Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling

Seddig, Daniel; Lomazzi, Vera

Postprint / Postprint
Zeitschriftenartikel / journal article

Accepted Manuscript

Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling

Daniel Seddig, Vera Lomazzi

PII: S0049-089X(18)30972-4

DOI: https://doi.org/10.1016/j.ssresearch.2019.102328

Article Number: 102328

Reference: YSSRE 102328

To appear in: *Social Science Research*

Received Date: 27 November 2018

Revised Date: 10 July 2019

Accepted Date: 15 July 2019

Please cite this article as: Seddig, D., Lomazzi, V., Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling, *Social Science Research* (2019), doi: https://doi.org/10.1016/j.ssresearch.2019.102328.

Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling

Daniel Seddig*
University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, Germany

Vera Lomazzi
Data Archive for the Social Sciences, GESIS, Cologne, Germany

* Corresponding author.

Daniel Seddig

University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology

Postal address:
University of Cologne, Institute of Sociology and Social Psychology
Albertus-Magnus-Platz
50923 Cologne, Germany

Email: dseddig@uni-koeln.de
Phone: +49/221/4703475

ORCID: 0000-0003-1595-6948

Vera Lomazzi

Data Archive for the Social Sciences, GESIS

Postal address:
GESIS – Leibniz Institute for the Social Sciences
Unter Sachsenhausen 6-8
50667 Cologne, Germany

Email: vera.lomazzi@gesis.org
Phone: +49/0221/47694479

ORCID: 0000-0003-2699-2768

Short biographies:

Daniel Seddig is a postdoctoral researcher at the Institute of Sociology and Social Psychology, University of Cologne, Germany. His research interests are cross-contextual and longitudinal data analysis with structural equation models, latent growth and mixture modeling, and Bayesian structural equation modeling. Substantive topics of his research are the value-attitude-behavior relationship, cross-cultural comparisons of attitudes, youth development, and deviance.

Vera Lomazzi is a senior researcher at the Data Archive for Social Sciences at GESIS – Leibniz Institute for Social Sciences in Cologne, Germany, and member of the Executive Committee of the European Values Study. Her substantive research mainly focuses on the cross-cultural study of gender equality and gender role attitudes, youth engagement, and collective identities. She has a specific interest in the quality of the instruments adopted by large cross-sectional survey programs and on their measurement equivalence.

Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling

Abstract

The current study explores the reasons for noninvariance of the measurements of gender role attitudes across countries. While previous studies have shown that noninvariance is a problem for comparative research and pointed out methods to alleviate the risks of drawing invalid conclusions, none has so far tried to explain why measurements of gender role attitudes are nonequivalent. Therefore, we use multilevel structural equation modeling to exploring measurement invariance and explain its absence. We use data assessing peoples views on the specialization of roles by gender and the consequences of female employment on family s well-being from the International Social Survey Programme. We can replicate the findings from prior research indicating that scalar measurement invariance across countries is absent. Furthermore, we use two country-level variables to explain the noninvariance of particular items. The cultural value embeddedness explains noninvariance to a considerable degree while the Gender Inequality Index from the United Nations Development Programme does not. Therefore, we conclude that issues of comparability of gender role attitudes are related mainly to cultural rather than structural differences between countries.

Keywords: gender role attitudes, comparative research, measurement noninvariance, multilevel structural equation modeling, cultural values, gender inequality

1. Introduction

The study of gender role attitudes attracts the interest of sociologists for several reasons. Attitudes toward gender roles are assumed as a proxy of the individual support for gender equality (Bergh, 2006). They refer to the underlying gender ideology, that is, the individual s level of support for a division of paid work and family responsibilities that is based on the notion of separated spheres (Davis and Greenstein, 2009, p. 89). Information concerning gender role attitudes has been collected since the 1970s also through several cross-national and repeated survey programs. The data collected allowed scholars to investigate the change of the individual support for traditional/egalitarian gender roles over time (Brewster and Padavic, 2000; Brooks and Bolzendahl, 2004; Kraaykamp, 2012; Kroska and Elman, 2009; Lomazzi, 2017; Mason and Lu, 1988; Misra, 1995; Scott and Braun, 2009) and make comparisons across countries (AndrØ, Gesthuizen, and Scheepers, 2013; Motiejunaite and Kravchenko, 2008; Scott and Braun, 2009; Sjöberg, 2004).

According to the literature (Kroska and Elman, 2009), individuals gender ideology is one relevant explanation in the actual household s arrangements concerning the division of paid and unpaid work between the partners. The measurement of gender role attitudes is therefore used also by scholars testing explicative models of the gendered division of the housework (Bianchi et al., 2012; Crompton et al., 2005; Fuwa, 2004; Geist and Cohen, 2011; Grunow et al., 2012). Finally, gender role attitudes are also used as predictor of societal elements, for example, by scholars investigating the effect of the individuals support for gender equality on female employment and other macro-level variables (Arpino et al., 2015; Cunningham, 2008; Stickney and Konrad, 2007).

The topic of gender role attitudes is relevant not only from a sociological perspective but also from a methodological one. Recently, scholars started being concerned about the quality of the measurements currently available to investigate gender role attitudes and about their suitability for comparative purposes (Braun, 1998, 2009; Constantin and Voicu, 2015; Lomazzi, 2018; AUTHOR, DATE; Walter, 2017). These studies related the lack of measurement invariance across countries with consequential risks of using of such measurements in cross-cultural studies. Measurement invariance is a relevant issue in comparative research. It refers to the basic assumption that the same object is compared across the different groups included in a study. In the case of survey research, this means assuming that respondents belonging to different groups perceive and interpreted items in the same way. However, methodological as well as cultural factors may cause equivalence biases and researchers risk to compare apples and oranges (Stegmueller, 2011). Noninvariance of measurements may then lead to substantial misinterpretations when comparing, for example, mean differences of gender role attitudes or correlations with other constructs across countries. The lack of measurement invariance can be alleviated using new and flexible approaches, such as alignment optimization (Asparouhov and MuthØn, 2014) or Bayesian structural equation modelng (BSEM) (MuthØn and Asparouhov, 2012). With these techniques, it is possible to ensure at least approximate invariance thus allowing for trustworthy comparisons of means. While the emergence of new methodologies provides applied researchers with tools to use large-scale cross-national surveys for comparative purposes and reduce the risk of drawing invalid conclusions, the pressing issue of *why* measurements are nonequivalent usually remains unsolved.

Therefore, in this article we provide, for the first time, an explanation of why measurements of attitudes toward gender roles are nonequivalent across countries. We apply a

multilevel structural equation modeling (MLSEM) strategy that uses country-level variables to account for differences in the measurement characteristics between countries. The country-level attributes we propose for explaining country bias in gender role attitudes are the cultural value embeddedness (EMB) and the gender inequality index (GII). We proceed by first introducing the theoretical framework of gender role attitudes. We then refer to the literature on its most common measurements and use in comparative studies. After describing the use of MLSEM to assess and explain measurement invariance, we continue with a description of the data for the current study. We then present our results and conclude with a summary and discussion.

2. Gender Role Attitudes: Theory and Measurement Issues

Gender role attitudes refer to the beliefs concerning the perceived appropriateness of social roles for men and women, in particular beliefs about the division of paid labor, childcare, and housework, on the basis of a gendered separation of tasks and responsibilities (Alwin, 2005; Davis and Greenstein, 2009; Schultz Lee et al., 2010). Traditional gender ideologies support the specialization of roles by gender. Historically, this means that social roles related to childcare and housework, and by extension, jobs in caring activities, are considered the most appropriate social roles for women. In contrast, roles in the public sphere connected to paid employment, authority, and also power positions by extension, are appropriate for men. Progressive attitudes toward gender roles tend to go beyond this gendered separation and express support for women s role in the public sphere as well as the men s role in the private one (Albrecht et al., 2000; Baxter and Kane, 1995; Cunningham et al., 2005).

Looking at the factors explaining differences in gender ideologies at the individual level, previous studies (AndrØ et al., 2013; Baxter and Kane, 1995; Bolzendahl and Myers, 2004;

Kroska and Elman, 2009; Scott et al., 1996; Sjøberg , 2004) provide consistent results. People that are young, more educated, who have access to higher economic resources, and live in urban contexts tend to express more egalitarian attitudes. For women, their participation in the labor market increases their support for progressive gender roles. On the opposite, people living in an institutionalized partnership with children and people who are integrated into a religious community tend to support traditional gender roles.

From a societal point of view, the differences across countries can be interpreted according to structural and cultural explanations. Countries differ in their demographic and educational structure, in their economic and technical development, and generally speaking in their opportunity structure. According to Inglehart s (1997) post-materialism theory, values change as a reflection of the modernization process. The economic and technological innovations bring changes in the behaviors and shape values and attitudes. The increase in female participation in the labor force is one of the biggest changes in the structure of the western societies, which also has an impact on the family structure and patterns of the division of paid and non-paid work between the partners. Several studies documented the impact of this structural change on individual gender role attitudes (Cunningham, 2008; Inglehart and Norris, 2003; Kraaykamp, 2012; Kroska and Elman, 2009; Motiejunaite and Kravchenko, 2008). These changes are directly connected with a change in the demand for welfare and family policies, which may also affect gender ideologies. Institutional factors, such as work-family balance policies, are relevant predictors of gender role attitudes: The availability of childcare services and parental leave schemes shapes the opportunity structure, affecting, therefore, individuals opinions (AndrØ et al., 2013; Kangas and Rostgaard,2007; Lomazzi et al., 2019; Sjøberg, 2004). Differences in societal structures can be considered, for example, along dimensions like the

demographic composition, educational level, economic and political performances within countries. Furthermore, several composite indicators are provided by transnational agencies to grasp the different structures. With a particular focus on gender equality, that is the substantive core issue of this study, objective measures of gender equality and of the status of women can provide insightful contribution to the understanding of the difference between countries and whether measurement biases can be rooted in such structural differences. For example, the United Nation Development Programme developed several indicators of gender equality in the conceptual framework of human development (Gaye et al., 2010; Klasen and Schüler, 2011; Permanyer, 2013). The most updated one allows for grasping differences across societies in the gendered access to health, political empowerment and economic participation. These aspects refer of structural differences of opportunity structures and gender role segregation in society. According to this index, a deep variety of structural gender equality exist (see Table 2): considering the countries included in the current study, structural inequality score (where 0 means equality) ranges from 0.466 (Venezuela) to 0,045 (Netherlands).

Alongside the different societal structures, countries also differ in their cultures and in particular in the societal value orientations. Researchers adopting cultural explanations shed light on the strict relation between attitudes toward gender roles and the broader value system that is predominant in society. For example, previous studies indicated the role of religious values and of the process of secularization on gender ideologies and their change over time (Inglehart, 1997; Lussier and Fish, 2016; Moore and Vanneman, 2003). Most religions support the gendered separation of roles, and the secularization process is believed to promote the liberalization of gender norms and the erosion of ascriptive principles and the value of self-realization (Inglehart and Norris, 2003; Kalmijn, 2003).

In addition to Inglehart s post-materialism theory (1997) two other approaches aimed at mapping cultural contexts have particular success in political, sociological and social psychology studies. Developing further the basic assumption of Inglehart s theory, Welzel (2013) proposes his evolutionary theory of emancipation. By combining a specific set of 12 items from the World Values Study (WVS), Welzel builds the Emancipative Values Index that aims at measuring people s support for freedom of choice and equality of opportunities in the domain of equity values (priority to gender equality over patriarchy), liberty values (priority to reproductive freedom over its restriction), autonomy values (priority to self-determination over obedience), and expression values (priority to give voice to people over order).The Emancipative Values Index, which is based on the WVS, is able to cover an extensive number of countries and allows cross-time comparisons. Substantively, it assumes a path-dependency approach, which links countryspecific cultures to their historical development and covers relevant values domains able to summarize the cultural differences across societies. For these reasons, Welzel s fascinating theory has found a broad consensus and has been used in many studies, also investigating gender equality in comparative perspective (Alexander and Welzel, 2015; Brieger et al., 2019). However, some aspects of Welzel s theory have been also criticized. The theory is based on the assumption that values are interpreted in the same way in all the cultures (Welzel, 2013 p. 41-43). As Sokolov (2018) points out, this assumption supports the idea that people who belong to different cultures give the same meaning to items aimed at investigating values in surveys and the country mean scores can be compared without any risk of bias. Studies in the field of comparative survey research (Davidov, Schmidt, & Billiet, 2011; Jowell, 1998; Stegmueller, 2011) argue that such a-priori assumptions of cross-cultural comparability need to be tested. In the case of the Emancipative Values Index, recent studies (AlemÆn & Woods, 2016; Sokolov,

2018) demonstrated that such an index is not invariant across all the countries, but it is equivalent across a subsample of Western post-industrial societies. An alternative approach is provided by Schwartz s theory of cultural value orientations (1994, 2006). Schwartz s approach is based on a theoretical conceptualization of values and of their relations as a system, which is then validated by empirical data. In contrast to Inglehart s perspective, which prioritizes the role of the economic change, and to Welzel s approach, which gives emphasis to the pursuit of freedom and democracy, Schwartz s theory provides a broader value spectrum concerning the shared conceptions of what is good and desirable in the culture (Schwartz, 2006, p.139). His theory assumes seven cultural value orientations: harmony; embeddedness; hierarchy; mastery; affective autonomy; intellectual autonomy; egalitarianism. Accordingly, cross-country differences in the attitudes toward gender roles can be linked to the differences in value priorities. Among these, gender role attitudes appear to be connected to embeddedness, hierarchy, and egalitarianism. In particular, AUTHOR (DATE) show that gender ideologies strongly correlate with the cultural value embeddedness. This value refers to the view that people are deeply embedded in the collectivity of society rather than autonomous entities. Societies, in which this value orientation is prevailing, emphasize intra-group social relationships and tradition, both reinforcing identification with the group. On the contrary, where embeddedness is not predominant, individuals are considered as autonomous-bounded entities that are expected to favor own preferences instead of those of the collective. Important individual level values that are the foundation of the cultural value orientation embeddedness are social order, respect for tradition, security, obedience, and wisdom (Schwartz, 2006, p. 140). Furthermore, societies in which people are strongly embedded in their collective consider maintaining the status quo important and people tend to express traditional gender role expectations as part of a traditional

system to be preserved. Based on a dimensional logic, the measurements developed on the basis of Schwartz s theory have been proved to be equivalent across societies (Schwartz, 1999, 2006) and offer therefore a valuable perspective to consider cultural differences at the societal level. Societies substantively differ in their expression of the value embeddedness, ranging from Sweden, the less embedded country (3.12), to Philippines and South Africa (4.03), the most embedded countries. To sum up we can state that societies around the world differ by their societal structure and prevailing cultural orientations. Moreover, societies display different gender roles and related gender role attitudes. However, can structural and cultural variability affect the measurement of gender role attitudes? In other words, to what extent are differences in gender role attitudes evidenced in survey research due to structural and/or cultural bias?

So far, little is known about the potential effect of different cultural structures on measurement equivalence. However, we know from previous studies (e.g., Braun 2008) that the measurement of gender role attitudes is particularly sensitive to cultural differences. Furthermore, access to resources and societal opportunity structures may provide respondents with different elements to be considered when expressing their gender role attitudes. Thus, social and cultural differences may be responsible for differential response patterns toward gender specific attitudes and therefore elicit biased responses to items although real differences in the underlying construct are absent. For example, the available childcare possibilities, average female labor market participation in a country, and cultural notions about embeddedness may affect the way people interpret, judge, and express opinions concerning the widely used item A preschool child is likely to suffer if his or her mother works .

However, at what point can social and cultural differences explain anything above and beyond the real differences that exist in the underlying construct across countries? First, many

instruments used by cross-sectional survey programs to measure gender role attitudes were developed in the 1970s in the U.S. and Western European countries and reflect the prevailing family patterns and views on gender roles of those contexts. Nowadays, respondents from different countries can perceive the wording used in such instruments differently (Braun, 1998, 2008, 2009) because of historically different developments of gender cultures (Pfau-Effinger, 2004), which are linked to different societal structures and general value orientations across countries. Second, ambiguity-based framing effects (Braun, 2009) can explain why semantically equivalent questionnaire items may be interpreted differently across countries, causing nonequivalence. Accordingly, such differences are to a large extend based on interactions between the characteristics of the question and the particular cultural and socio-economic background of the respondents. Although the cognitive and motivational foundations of response behavior may be the same across countries and cultures, respondents perceptions of the societal and cultural realities in their respective contexts may render the interpretation of questions. Even if the perceived realities hardly overlap with the content of questionnaire items, interpretations will be guided by the information provided by the social and cultural context. This is especially the case when item formulations leave room for ambiguities, for example, when items are formulated vaguely and lack enough information to be properly understood. As an example, consider the item A working mother can establish j ust as warm and secure a relationship with her children as a mother who does not work . The in formation may be deemed ambiguous with regard to several aspects, such as the capabilities of the mother, the family and marital situation, and the needs of the child. Braun (1998) investigated differences in the item interpretation between German and British respondents and found that Anglo-Saxons tend to evaluate the needs of the child, while Germans assess the mother s capabilities. He relates the results to the

different cultural orientations providing a different framework for interpreting the question. Thus, when specific information is missing, which is otherwise needed for activating an underlying attitude position, respondents tend to fill the informational gaps with knowledge that is specific to their societal and cultural background. Furthermore, the degree to which intercultural differences will affect the interpretation of questions depends on the salience of concepts in a country. Immediate accessibility of social and cultural information related to a particular concept enhances the chance that respondents more readily (and automatically) fill the gaps with the accessible information from their background.

Based on these insights, researchers who adopt a comparative approach to the study of gender role attitudes have become concerned about the threats to comparability caused by cultural sensitivity. Scholars have begun to investigate such biases by addressing the issue of equivalence as a basic step prior to any meaningful cross-cultural comparison of gender role attitudes (Constantin and Voicu, 2015; Lomazzi, 2018; van Vlimmeren et al., 2016; AUTHOR, DATE). While these studies pointed out the lack of equivalence across countries, the possible reasons for this bias have not been investigated yet.

3. Measurement Invariance and Multilevel Structural Equation Modeling

A prerequisite for valid comparisons of correlations, regression coefficients, or means across cultures (or groups in general) is that the measures of the theoretical construct are comparable. Therefore, it is essential to test empirically whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute (Horn and McArdle, 1992, p. 117). Measurement invariance means that items in a measurement instrument are perceived and used in a similar way across cultures (Byrne and van

de Vijver, 2010). In other words, measurement invariance ensures that respondents with the same true value on an underlying latent dimension have the same expected score on an observed variable, irrespective of the group or culture they belong to. In the case of noninvariance, other variables may systematically influence the observed items scores and thus bias comparisons across cultures.

Many methods exist to test measurement invariance (see Davidov et al., 2018b). One of the most frequently used methods is multiple-group confirmatory factor analysis (MGCFA) (Bollen, 1989; Jöreskog, 1971). MGCFA allows testin g a sequence of more or less restrictive measurement models that differ with regard to the constraints on the measurement parameters across groups (Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000). In cross-cultural research, usually three levels of measurement invariance are distinguished. Configural invariance represents the invariance of the model structure across groups. This means that the same items in each group measure the same factors. Metric invariance refers to the equality of the factor loadings across groups and is a prerequisite for the comparison of factor (co)variances or regression coefficients. Scalar invariance requires that, in addition to the model structure and factor loadings, the item intercepts are equal across groups. When scalar invariance holds in the data, factor means can be compared across groups. Testing the least restrictive (configural) against the more restrictive (metric and scalar) models can be done by comparing model fit statistics (Chen, 2007). In many large-scale comparative studies, scalar measurement invariance often lacks empirical support suggesting that mean comparisons of theoretical constructs across countries or cultures are precluded. Davidov et al. (2018a) even claimed that "scalar noninvariance constitutes one of the most serious threats to cross-cultural research" (p. 5).

Various strategies deal with measurement noninvariance. While some authors suggested ignoring it (e.g., Welzel and Inglehart, 2016; Brunkert et al., Forthcoming), others proposed methods that aim at reducing it or relaxing the strict equality requirements of MGCFA. For example, researchers may exclude groups that are noninvariant or free noninvariant parameters from equality constraints. The exclusion of groups can seriously limit the substantive scope of a study, because important information may be ignored. Freeing parameters from equality constraints may yield partial measurement invariance (Byrne et al., 1989; Steenkamp and Baumgartner, 1998), meaning that some (at least two) indicators of a factor are invariant across groups. While some author argued that partial invariance is sufficient for drawing valid comparisons (Byrne et al., 1989; Steenkamp and Baumgartner, 1998), others suggest that this is not the case (de Beuckelaer and Swinnen, 2018; Steinmetz, 2018).

Recent approaches aim at relaxing the strict criteria of exact equality of measurement parameters across groups and replace them with approximate equality constraints (MuthØn and Asparouhov, 2013). Approximate measurement invariance can be assessed with the alignment procedure (Asparouhov and MuthØn, 2014; Cieciuch etal., 2018; Marsh et al., 2017; Munck et al., 2017; MuthØn and Asparouhov, 2014) or BSEM (Cieciuch et al., 2014, 2017; MuthØn and Asparouhov, 2013; Seddig and Leitgöb, 2018a, 2018b; van de Schoot et al., 2013).

So far, scholars mainly assessed the measurement invariance of gender role attitudes adopting MGCFA (Constantin and Voicu, 2015; van Vlimmeren et al., 2016). Even if there are still few studies in this field, all of them come to the conclusion that when assessing measurement invariance of gender role attitudes across a large number of countries, the attempt fails regardless of the survey analyzed. More recent studies (Lomazzi, 2018; AUTHOR, DATE) have applied the novel alignment optimization procedure (Asparouhov and MuthØn, 2014) and

proposed this technique as a potential alternative to MGCFA. In fact, it allows overcoming the limits of MGCFA, such as its strictness and the arbitrary pathway to establish partial invariance (Asparouhov and MuthØn, 2014; Davidov et al., 2008,2014; van de Schoot et al., 2013), and provides reliable factor means for comparisons of a larger number of groups.

The application of new procedures to assess invariance is relevant for the future of survey research and both strategies suggest amending the degree of noninvariance until meaningful comparisons are deemed appropriate. However, they lack an explanation as to *why* measurement invariance is absent. Such information can be important to broaden our understanding of measurement differences across countries and for scale construction in future research. Therefore, several authors have proposed to use multilevel SEM (MLSEM) to *explain* measurement noninvariance (Davidov et al., 2012, 2018a; Jak, 2014, 2017; Jak et al., 2013, 2014a, 2014b). We do not claim that MLSEM is superior in any way over MGCFA in testing measurement invariance. However, it is superior in *explaining* noninvariance by using macro-level variables, which are derived from theoretical assumptions.

In multilevel confirmatory factor analysis (Cheung and Au, 2005; Hox, 2010; MuthØn, 1985, 1994) the variability of the indicators is decomposed into within-level variability (e.g., across individuals) and between-level variability (e.g., across groups). Suppose a two-level confirmatory factor analysis (see MuthØn, 1991, p.344):

Level 1 (within): $y_{ijk} = \alpha_{jk} + \lambda_{Wk}\eta_{Wij} + \varepsilon_{Wijk}$ Level 2 (between): $\alpha_{jk} = \nu_k + \lambda_{Bk}\eta_{Bj} + \varepsilon_{Bjk}$ (1)

On the within-level, y_{ijk} is the observed value of individual *i* in group *j* on indicator variable *k*, α_{jk} is the intercept of indicator variable *k* in group *j*, λ_{Wk} is the within-level factor loading of indicator variable *k*, η_{Wij} is the score of individual *i* of group *j* on the within-level factor, and ε_{Wijk}

is the within-level error term for individual *i* in group *j* on indicator variable *k*. The key to connect the within- and between-levels is the group-specific item intercept α_{jk}, which constitutes the information used on the left-hand side of the between-level equation. Furthermore, the between-level consists of υ_k, which is the grand intercept of indicator variable *k* across groups, λ_{Bk}, which is the between-level factor loading of indicator variable *k*, η_{Bj}, which is the score of group *j* on the between-level factor, and ε_{Bjk}, which is the between-level error term (or random intercept) for group *j* on indicator variable *k*.

In a multilevel framework, metric measurement invariance is achieved when the factor loadings are equal across levels ($\lambda_{Wk} = \lambda_{Bk}$) (Jak et al., 2013). Equal factor loadings ensure that the common factor has the same meaning on the within- and the between-level (Guenole, 2016; Tay et al., 2014).[1] Scalar measurement invariance is achieved when, in addition to equal factor loadings across levels, the residual variance at the between-level is zero ($\varepsilon_{Bjk} = 0$) (Jak et al., 2013). When scalar measurement invariance holds, all observed differences between groups are due to differences in the common factor. When scalar measurement invariance does not hold, the residual variance at the between-level will be significantly larger than zero. The implication is that variables other than the common factor may be responsible for the differences in the indicators across groups. Therefore, if the residual variance is unaccounted for, any comparison of means across groups may be biased. In other words, when the error variances substantially deviate from zero, the means of the group-specific items deviate from what can be expected given the between-level factor mean. In MLSEM, observed between-level variables can be added

[1] Some authors suggest that this assumption is too strict for many sociological application (Davidov et al., 2018a) and that aim of the multilevel approach is different model to a MGCFA (Chen, et al., 2005; Kim et al., 2015). We agree with their position. However, we begin our analysis with the more parsimonious approach and constrain the loadings to be equal across levels.

to explain differences in the common factor and group-specific differences in specific items. Therefore, Equation (1) is extended (e.g., Davidov et al., 2018a):

Level 1 (within): Level 2 (between): (2)

$$y_{ijk} = \alpha_{jk} + \lambda_{Wk}\eta_{Wij} + \varepsilon_{Wijk} \qquad \alpha_{jk} = \nu_k + \lambda_{Bk}\eta_{Bj} + \sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{kl} * z_{jl} + \varepsilon_{Bjk}$$

$$\eta_{Bj} = \mu_B + \sum_{l=1}^{L}\gamma_{l0} * z_{jl} + \zeta_{Bj}$$

On the between-level, z_{jl} refers to score of group *j* on between-level predictor variable *z*. The coefficients γ_{kl} and γ_{l0} refer to the effects of the between-level predictor variable *z* on the group-specific item intercept α_{jk} and the between-level factor η_{Bj}, respectively.

We adopt an already well established modeling strategy proposed by Jak (2017), which involves several consecutive steps. First, we are interested in whether differences on the observed variables exist between countries (Step 1). Thus, we test whether sufficient variance can be found at the between-level to vindicate the use of multilevel techniques. This involves testing a model that is saturated on the within-level (all indicator variances and covariances are estimated) and a null model on the between-level (all indicator variances and covariances are fixed to zero). Since the within-level is just identified, any lack of model fit can be ascribed to the restrictions on the between-level. Furthermore, the intraclass correlations (ICC) quantify the amount of between-group variation. Then we fit the assumed measurement model to the data on the within-level and additionally fit a saturated model on the between-level (Step 2). Next, a two-level model with equal factor loadings across levels and zero residual variances at the between-level is estimated to test whether scalar measurement invariance holds (Step 3). Given the results from prior research, we expect that this is not the case. Therefore, the residual variances at the between-level are estimated to test whether they differ from zero. In this step, we can also

determine proportion of the variation on the between-level that is residual variation (Jak, 2014). Finally, we add observed between-level variables (the cultural value embeddedness [EMB] and the gender inequality index [GII]) to explain variation in the common factor and group-specific variation in specific items (Step 4).

4. Measurements and Data Sources

Several cross-sectional surveys, such as European Social Survey (ESS), European Values Study (EVS), World Values Survey (WVS), and International Social Survey Programme (ISSP) include measurements of gender role attitudes. Despite this wide offer, these surveys present some limits for our study. Compared to the other measurements, the one included in the ESS provides a poorer coverage of the concept (only one item investigates gender beliefs). The most recent full dataset of EVS is quite old (2008) and the currently available data from the last wave (2017) have been released only for 16 countries. The geographical coverage is bigger in the last wave of ISSP (36 countries) and WVS (60 countries). WVS measurement invariance has been already assessed in previous research (Lomazzi, 2018). Furthermore, according to the theoretical conceptualization of gender role attitudes (Davis and Greenstein, 2009; Walter, 2017), ISSP offers a better operationalization. For the present study, the most recent wave of ISSP (2012) therefore represents a valuable compromise, allowing a good conceptualization and coverage of cultural contexts. Moreover, the current research is a follow up of a previous study (AUTHOR, DATE) that also employed the same measurement from ISSP.

For more than 30 years, the ISSP has been collecting information on individual behaviors, preferences, opinions, and attitudes among population samples across the world through the implementation of thematic modules. In 1988, 1994, 2002, and 2012, information on

several gender issues has been collected by the module Family and Changing Gender Roles . We use the measurement of gender role attitudes included in the most recent edition (ISSP Research Group, 2016). Attitudes toward gender roles are investigated by a scale asking the respondents to express their agreement with seven statements[2], mainly concerning the specialization of roles by gender and the consequences of female employment on family s well-being. Respondents could rate their agreement from 1 (*strongly agree*) to 5 (*strongly disagree*). Table 1 shows the items and their descriptive statistics.

Table 1 about here

Regardless of the widespread use of this measurement in several studies (Scott et al., 1996; Sjöberg, 2004; Stickney and Konrad, 2007 ; Motiejunaite and Kravchenko, 2008; Braun, 2009), the investigation of gender role attitudes through such items can be controversial. These items have been criticized (Braun, 1998) for producing conceptual problems, for example, by introducing two aspects in the same questions (as in items *warmrelation* and *childsuffers*) to which respondents can react. Other items, such as item *childsuffers*, are so strictly related to the current availability of contextual resources for childcare to potentially introduce a bias related to the variability of existent institutional provisions in different countries. Again, items with a strong traditionalist wording (as item *womensjob*) can be interpreted differently across countries which developed different gender cultures, and respondents living in more egalitarian societies could reject

[2] The seven items are displayed in the questionnaire by two batteries. The first battery contains the items *warmrelation*, *childsuffers*, *familysuffers*, *family&kids*; the second contains the items *bothincome* and *womensjob*.

the statements, while those living in countries with a more traditional gender culture could find it in line with their experience (Braun, 2008). These issues suggest that first of all the measurement invariance of the ISSP gender role attitudes scale needs to be assessed. Furthermore, the criticism articulated by Braun (1998, 2008, 2009) raise the issue whether these items are affected by cultural bias which compromises the measurement invariance across different countries.

Following the results of an earlier exploratory factor analysis (AUTHOR, DATE), our measurement includes only the items with the highest factor loadings: *childsuffers* (0.74), *familysuffers* (0.75), *home&kids* (0.58), *womensjob* (0.59). Although the loading of *warmrelation* is 0.43, a deeper country-by-country investigation reveals that this item shows very poor factor loadings (below 0.3) in many countries. We therefore decided that the working mother domain is sufficiently captured by the stronger loadings of items *childsuffer* and *familysuffers displayed in most of the countries*. All other factor loadings were below 0.40.

In our investigation of the cultural bias that may affect the measurement invariance, we consider two elements related to the structural and the cultural explanation of the variability of gender role attitudes across societies. The Gender Inequality Index (GII) was developed in 2010 by the United Nations Development Programme (UNPD) with the aim of measuring the gender inequalities in three domains of human development: reproductive health, political empowerment, and economic participation. As an index to purely measure gender inequality, the GII has been criticized because, considering how it is constructed, it neglects some other areas where gender inequality takes place, maintaining its focus on incorporating the gender perspective in the measure of the human development more than defining gender inequality. In fact some dimensions, such as health, include indicators valid only for women, while a comparison between men s and women s conditions is taken in account in other dimensions (for

an overview of the debate on this index, see Gaye et al., 2010; Klasen and Schüler, 2011; Permanyer, 2013). However, for the same reasons, this index can be used to grasp societal structural aspects related to gender inequality. The GII is in fact built by computing the maternal mortality ratio and adolescent birth rates (health dimension); proportion of parliamentary seats occupied by females and proportion of adult females and males aged older than 25 years with at least some secondary education (political empowerment); economic participation of women compared to that of men in the country. We therefore use the GII 2012 (Malik, 2013)[3] as an indicator of the structural variability across countries, which can explain the lack of measurement invariance. GII scores range from 0, which indicates equality, to 1, meaning the highest inequality possible.

To include in our analytical framework the cultural explanation, we refer to Schwartz s (2006) theory of cultural values. We focus our attention on the cultural value of embeddedness (EMB), which shows the strongest correlation with gender role attitudes (AUTHOR, DATE). The embeddedness scores are provided by Schwartz (2008), who reports on a study of value orientations based on the administration of the 56- to 57-item Schwartz Value Survey (SVS) among schoolteachers and college students from 58 and 64 national groups, respectively, between 1988 and 2007 (for details see Schwartz [2006, p. 145]). The 56-57 values were measured on a 9-point scale (-1 to +7) and subsequently combined to seven cultural value orientations. Thus, higher values indicate higher degree of embeddedness. Furthermore, each country s score was centered around the group mean score. Thus, a cultural value score within a group represents the importance of the

[3] Taiwan is not recognized by the United Nations. We used the score provided by the National Statistics of Taiwan (China), which has calculated the index independently by adopting the same formula of GII. Source: https://eng.stat.gov.tw

value relative to the other values (Schwartz, 2008). In our analysis we include the country mean scores (scores not available for Iceland and Lithuania).

Our study concerns the 34 countries for which all the measurements are available. Table 2 provides an overview of variation of the GII and EMB scores by country. Gender inequality connected to structural aspects is greater in Venezuela, South Africa, and the Philippines. In the Netherlands, Taiwan, and Sweden, the gender disadvantages are the lowest. South Africa, the Philippines, and Croatia display the highest levels of embeddedness in the collectivity and consider the maintenance of the status quo to be very important. Germany and Sweden show the lowest embeddedness. According to Schwartz s theory (2006), these countries, together with the Netherlands, Denmark, and Switzerland (all at third position), place less emphasis on the maintenance of the traditional order.

Table 2 about here

5. Results

We used the software package Mplus Version 8 (MuthØn and MuthØn, 1998-2017) in conjunction with a maximum likelihood estimator that provides adjusted and robust standard errors (MLR). For model evaluation, we inspect the chi-square fit statistics as well as the comparative fit index (CFI) and the root mean error of approximation (RMSEA). A CFI above 0.95 and an RMSEA below 0.05 indicate good fit of the model to the data (Hu and Bentler, 1999; Marsh et al., 2004; West et al., 2012). A CFI between 0.90 and 0.95 and an RMSEA between 0.05 and 0.08 are still acceptable.[4]

[4] Upon request, the authors will provide all Mplus syntax and a description of the data preparation as supplementary material.

Step 1. We tested whether variation on the observed variables exists between countries. The model with a saturated within-level structure and a fully constrained between-level structure had a significant very large chi-square test statistic: $\chi\dagger = 22466.0$ ($df = 10$). The ICC statistics for the items were $ICC_{childsuffers} = 0.147$, $ICC_{familysuffers} = 0.129$, $ICC_{home\&kids} = 0.146$, and $ICC_{womensjob} = 0.218$. Thus, sufficient variation exists on the between-level, and the use of multilevel techniques seems appropriate.

Step 2. Next we fitted the measurement model to the within-level and a saturated model to the between-level. The model fit was not adequate: $\chi\dagger = 716.4$ ($df = 2$), CFI = 0.865, RMSEA = 0.078. Modification indexes suggested adding a residual covariance between items *home&kids* and *womensjob*. This corresponds with earlier findings (AUTHOR, DATE) and is interpreted as a similarity of both items with regard to the domain of home life. We considered the residual covariance as an additional latent variable with both factor loadings (for *home&kids* and *womensjob*) fixed to 1.0 and no correlations with other items or factors (Jak et al., 2014a). The variance of the additional factor corresponds to the residual covariance of the items. Thereby, we avoided altering the model structure across levels. For example, testing for scalar invariance requires fixing the residuals variances on the between-level to zero. The model with the additional factor fitted the data well: $\chi\dagger = 8.0$ ($df = 1$), CFI = 0.999, RMSEA = 0.011. Standardized within-level factor loadings were $\lambda_{Wchildsuffers} = 0.710$, $\lambda_{Wfamilysuffers} = 0.800$, $\lambda_{Whome\&kids} = 0.448$, and $\lambda_{Wwomensjob} = 0.428$.

Step 3. We tested for scalar measurement invariance by constraining the factor loadings to be equal across levels and fixing the residual variances at the between-level to zero. This model assumed that all country-mean differences of the factor within the countries are accounted for by the common factor at the between-level. However, the model did not fit the data,

indicating that a considerable proportion of the country-mean differences of the factor within the countries are *not* accounted for by the common factor at the between-level: $\chi^2 = 3918.8$ ($df = 9$), CFI = 0.259, RMSEA = 0.086. In other words, scalar measurement invariance does not hold. The proportion of country-mean differences that is left unaccounted for is assumed to be present as residual variance at the between-level. Thus, we estimated the residual variances and the model fit improved: $\chi^2 = 37.8$ ($df = 5$), CFI = 0.994, RMSEA = 0.011. Significant residual variances were found for items v6 and v11. The proportion of the variance on the between-level that is residual variance (i.e., bias) was 0.187 (*childsuffers*), 0.043 (*familysuffers*), 0.017 (*home&kids*), and 0.609 (*womensjob*). Thus, the item intercepts for *childsuffers* and *womensjob* show considerable variation across countries, a finding that precludes valid comparisons.

Step 4. We used GII and EMB to explain the residual variances of *childsuffers* and *womensjob* and correlated them with the between-level common factor. The first model revealed an additional significant residual covariance of item *home&kids*. This item was also identified as being noninvariant in prior research (AUTHOR, DATE). Thus, we included it also here and explained its residual variance in the final model. This model fit the data well: $\chi^2 = 24.1$ ($df = 5$), CFI = 0.998, RMSEA = 0.008. The GII had no significant effects on the residual variances, but was positively correlated with the common factor on the between-level (Table 3). Thus, positive attitudes toward traditional roles of women coincide with a systematic structural disadvantage of women. However, the GII does not account for the proportion of the country-mean differences that is not accounted for by the common factor at the between-level (i.e., measurement noninvariance). The value EMB had positive and significant effects on the residual variances and is positively correlated with the common factor on the between-level. Thus, positive attitudes toward traditional roles of women are related to a value preference that is concerned with

maintaining the status quo and the traditional order. Furthermore, EMB can explain at least to a certain degree measurement noninvariance in *childsuffers*, *home&kids*, and *womensjob*. For equal levels of gender role attitudes between countries, more embedded countries have higher scores on variables that express disapproval with working mothers (*childsuffers* and *home&kids*) and approval of the traditional divide of domains (*womensjob*). Thus, noninvariance of the three items rests on a cultural explanation rather than a structural explanation. Figure 1 shows a graphical representation of the final model.

Table 3 about here

6. Summary and Conclusion

Gender role attitudes attract the interest of sociologists for theoretical and methodological reasons. From a theoretical perspective, attitudes toward gender roles are important predictors and proxy variables of the individual support for gender equality (Bergh, 2006). Such attitudes refer to the individual's preferences for separate life spheres with regard to the division of paid work and family responsibilities. From a methodological perspective, the measurement of gender role attitudes has to be in accordance with the requirement for invariance across countries in order to make valid comparisons. In case measurement invariance across countries is absent or not tested at all, there is a risk of drawing invalid and biased conclusions or simply being uninformed whether or not constructs may be compared. However, strict measurement invariance may be rather unrealistic in many situations and recent approaches (e.g., alignment or BSEM) have been proposed to soften the strict measurement invariance paradigm. However, none of the MGCFA based methods can provide an answer to the important question of *why*

measurements are nonequivalent. Offering substantive information as to why people from different (social, cultural, or economic) backgrounds respond differently to survey questions is vital to improve the quality of cross-cultural and survey research in general. Therefore, we applied an MLSEM approach to provide an explanation of noninvariance of measures of attitudes toward gender roles. In MLSEM, differences in measurement parameters (e.g., intercepts) across countries are assumed to be present as residual variances at the between-level, which can be regressed on country-level predictors. First, we were able to replicate the results of a previous study on measurement invariance that used the MGCFA approach (AUTHOR, DATE). With the current MLSEM approach, we showed that the basic measurement model for gender role attitudes was valid across countries. Furthermore, it was obvious that significant residual variance existed between countries. We interpreted this as an indication for the lack of scalar measurement invariance, which is a prerequisite for comparing factor means across countries. In particular, measurement noninvariance was indicated for items addressing the aspect of a preschool child suffering from a working mother and the domain of home life. We included two country-level variables to explain noninvariance of these items. The Gender Inequality Index from the UNPD, which refers to a structural measure assessing gender inequalities in the domains of reproductive health, political empowerment, and economic participation, was positively related to gender role attitudes at the country level. Thus, countries with higher structural gender inequalities show a higher degree of approval for traditional gender roles. However, the structural information did not contribute to explaining noninvariance of the measures. The cultural value embeddedness, which refers to a measure of the importance to maintain the status quo and the traditional system of role separation in a society, was positively related to gender role attitudes at the country level. This indicates that countries emphasizing the

importance of maintaining the societal status quo show a higher degree of approval for traditional gender roles. Embeddedness also explained *why* the measurements were partly nonequivalent across countries. The differing societal views on and importance given to the maintenance of the traditional order and status quo provide different cultural frameworks for the interpretation of the item wording. Furthermore, the process of ambiguity-based framing (Braun, 2009) can explain how societal value orientations function to fill the informational gaps left by vaguely formulated items (as in the case of item A preschool child is likely to suffer if his or her mother works). Ambiguities that make it difficult to properly understand and interpret questionnaire item increase measurement biases across countries because of the different cultural information that is used to fill the gaps. Thus, even if items are semantically equivalent and translated very well, culturally induced measurement invariance may be an issue.

Thus, it can be expected that the comparability of measures assessing gender role attitudes is likely to be limited when researchers compare countries from very different cultural backgrounds. However, what does it mean that noninvariance is accounted for only by the cultural value embeddedness? The ability to explain between-level residual variance refers to accounting for item specific differences that are not due to differences on the common factor. This means that apart from differences on the latent dimension (gender role attitudes) respondents from different countries will respond differently due to country-differences in the degree to which collective goals, the status-quo, and traditional gender separation are favored over personal goals, change, and autonomy. Thus, the tendency to respond more in favor of a traditionalist gender role view despite any differences on the latent dimension represents a cultural bias (Braun, 1998, 2008, 2009), meaning that particular items are culturally charged and a little extra is added to the response.

Such a bias could not be found for the Gender Inequality Index. On the one hand, this may be explained by a predominance of culture in shaping attitudes and providing information for addressing interpretative ambiguities. On the other hand, structural characteristics referring to opportunity structures and gender role segregation may be interpreted as manifestations of cultural differences, which govern differences between countries in their decisions to implement, for example, more traditional or more liberal gender policies. This is view is also supported if we use only GII as a between-level predictor. In this case, the GII *can significantly explain* noninvariance in some items.[5] While in this study we focused on the GII as a structural indicator particularly related to the substantive issue of gender equality, future research may consider also other elements of the opportunity structure connected to individual gender role attitudes, such as the female economic participation and work-family balance policies (AndrØ et al., 2013; Kangas and Rostgaard, 2007; Lomazzi et al., 2019; Sjöberg, 2004), which can potentially contribute to the explanation of their noninvariance.

In order to facilitate comparative analysis in the area, future research is required to develop measures that are less sensitive to cultural differences in the way respondents address certain items. For example, to reduce the risk of different interpretations across societies, items could be improved by providing enough specifications and adopting less traditionalist wordings. According to the construal model of attitudes (Schwarz, 2007; Tourangeau and Rasinski, 1988; Wilson and Hodges, 1992), respondents make use of the most recent available information to interpret the question and express their judgment. As Braun (2009) argues, items such as A preschool child is likely to suffer if his or her mother works (*childsuffers*) leave respondents

[5] The same line of reasoning may also explain why noninvariance appeared only in some items. Although the *familysuffers* item also addresses the working mother aspect, it did not display significant noninvariance. Other items may be more sensitive toward differences in embeddedness. For example, the *childsuffers* item may tap more precisely into the domain of a homemaker s duties (e.g., child rearing) than the *familysuffers* item, thus eliciting the cultural bias.

with many open issues, concerning, for example, the age of the child, the mother s working time, the presence and the role of the father, and the possible availability of childcare services. Without this information, respondents tend to fill these gaps automatically, for example making use of information coming from their cultural context, which can differ by female employment rate, childcare provisions, and so on. Also items with strong traditionalist wordings, as in our case the item A man's job is to earn money; a woman's job is to look after the home and family (*womenjob*), as well strongly progressive options can be problematic (Braun, 2008, 2009; Walter, 2017) because both cases may produce problems of noninvariance due to cultural embeddedness and the related societal views on family patterns and the role of fathers and mothers.

Regarding the statistical methodology, we also encourage researchers interested in focusing more strictly on the item-level in examining the reasons for non-equivalence to use methods based in item response theory (e.g., Fox and Verhagen, 2018; Janssen, 2018; Quandt, 2018). These methods take into account that items are truly categorical and may provide a deeper understanding why items function differentially across countries.

Acknowledgments

The authors would like to thank Lisa Trierweiler for the English proof of the manuscript.

References

Albrecht, J.W., Edin, P.-A., Vroman, S. B., 2000. A cross-country comparison of attitudes towards mothers working and their actual labor market experience. LABOUR, 14(4), 591–607. https://doi.org/10.1111/1467-9914.00147

AlemÆn, J., Woods, D. 2016. Value Orientations From the World Values Survey How Comparable Are They Cross-Nationally? Comparative Political Studies, 49(8), 1039–1067. https://doi.org/10.1177/0010414015600458

Alexander, A. C., Welzel, C. 2015. Eroding patriarchy: the co-evolution of women s rights and emancipative values. International Review of Sociology, 25(1), 144–165. https://doi.org/10.1080/03906701.2014.976949

Alwin, D.F., 2005. Attitudes, beliefs, and childbearing. In A. Booth & A.C. Crouter (Eds.), The New Population Problem: Why Families in Developed Countries are Shrinking and What it Means. Mahwah, NJ: Lawrence Erlbaum Associates, 115–126.

AndrØ, S., Gesthuizen, M., Scheepers, P., 2013. Support for Traditional Female Roles across 32 Countries: Female Labour Market Participation, Policy Models and Gender Differences. Comparative Sociology, 12(4), 447–476. https://doi.org/10.1163/15691330-12341270

Arpino, B., Esping-Andersen, G., Pessin, L., 2015. How do changes in gender role attitudes towards female employment influence fertility? A macro-level analysis. European Sociological Review, 31(3), 370–382. https://doi.org/10.1093/esr/jcv002

Asparouhov, T., MuthØn, B.O., 2014. Multiple-groupfactor analysis alignment. Structural Equation Modeling 2, 495–508. http://dx.doi.org/10.1080/10705511.2014.919210

Baxter, J., Kane, E.W., 1995. Dependence and independence. A cross-national analysis of gender inequality and gender attitudes. Gender & Society, 9(2), 193–215. https://doi.org/10.1177/089124395009002004

Bergh, J., 2006. Gender attitudes and modernization processes. International Journal of Public Opinion Research, 19(1), 5–23. https://doi.org/10.1093/ijpor/edl004

Bianchi, S. M., Sayer, L. C., Milkie, M. A., & Robinson, J. P., 2012. Housework: Who did, does or will do it, and how much does it matter? Social Forces, 91(1), 55–63. https://doi.org/10.1093/sf/sos120

Bollen, K.A., 1989. Structural Equations with Latent Variables. New York, NY: Wiley.

Bolzendahl, C. I., Myers, D. J., 2004. Feminist attitudes and support for gender equality: Opinion change in women and men, 1974-1998. Social Forces, 83(2), 759–789. https://doi.org/10.1353/sof.2005.0005

Braun, M., 1998. Gender roles. In Van Deth, JW (Ed.), Comparative Politics: The Problem of Invariance. London, England: Routledge, 111–134.

Braun, M., 2008. Using egalitarian items to measure men s and women s family roles. Sex Roles, 59(9–10), 644–656. https://doi.org/10.1007/s11199-008-9468-5

Braun, M., 2009. The role of cultural contexts in item interpretation: the example of gender roles. In M. Haller, R. Jowell, T.W. Smith (Eds.), The International Social Survey Programme, 1984-2009: Charting the globe. London/New York: Routledge, 395–408.

Brewster, K.L., Padavic, I., 2000. Change in gender-ideology, 1977–1996: The contributions of intracohort change and population turnover. Journal of Marriage and Family, 62(2), 477–487. https://doi.org/10.1111/j.1741-3737.2000.00477.x

Brieger, S. A., Francoeur, C., Welzel, C., Ben-Amar, W. 2019. Empowering Women: The Role of Emancipative Forces in Board Gender Diversity. Journal of Business Ethics, 155(2), 495–511. https://doi.org/10.1007/s10551-017-3489-3

Brooks, C., Bolzendahl, C., 2004. The transformation of US gender role attitudes: cohort replacement, social-structural change, and ideological learning. Social Science Research, 33(1), 106–133. https://doi.org/10.1016/S0049-089X(03)00041-3

Brunkert, L., R. Inglehart, S. Kruse and Welzel C.. Forthcoming. Measurement Equivalence? A Tale of False Obsessions and a Cure. World Values Research 11 (3): 54-84

Byrne, B.M., van de Vijver, F.F., 2010. Testing for measurement and structural invariance in large-scale cross-cultural studies: Addressing the issue of noninvariance. International Journal of Testing, 10, 107–132. https://doi.org/10.1080/15305051003637306

Byrne, B.M., Shavelson, R.S., MuthØn, B. 1989. Testing for the invariance of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin 105, 456–466.

Chen, F.F. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling 14, 464–504. https://doi.org/10.1080/10705510701301834

Chen, G., Bliese, P. D., Mathieu, J. E., 2005. Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. Organizational Research Methods, 8, 375–409. https://doi.org/10.1177/1094428105280056

Cheung, M.W.-L., Au, K., 2005. Applications of multilevel structural equation modeling to cross-cultural research. Structural Equation Modeling 12, 598–619. https://doi.org/10.1207/s15328007sem1204_5

Cieciuch, J., Davidov, E., Schmidt, P., 2018. Alignment optimization: Estimation of the most trustworthy means in cross-cultural studies even in the presence of noninvariance. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. New York: Routledge, 571–593.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., Schwartz, S.H., 2014. Comparing results of an exact versus an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. Frontiers in Psychology 5, 982. https://doi.org/10.3389/fpsyg.2014.00982

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P., 2017. Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research. https://doi.org/10.1177/0049124117701478

Constantin, A., Voicu, M., 2015. Attitudes towards gender roles in cross-cultural surveys: Content validity and cross-cultural measurement invariance. Social Indicators Research, 123(3), 733–751. https://doi.org/10.1007/s11205-014-0758-8

Crompton, R., Brockmann, M., Lyonette, C., 2005. Attitudes, women s employment and the domestic division of labour a cross-national analysis in two waves. Work, Employment & Society, 19(2), 213–233. https://doi.org/10.1177/0950017005053168

Cunningham, M., 2008. Influences of gender ideology and housework allocation on women s employment over the life course. Social Science Research, 37(1), 254–267. https://doi.org/10.1016/j.ssresearch.2007.01.003

Cunningham, M., Beutel, A. M., Barber, J. S., Thornton, A., 2005. Reciprocal relationships between attitudes about gender and social contexts during young adulthood. Social Science Research, 34(4), 862–892. https://doi.org/10.1016/j.ssresearch.2005.03.001

Davidov, E., Meuleman, B., Billiet, J., Schmidt, P., 2008. Values and support for immigration: A cross-country comparison. European Sociological Review, 24(5), 583–599. https://doi.org/10.1093/esr/jcn020
Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., Billiet, J., 2014. Measurement invariance in cross-national research. Annual Review of Sociology, 40, 50–75. https://doi.org/10.1146/annurev-soc-071913-043137
Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., Meuleman, B., 2012. Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. Journal of Cross-Cultural Psychology, 43, 558–575. https://doi.org/10.1177/0022022112438397
Davidov, E., Dülmer, H., Cieciuch, J., Kuntz, A., S eddig, D., Schmidt, P., 2018a. Explaining measurement noninvariance using multilevel structural equation modeling: The case of attitudes toward citizenship rights. Sociological Methods & Research, 47(4), 729-760. https://doi.org/10.1177/0049124116672678
Davidov, E., Schmidt, P., Billiet, J., Meuleman, B., 2018b. Cross-Cultural Analysis. 2nd Edition. New York, NY: Routledge.
Davis, S.N., Greenstein, T.N., 2009. Gender ideology: Components, predictors, and consequences. Annual Review of Sociology, 35(1), 87–105. https://doi.org/10.1146/annurev-soc-070308-115920
De Beuckelaer, A., Swinnen, G., 2018. Biased latent variable mean comparisons due to measurement noninvariance: A simulation study. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. New York: Routledge, 127–156.
Fox, J.-P., Verhagen, J., 2018. Random item effects modeling for cross-national survey data. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. 2nd Edition. New York: Routledge, 529–550.
Fuwa, M., 2004. Macro-level gender inequality and the division of household labor in 22 countries. American Sociological Review, 69(6), 751–767. https://doi.org/10.1177/000312240406900601
Gaye, A., Klugman, J., Kovacevic, M., Twigg, S., Zambrano, E., 2010. Measuring Key Disparities in Human Development: The Gender Inequality Index (Human Development Research Papers (2009 to present) No. HDRP-2010-46). Human Development Report Office (HDRO), United Nations Development Programme (UNDP). Retrieved from https://ideas.repec.org/p/hdr/papers/hdrp-2010-46.html
Geist, C., Cohen, P.N., 2011. Headed toward equality? Housework change in comparative perspective. Journal of Marriage and Family, 73(4), 832–844. https://doi.org/10.1111/j.1741-3737.2011.00850.x
Grunow, D., Schulz, F., Blossfeld, H.-P., 2012. What determines change in the division of housework over the course of marriage? International Sociology, 27(3), 289–307. https://doi.org/10.1177/0268580911423056
Guenole, N., 2016. The importance of isomorphism for conclusions about homology: A Bayesian multilevel structural equation modeling approach with ordinal indicators. Frontiers in Psychology 7, 289. https://doi.org/10.3389/fpsyg.2016.00289
Horn, J.L., McArdle, J.J., 1992. A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research 18, 117–144. https://doi.org/10.1080/03610739208253916
Hox, J., 2010. Multilevel Analysis. Techniques and Applications. 2nd Edition. New York, NY: Routledge.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6: 1–55. https://doi.org/10.1080/10705519909540118
Inglehart, R., 1997. Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies (Vol. 19). Cambridge University Press.
Inglehart, R., Norris, P., 2003. Rising Tide: Gender Equality and Cultural Change Around the World. Cambridge, UK; New York: Cambridge University Press.
ISSP Research Group, 2016. International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012. GESIS Data Archive, Cologne. ZA5900 Data file Version 4.0.0.
Jak, S., Oort, F.J., Dolan, C.V., 2013. A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. Structural Equation Modeling 20, 265–282. https://doi.org/10.1080/10705511.2013.769392
Jak, S., Oort, F.J., Dolan, C.V., 2014a. Measurement bias in multilevel data. Structural Equation Modeling 21, 31–39. https://doi.org/10.1080/10705511.2014.856694
Jak, S., Oort, F.J., Dolan, C.V., 2014b. Using two-level factor analysis to test for cluster bias in ordinal data. Multivariate Behavioral Research 49, 544–553. https://doi.org/10.1080/00273171.2014.947353
Jak, S., 2014. Testing strong factorial invariance using three-level structural equation modeling Frontiers in Psychology 5, 745.
Jak, S., 2017. Testing and explaining differences in common and residual factors across many countries. Journal of Cross-Cultural Psychology 48, 75–92. https://doi.org/10.1177/0022022116674599
Janssen, R., 2018. Using a differential item functioning approach to investigate measurement invariance. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. 2nd Edition. New York: Routledge, 483–500.
Jöreskog, K.G. 1971. Simultaneous factor analysis i n several populations. Psychometrika 36, 409–26. https://doi.org/10.1007/BF02291366
Jowell, R. 1998. How comparative is comparative research? American Behavioral Scientist, 42(2), 168–177. https://doi.org/10.1177/0002764298042002004
Kalmijn, M., 2003. Country differences in sex-role attitudes: cultural and economic explanations. In Arts, W. Halman, L. (Eds.), The Cultural Diversity of European Unity: Findings, Explanations and Reflections from the European Values Study. Leiden: Brill, 311–336.
Kangas, O., Rostgaard, T., 2007. Preferences or institutions? Work family life opportunities in seven European countries. Journal of European Social Policy, 17(3), 240–256. https://doi.org/10.1177/0958928707078367
Kim, E. S., Yoon, M., Wen, Y., Luo, W., Kwok, O., 2015. Within-level group factorial invariance with multilevel data: Multilevel factor mixture and multilevel MIMIC models. Structural Equation Modeling, 22, 603–16. https://doi.org/10.1080/10705511.2014.938217
Klasen, S., Schüler, D., 2011. Reforming the gender -related development index and the gender empowerment measure: Implementing some specific proposals. Feminist Economics, 17(1), 1–30. https://doi.org/10.1080/13545701.2010.541860
Kraaykamp, G., 2012. Employment status and family role attitudes: A trend analysis for the Netherlands. International Sociology, 27(3), 308–329. https://doi.org/10.1177/0268580911423046

Kroska, A., Elman, C., 2009. Change in attitudes about employed mothers: Exposure, interests, and gender ideology discrepancies. Social Science Research, 38(2), 366–382. https://doi.org/10.1016/j.ssresearch.2008.12.004
Lomazzi, V., 2017. Gender role attitudes in Italy: 1988–2008. A path-dependency story of traditionalism. European Societies, 19(4), 370–395. https://doi.org/10.1080/14616696.2017.1318330
Lomazzi, V., 2018. Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. Methods, Data, Analyses (Mda), 12(1), 77–103. https://doi.org/10.12758/mda.2017.0x
Lomazzi, V., Israel, S., Crespi, I. 2019.. Gender equality in Europe and the effect of work-family balance policies on gender role attitudes. Social Sciences, 8, 5. https://doi.org/10.3390/socsci8010005
Lussier, D.N., Fish, M.S., 2016. Men, Muslims, and attitudes toward gender inequality. Politics and Religion, 9(1), 29–60. https://doi.org/10.1017/S1755048315000826
Malik, K., 2013. Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World. Ney York, NY: United Nations Development Programme.
Mason, K.O., Lu, Y.-H., 1988. ATTITUDES TOWARD WOMEN S FAMILIAL ROLES: Changes in the United States, 1977-1985. Gender & Society, 2(1), 39–57. https://doi.org/10.1177/089124388002001004
Marsh, H.W., Guo, J., Parker, P.D., Nagengast, B., Asparouhov, T., MuthØn B.O., Dicke, T., 2017. What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. Psychological Methods. Advance online publication. http://dx.doi.org/10.1037/met0000113
Marsh, H.W., Hau, K.-T., Wen, Z., 2004. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler s (1999) findings. Structural Equation Modeling 11, 320–341. https://doi.org/10.1207/s15328007sem1103_2
Misra, R., 1995. CHANGE IN ATTITUDES TOWARD WORKING WOMEN: A COHORT ANALYSIS. International Journal of Sociology and Social Policy, 15(6), 1–20. https://doi.org/10.1108/eb013213
Moore, L. M., Vanneman, R., 2003. Context matters: Effects of the proportion of fundamentalists on gender attitudes. Social Forces, 82(1), 115–139. https://doi.org/10.1353/sof.2003.0099
Motiejunaite, A., Kravchenko, Z., 2008. Family policy, employment and gender-role attitudes: a comparative analysis of Russia and Sweden. Journal of European Social Policy, 18(1), 38–49. https://doi.org/10.1177/0958928707084453
Munck, I., Barber, C., Torney-Purta, J., 2017. Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. Sociological Methods & Research. https://doi.org/10.1177/0049124117729691
MuthØn, B.O., 1985. A method for studying the homogeneity of test items with respect to other relevant variables. Journal of Educational Statistics 10, 121–32. https://doi.org/10.3102/10769986010002121
MuthØn, B.O., 1991. Multilevel factor analysis of class and student achievement components. Journal of Educational Measurement 28, 338–54. https://doi.org/10.1111/j.1745-3984.1991.tb00363.x

MuthØn, B.O., 1994. Multilevel covariance structureanalysis. Sociological Methods & Research 22, 376–98. https://doi.org/10.1177/0049124194022003006
MuthØn, B.O., Asparouhov, T., 2012. Bayesian structural equation modeling: A more flexible representation of substantive theory. Psychological Methods, 17(3), 313–335.
MuthØn, B.O., Asparouhov, T., 2013. BSEM measurement invariance analysis. Mplus Web Notes 17. Retrieved from http://www.statmodel.com/
MuthØn, B.O., Asparouhov, T., 2014. IRT studies ofmany groups: the alignment method. Frontiers in Psychology 5, 978. Advance online publication. https://doi.org/10.3389/fpsyg.2014.00978
MuthØn, L.K., MuthØn, B.O., 1998-2017. Mplus user sguide. 8th Edition. Los Angeles, CA: MuthØn & MuthØn.
Permanyer, I., 2013. A critical assessment of the UNDP s Gender Inequality Index. Feminist Economics, 19(2), 1–32. https://doi.org/10.1080/13545701.2013.769687
Pfau-Effinger, B., 2004. Socio-historical paths of the male breadwinner model – an explanation of cross-national differences. The British Journal of Sociology, 55(3), 377–399. https://doi.org/10.1111/j.1468-4446.2004.00025.x
Quandt, M., 2018. Using the mixed Rasch model in the comparative analysis if attitudes. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. 2nd Edition. New York: Routledge, 501–527.
Schultz Lee, K.S., Tufiş, P.A., Alwin, D.F., 2010. Separate spheres or increasing equality? Changing gender beliefs in postwar Japan. Journal of Marriage and Family, 72(1), 184–201. https://doi.org/10.1111/j.1741-3737.2009.00691.x
Schwartz, S., 1994. Beyond Individualism/Collectivism. New cultural dimensions of values. In Kim, U., Triandis, H. C., Kagitçibasi, Ç, Choi, S.-C., Yoon, G. (Eds.), Individualism and Collectivism. Theory, Method, and Applications. London: Thousand Oaks, 85–119.
Schwartz, S.H., 2006. A Theory of Cultural Value Orientations: Explication and Applications. Comparative Sociology, 5(2), 137–182. https://doi.org/10.1163/156913306778667357
Schwartz, S.H. 2008. National cultural value orientation scores. DOI:10.13140/RG.2.1.3313.3040
Schwarz, N., 2007. Attitude construction: Evaluation in context. Social Cognition, 25(5), 638–656. https://doi.org/10.1521/soco.2007.25.5.638
Scott, J., Braun, M., 2009. Changing public views of gender roles in seven nations: 1988-2002. In Haller, M., Jowell, R., Smith, T. W. (Eds.), The International Social Survey Programme, 1984-2009: charting the globe. New York: Routledge. Retrieved from http://www.repository.cam.ac.uk/handle/1810/242230
Scott, J., Alwin, D.F., Braun, M., 1996. Generational changes in gender-role attitudes: Britain in a cross-national perspective. Sociology, 30(3), 471–492. https://doi.org/10.1177/0038038596030003004
Seddig, D., Leitgöb, H., 2018a. Approximate measure ment invariance and longitudinal confirmatory factor analysis: concept and application with panel data. Survey Research Methods 12, 29–41. http://doi.org/10.18148/srm/2018.v12i1.7210
Seddig, D., Leitgöb, H., 2018b. Exact and Bayesian approximate measurement invariance. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-Cultural Analysis: Methods and Applications. 2nd Edition. New York: Routledge, 553–579.
Sjöberg, O., 2004. The role of family policy instit utions in explaining gender-role attitudes: A comparative multilevel analysis of thirteen industrialized countries. Journal of European Social Policy, 14(2), 107–123. https://doi.org/10.1177/0958928704042003

Sokolov, B. 2018. The Index of Emancipative Values: Measurement Model Misspecifications. American Political Science Review, 112(2), 395–408. https://doi.org/10.1017/S0003055417000624
Steenkamp, J.-B., Baumgartner, H., 1998. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research 25, 78–90. https://doi.org/10.1086/209528
Stegmueller, D. 2011. Apples and Oranges? The Problem of Equivalence in Comparative Research. Political Analysis, 19(4), 471–487. https://doi.org/10.1093/pan/mpr028
Steinmetz, H., 2018. Estimation and comparison of latent means across cultures. In Davidov, E., Schmidt, P., Billiet, J., Meuleman, B. (Eds.), Cross-cultural Analysis: Methods and Applications. 2nd Edition. New York: Routledge, 95–126.
Stickney, L.T., Konrad, A.M., 2007. Gender-role attitudes and earnings: A multinational study of married women and men. Sex Roles, 57(11–12), 801–811. https://doi.org/10.1007/s11199-007-9311-4
Tay, L., Woo, S.E., Vermunt, J.K., 2014. A conceptual framework of cross-level isomorphism: Psychometric validation of multilevel constructs. Organizational Research Methods 17, 77-106. https://doi.org/10.1177/1094428113517008
Tourangeau, R., Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. Psychological Bulletin, 103(3), 299.
Van de Schoot, R., Kluytmans, A., Tummers, A., Lugtig, P., Hox, J.J., MuthØn, B.O., 2013. Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. Frontiers in Psychology 4, 770. https://doi.org/10.3389/fpsyg.2013.00770
van Vlimmeren, E., Moors, G.B.D., Gelissen, J.P.T.M., 2016. Clusters of cultures: diversity in meaning of family value and gender role items across Europe. Quality & Quantity, 1–24. https://doi.org/10.1007/s11135-016-0422-2
Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods 3, 4–70. https://doi.org/10.1177/109442810031002
Walter, J.G., 2017. The adequacy of measures of gender roles attitudes: a review of current measures in omnibus surveys. Quality & Quantity, 52(2), 284–248. https://doi.org/10.1007/s11135-017-0491-x
Welzel, C. 2013. Freedom rising: Human empowerment and the quest for emancipation. New York: Cambridge University Press.
Welzel, C., Inglehart, R., 2016. Misconceptions of measurement invariance: Time for a paradigm shift. Comparative Political Studies 49, 1068–1094. https://doi.org/10.1177/0010414016628275
West, S.G., Taylor, A.B., Wu, W., 2012. Model fit and model selection in structural equation modeling. In Hoyle, R.H. (Ed.), Handbook of structural equation modeling. New York, NY: Guilford Press, 209–231.
Wilson, T., Hodges, S., 1992. Attitudes as temporary constructions. In Martin, L.L., Tesser, A. (Eds.), The Construction of Social Judgments. Hillsdale: Erlbaum, 37–61.

Table 1. Items in the ISSP measuring gender role attitudes (*N* = 61,754)

Item (variable name)	Wording	Mean	SD	% missing
warmrelation (v5)	A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.	2.26	1.14	6.31
childsuffers (v6)	A preschool child is likely to suffer if his or her mother works.	3.03	1.22	6.81
familysuffers (v7)	All in all, family life suffers when the woman has full-time job.	2.94	1.24	6.68
home&kids (v8)	A job is all right, but what most women really want is a home and kids.	3.12	1.21	8.70
housework as fulfilling (v9)	Being a housewife is just as fulfilling as working for pay.	3.19	1.17	9.95
bothincome (v10)	Both the man and woman should contribute to the household income.	2.00	0.93	5.79
womensjob (v11)	A man's job is to earn money; a woman's job is to look after the home and family.	2.73	1.29	5.95

Table 2. Between-level variables: Gender Inequality Index (GII) and Embeddedness (EMB) scores

Country	GII (UNPD)	EMB
Argentina	0.380	3.52
Australia	0.115	3.59
Belgium	0.098	3.25
Bulgaria	0.219	3.87
Canada	0.119	3.3
Chile	0.360	3.64
China	0.213	3.74
Croatia	0.179	4
Czech Republic	0.122	3.59
Denmark	0.057	3.19
Finland	0.075	3.37
France	0.083	3.2
Germany	0.075	3.09
Great Britain	0.205	3.34
Hungary	0.256	3.6
Ireland	0.121	3.41
Israel	0.144	3.85
Japan	0.131	3.49
Korea	0.153	3.68
Latvia	0.216	3.83
Netherlands	0.045	3.19
Philippines	0.418	4.03
Poland	0.140	3.86
Portugal	0.114	3.43
Russia	0.312	3.81
Slovakia	0.171	3.82
Slovenia	0.080	3.71
South Africa	0.462	4.03
Sweden	0.055	3.12
Switzerland	0.057	3.19
Taiwan	0.053	3.82
Turkey	0.366	3.77
United States	0.256	3.67
Venezuela	0.466	3.74

Table 3. Parameter estimates for the final two-level structural equation model (N = 59,119).

	Within-level				Between-level			
Par.	Est.	SE	p	Std.	Est.	SE	p	Std.
Factor loadings								
$\lambda_{vchildsuffers\_GRA}$	0.806	0.030	0.000	0.710	0.806	0.030	0.000	0.799
$\lambda_{familysuffers\_GRA}$	0.927	0.029	0.000	0.799	0.927	0.029	0.000	0.954
$\lambda_{vhome\&kids\_GRA}$	0.501	0.029	0.000	0.448	0.501	0.029	0.000	0.484
$\lambda_{vwomensjob\_GRA}$	0.489	0.035	0.000	0.428	0.489	0.035	0.000	0.383
$\lambda_{home\&kids\_RCF}$	1.000			0.469	1.000			0.388
$\lambda_{womensjob\_RCF}$	1.000			0.459	1.000			0.314
Factor variances								
$\varphi_{\_GRA}$	1.000			1.000	0.214	0.048	0.000	1.000
$\varphi_{\_RCF}$	0.275	0.019	0.000	1.000	0.034	0.014	0.014	1.000
Residual variances								
$\theta_{childsuffers}$	0.640	0.026	0.000	0.496	0.018	0.008	0.032	0.082
$\theta_{amilysuffers}$	0.486	0.032	0.000	0.361	0.018	0.010	0.071	0.090
$\theta_{home\&kids}$	0.724	0.029	0.000	0.580	0.024	0.011	0.032	0.106
$\theta_{womensjob}$	0.791	0.044	0.000	0.606	0.052	0.014	0.000	0.149
Covariances								
$cov_{(GRA\_GII)}$					0.039	0.010	0.000	0.679
$cov_{(GRA\_EMB)}$					0.069	0.019	0.000	0.536
$cov_{(GII\_EMB)}$					0.021	0.005	0.000	0.618
Regression coefficients								
$b_{(childsuffers\_GII)}$					0.244	0.333	0.464	0.065
$b_{(home\&kids\_GII)}$					-0.009	0.534	0.986	-0.002
$b_{(womensjob\_GII)}$					0.403	0.589	0.494	0.085
$b_{(childsuffers\_EMB)}$					0.308	0.144	0.032	0.184
$b_{(home\&kids\_EMB)}$					0.865	0.238	0.000	0.502
$b_{(womensjob\_EMB)}$					1.127	0.276	0.000	0.530

Note: Par. = Parameter, Est. = estimate, SE = standard error, p = p-value, Std. = standardized parameter estimate, GRA = gender role attitudes, RCF = residual correlation factor, GII = gender inequality index, EMB = embeddedness.

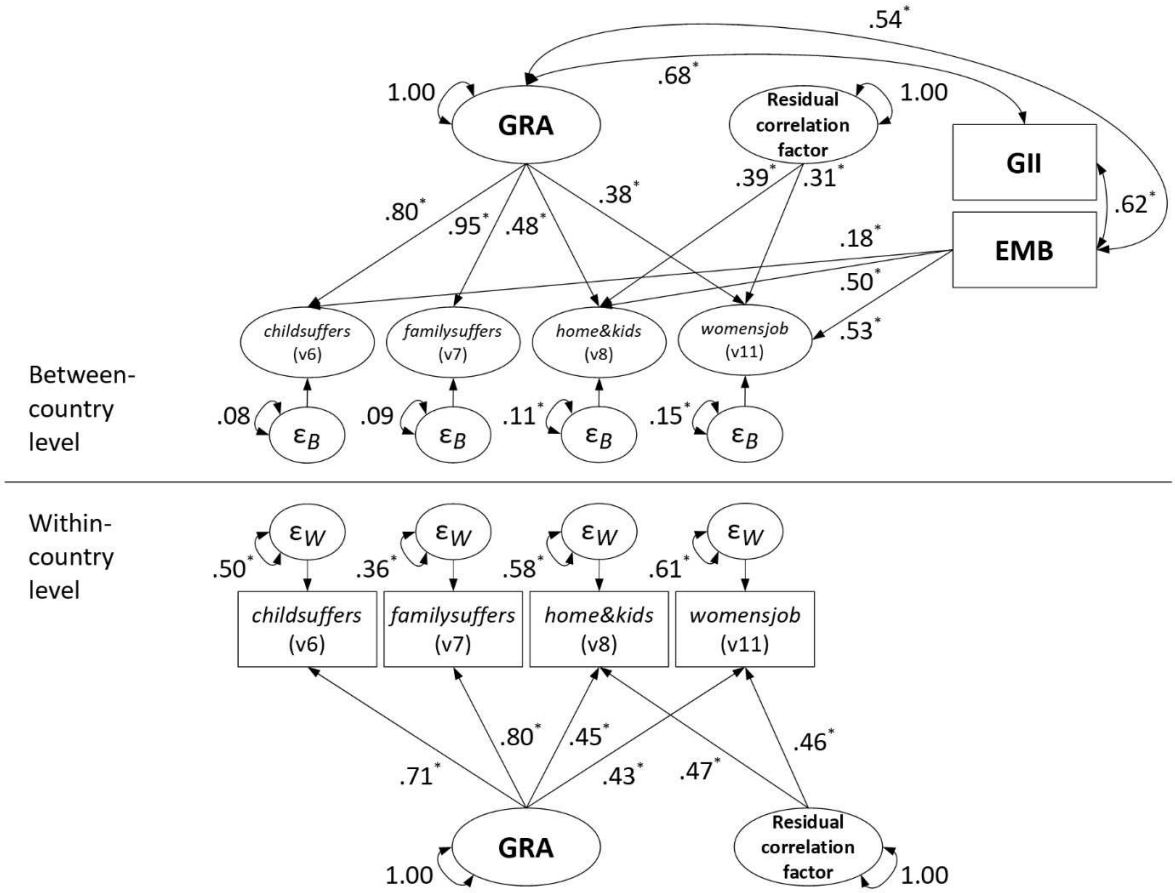

Figure 1. Two-level structural equation model for gender role attitudes (GRA) and country-level variables gender inequality index (GII) and embeddedness (EMB). The figure shows the completely standardized solution. * $p < .05$. Nonsignificant effects of GII on v6, v8, and v11 are not displayed.