political efficacy (“*Alternative Measures of Political Efficacy: The Quest for Cross-Cultural Invariance With Ordinally Scaled Survey Items*”). They propose to distinguish between internal and external efficacy. In representative samples of respondents in the United States and Great Britain, they find equivalence of loadings and thresholds for their measurement model and thus conclude that differences in latent variable means can be interpreted meaningfully. Concretely, British respondents are found to have lower levels of internal and external efficacy than American respondents.

Sokolov tests for measurement invariance of two recently introduced measures of attitudes toward democracy in the World Values Survey’s sixth round, the liberal and authoritarian notions of democracy. His analyses show that both measures can be considered reliable comparative measures of democratic attitudes, although for different reasons. Sokolov points out that some survey-based constructs, e.g., authoritarian notions of democracy, do not follow the reflective logic of construct development. Instead, Sokolov claims that these notions should be regarded as formative measures.

Heyder et al. propose a revised version of the Group Focused Enmity (GFE) syndrome as a two-dimensional concept: an ideology of inequality (generalized attitudes) and social prejudice (specific attitudes). The measurement models are empirically tested using data from the GFE panel (waves 2006, 2008) as well as the representative GFE surveys (cross-sections 2003, 2011) conducted in Germany. To test for external

TABLE 1 Overview of articles in this special issue\*.

Author and title	Topic, statistical method, goal of analysis	Countries and data
Billiet, J., Meeusen, C., and Abts, K. The relationship between (sub) national identity, citizenship conceptions, and perceived ethnic threat in Flanders and Wallonia for the period 1995–2020: a measurement invariance testing strategy	<b>Topic:</b> Relationship between (sub)national identity and attitudes toward immigrants in the multinational context of Belgium <b>Methods:</b> Measurement invariance testing strategy using confirmatory factor analysis <b>Goal of analysis:</b> Verifying whether the relationship between (sub)national units can be attributed to different conceptions of community membership	<b>Belgian</b> National Election Studies 1995–2020
Heyder, A., Anstötz, P., Eisentraut, M., and Schmidt, P. “20 years after. . .” GFE 2.0: a theoretical revision and empirical testing of the concept of “group-focused enmity” based on longitudinal data	<b>Topic:</b> Relationship between group-focused enmity, ideologies, attitudes, stereotypes, and social prejudice <b>Methods:</b> Measurement invariance testing strategy using higher-order confirmatory factor analysis <b>Goal of analysis:</b> Explaining group-focused enmity as a two-dimensional concept of ideologies of inequality and social prejudice	<b>German</b> Group-Focused Enmity (GFE) Panel Study (2006, 2008) and GFE cross-sections (2003, 2011)
Lomazzi, V. Can we compare solidarity across Europe? What, why, when, and how to assess exact and approximate equivalence of first- and second-order factor models	<b>Topic:</b> Analysis of the three dimensions of solidarity—social, local, and global—using data from 34 countries <b>Methods:</b> Multigroup confirmatory factor analysis, frequentist alignment approach <b>Goal of analysis:</b> Discussion of the implications of formative and reflexive approaches to measurement invariance	Last Wave of the European Value Survey from <b>34 European countries</b> (2017–2020)
Maskileyson, D., Seddig, D., and Davidov, E. The EURO-D measure of depressive symptoms in the aging population: comparability across European countries and Israel	<b>Topic:</b> Examine measurement invariance of self-reported depressive symptoms among older populations <b>Methods:</b> Multigroup confirmatory factor analysis and frequentist alignment approach <b>Goal of analysis:</b> Cross-country validation of existing depression measures using samples of the older population	Sixth wave (2015) of the Survey on Health, Aging, and Retirement in Europe (SHARE) from <b>17 European countries and Israel</b>
Scotto, T. J., Xena, C., Reifler, J. Alternative Measures of political efficacy: the quest for cross-cultural invariance with ordinally scaled survey items	<b>Topic:</b> Internal and external political efficacy among Americans and British <b>Methods:</b> Multi-Group confirmatory factor analysis for ordinally scaled survey items <b>Goal of analysis:</b> Comparison of two sets of efficacy indicators	Representative Surveys from <b>Great Britain and the United States</b>
Sokolov, B. Measurement invariance of liberal and authoritarian notions of democracy: evidence from the world values survey and additional methodological considerations	<b>Topic:</b> Review of the conceptual foundations of measurement invariance using an analysis of liberal and authoritarian notions of democratic attitudes <b>Methods:</b> Multi-Group confirmatory factor analysis, bayesian approximate measurement invariance, mgcfa alignment optimization <b>Goal of analysis:</b> An empirical illustration of the key concepts and methods from the MGCFa-measurement invariance literature	Sixth round of the World Values Survey from <b>60 countries</b>

\*We thank Oliver Platt for compiling this table.

validity, they have included a social dominance orientation (SDO). Additionally, the methodological focus of the study is to test for several forms of measurement invariance in the context of higher-order factor models considering the issue of multidimensionality of latent variables. The empirical results support the idea that GFE is a bi-dimensional concept consisting of an ideology of inequality and social prejudice. Moreover, SDO is demonstrated to be empirically distinct from both dimensions and correlates more strongly with the ideology of inequality in comparison to social prejudice. The bi-dimensional GFE conceptualization proves to be at least metric invariant both between and within individuals. Finally, the impact of the proposed conceptualization and empirical findings are discussed in the context of international research on ideologies, attitudes, and prejudices.

Finally, Lomazzi's study offers a theoretical overview of the key issues concerning the measurement and comparison of socio-political values and aims to answer questions of what, why, and when they must be evaluated, and how measurement equivalence can be assessed in practice. Furthermore, she discusses the implications of formative and reflective approaches to the measurement of socio-political values. Exact and approximate approaches to equivalence are described as well as their empirical translation into multigroup confirmatory factor analysis (MGCFA) and the frequentist alignment method. Her study investigates the construct of solidarity as measured by the European Values Study (EVS) and uses data collected in 34 countries in the last wave of the EVS (2017–2020). The concept is captured through a battery of nine items reflecting three dimensions of solidarity: social, local, and global. Two measurement models are hypothesized: a first-order factor model, in which the three independent dimensions of solidarity are correlated, and a second-order factor model, in which solidarity is conceived according to a hierarchical principle and the construct of solidarity is reflected in the three sub-factors. Employing MGCFA the results indicated that metric invariance was achieved. The alignment method supported approximate equivalence only when the model was reduced to two factors excluding global solidarity. The second-order factor model fits the data for only seven of the 34 countries.

In a nutshell, our conclusions from these studies and previous research are as follows:

1. Contrary to the position of Welzel and Inglehart (2016) and Welzel et al. (2021), these studies argue that one needs to reach at least partial metric invariance to get unbiased regression coefficients in comparative research involving two or more groups of countries (Meuleman et al., 2022; Pokropek and Pokropek, 2022).
2. If one wants to compare means one must employ latent means to correct for measurement error and must reach partial scalar invariance with at least two equal loadings and two intercepts of the same items (Meuleman et al., 2022; Pokropek and Pokropek, 2022).
3. If partial invariance fails one can use more liberal approximate techniques like Bayesian CFA (Seddig and Leitgöb, 2018) in the case of few groups (<10) and alignment in the case of many groups (>10) (Muthén and Asparouhov, 2014; Cieciuch et al., 2018).
4. The choice of model specification in the case of measurement models as reflective or formative must be founded on theoretical arguments as one cannot test them against each other in cross-sectional models. The reason is that the two model specifications are neither nested nor equivalent (Asparouhov and Muthén, 2019).
5. As measurement invariance is only a necessary but not sufficient condition it is advisable to employ additional cognitive interviews or web probing before the main study is executed (Meitinger, 2017; Behr et al., 2020; Meitinger et al., 2020).

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alemán, J., and Woods, D. (2016). Value orientations from the World Values Survey. *Comp. Polit. Stud.* 49, 1039–1067. doi: 10.1177/0010414015600458
- Ariely, G., and Davidov, E. (2012). Assessment of measurement equivalence with cross-national and longitudinal surveys in political science. *Euro. Polit. Sci.* 11, 363–377. doi: 10.1057/eps.2011.11
- Asparouhov, T., and Muthén, B. (2019). Nesting and equivalence testing for structural equation models. *Struct. Equat. Model. Multidiscipl. J.* 26, 302–309. doi: 10.1080/10705511.2018.1513795
- Behr, D., Meitinger, K., Braun, M., and Kaczmirek, L. (2017). *Web Probing—Implementing Probing Techniques From Cognitive Interviewing in Web Surveys With the Goal to Assess the Validity of Survey Questions (Version 1.0) (GESIS Survey Guidelines)*. Mannheim: GESIS—Leibniz-Institut für Sozialwissenschaften.
- Behr, D., Meitinger, K., Braun, M., and Kaczmirek, L. (2020). “Cross-national web probing: an overview of its methodology and its use in cross-national studies,” in *Advances in Questionnaire Design, Development, Evaluation and Testing*, eds P. Beatty, D. Collins, L. Kaye, J. L. Padilla, G. Willis, and A. Wilmot (Hoboken, NJ: Wiley) 521–543. doi: 10.1002/9781119263685.ch21
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research, 2nd Edn*. New York, NY: Guilford Press.
- Cieciuch, J., Davidov, E., Algesheimer, R., and Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European social survey. *Sociol. Methods Res.* 47, 665–686. doi: 10.1177/0049124117701478
- Coromina, L., and Davidov, E. (2013). Evaluating measurement invariance for social and political trust in western europe over four measurement time points (2002–2008). *Ask Res. Methods* 22, 37–54. doi: 10.5167/uzh-92161
- Davidov, E. (2009). Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. *Polit. Anal.* 17, 64–82. doi: 10.1093/pan/mpn014
- Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., and Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European social survey: exact versus approximate measurement equivalence. *Public Opin. Q.* 79, 244–266. doi: 10.1093/poq/nfv008
- Davidov, E., Cieciuch, J., and Schmidt, P. (2018). The cross-country measurement comparability in the immigration module of the European social survey 2014–2015. *Surv. Res. Methods* 12, 15–27. doi: 10.18148/srm/2018.v12i1.7212
- Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., and Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *J. Cross Cult. Psychol.* 43, 558–575. doi: 10.1177/0022022112438397
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., and Billiet, J. (2014). Measurement equivalence in cross-national research. *Annu. Rev. Sociol.* 40, 55–75. doi: 10.1146/annurev-soc-071913-043137
- Jak, S., and Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Front. Psychol.* 8, 1640. doi: 10.3389/fpsyg.2017.01640
- Jak, S., Oort, F. J., and Dolan, C. V. (2013). A test for cluster bias: detecting violations of measurement invariance across clusters in multilevel data. *Struct. Equation Model. Multidiscipl. J.* 20, 265–282. doi: 10.1080/10705511.2013.769392
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika* 36, 409–426. doi: 10.1007/BF02291366
- Kim, E. S., Joo, S.-H., Lee, P., Wang, Y., and Stark, S. (2016). Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. *Struct. Equation Model. Multidiscipl. J.* 23, 870–887. doi: 10.1080/10705511.2016.1196108
- Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., and Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychol. Methods* 22, 486–506. doi: 10.1037/met0000075
- Meitinger, K. (2017). Necessary but insufficient: why measurement invariance tests need online probing as a complementary tool. *Public Opin. Q.* 81, 447–472. doi: 10.1093/poq/nfx009
- Meitinger, K., Davidov, E., Schmidt, P., and Braun, M. (2020). Measurement invariance: test-ing for it and explaining why it is absent. *Surv. Res. Methods* 14, 345–349. doi: 10.18148/srm/2020.v14i4.7655
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825
- Meuleman, B., Zóltak, T., Pokropek, A., Davidov, E., Muthén, B., Oberski, D. L., et al. (2022). Why measurement invariance is important in comparative research. A response to Welzel et al. (2021). *Sociol. Methods Res.* 1–33. doi: 10.1177/00491241221091755
- Muthén, B., and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol. Methods* 17, 313. doi: 10.1037/a0026802
- Muthén, B., and Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Front. Psychol.* 5, 978. doi: 10.3389/fpsyg.2014.0978
- Pokropek, A., and Pokropek, E. (2022). Deep neural networks for detecting statistical model misspecifications. The case of measurement invariance. *Struct. Equat. Model. Multidiscipl. J.* 29, 394–411. doi: 10.1080/10705511.2021.2010083
- Roover, K. (2021). Finding clusters of groups with measurement invariance: unraveling intercept non-invariance with mixture multigroup factor analysis. *Struct. Equat. Model. Multidiscipl. J.* 28, 663–683. doi: 10.1080/10705511.2020.1866577
- Seddig, D., and Leitgöb, H. (2018). “Exact and bayesian approximate measurement invariance,” in *Cross-Cultural Analysis: Methods and Applications*, ed E. Davidov, P. Schmidt, J. Billiet, and B. Meuleman (New York, NY: Routledge), 553–570. doi: 10.4324/9781315537078-20
- Sokolov, B. (2018). The index of emancipative values: measurement model misspecifications. *Am. Polit. Sci. Rev.* 112, 395–408. doi: 10.1017/S0003055417000624
- Stadler, M., Sailer, M., and Fischer, F. (2021). Knowledge as a formative construct: a good alpha is not always better. *New Ideas Psychol.* 60, 100832. doi: 10.1016/j.newideapsych.2020.100832
- Steenkamp, J.-B. E. M., and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *J. Cons. Res.* 25, 78–90. doi: 10.1086/209528
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtj, P., Hox, J., and Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front. Psychol.* 4, 770. doi: 10.3389/fpsyg.2013.00770
- van de Vijver, F. J., Avvisati, F., Davidov, E., Eid, M., Fox, J.-P., Le Donné, N., et al. (2019). *Invariance Analyses in Large-Scale Studies. OECD Education Working Papers, No. 201*. Paris: OECD Publishing.
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3, 4–70. doi: 10.1177/109442810031002
- Verhagen, J., Levy, R., Millsap, R. E., and Fox, J.-P. (2016). Evaluating evidence for invariant items: a Bayes factor applied to testing measurement invariance in IRT models. *J. Math. Psychol.* 72, 171–182. doi: 10.1016/j.jmp.2015.06.005
- Welzel, C., Brunkert, L., Kruse, S., and Inglehart, R. F. (2021). Non-invariance? An overstated problem with misconceived causes. *Sociol. Methods Res.* 1–19. doi: 10.1177/0049124121995521
- Welzel, C., and Inglehart, R. F. (2016). Misconceptions of measurement equivalence: time for a paradigm shift. *Comp. Polit. Stud.* 49, 1068–1094. doi: 10.1177/0010414016628275
- Zercher, F., Schmidt, P., Cieciuch, J., and Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European social survey: exact vs. approximate measurement invariance. *Front. Psychol.* 6, 733. doi: 10.3389/fpsyg.2015.00733