Detecting and Explaining Missing Comparability in Cross-National Studies: The Case of Citizen Evaluation of Patriotism

Katharina Meitinger[1] · Peter Schmidt[2,3] · Michael Braun
[1]Utrecht University
[2]University of Giessen
[3]University of Mainz

Measurement invariance tests are an important precondition to analyze cross-national data. However, the traditional approach of multigroup confirmatory factor analysis (MGCFA) has been criticized as too strict and more liberal approaches, such as alignment, have been proposed. However, both approaches can only detect but cannot explain why there are comparability issues. Mixed methods approaches combining quantitative and qualitative insights from web probing provide a powerful tool to detect and explain a lack of comparability of measures. For this study, we selected the 2013 International Social Survey Program item battery on "Citizen Evaluation Of Patriotism" and assessed the comparability for Germany (N=1717), Great Britain (N=904), the U.S. (N=1274), Mexico (N=1062), and Spain (N=1225) and combined it with web probing results from an online survey conducted in 2014 in the five countries (N=2685). Strict measurement invariance tests using MGCFA failed to show scalar measurement invariance but with an approximate approach of alignment estimation unbiased equal factor loadings and latent means could be estimated for all countries. In line with MGCFA results, qualitative web probing detected issues that question the comparability of results.

Keywords: National identity; citizen evaluation of patriotism; measurement invariance; approximate measurement invariance; alignment; web probing; construct bias; mixed methods; ISSP

1 Assessing Cross-national Data

Given the tremendous increase in cross-national data production in recent decades, studies comparing many countries are increasingly popular (Smith, 2010). However, cross-national differences can be methodological artifacts due to various reasons, also called biases. Biases are "nuisance factors that jeopardize the validity of instruments applied in different cultures" (He van de Vijver, 2012, p. 3). *Construct bias* means that the measured construct differs across cultures (Van de Vijver Poortinga, 1997), whereas distorting effects through specific methods and the context of the measurement can create a *method bias*, e.g., due to differences in sampling procedures. Additionally, *item bias* can appear due to poor item translation, ambiguous source

items or inapplicability of item contents (He van de Vijver, 2012).

Therefore, a precondition of drawing substantive conclusions is to assess measurement invariance. Otherwise, researchers risk confusing ambiguous and erroneous data as "real" substantive differences across countries (Steenkamp Baumgartner, 1998).

1.1 From Strict to Approximate Measurement Invariance

Exact measurement invariance (MI) tests that use multigroup confirmatory factor analysis (MGCFA) (Jöreskog, 1971) or IRT models (Muthén Asparouhov, 2014; Fox Verhagen, 2018) are widespread approaches to assess the cross-national comparability of survey data. They provide insights whether the constructs (configural invariance), the coefficients (metric invariance) and the latent means of a construct (scalar invariance) are cross-nationally comparable

Corresponding author: Katharina Meitinger, Utrecht University, Padualaan 14, 3584 CH Utrecht, Netherlands (Email: k.m.meitinger@uu.nl)

(Meredith, 1993; Millsap, 2011; Steenkamp Baumgartner, 1998; Vandenberg Lance, 2000).

However, scalar MI tests often fail and approaches such as multilevel analyses (Hox et al., 2017) or a country ranking of the constructs' mean values cannot be performed with confidence anymore. Since especially scalar MI is rarely established, the exact approach has been criticized as too strict.

More liberal approaches, such as Bayesian structural equation modeling (BSEM) (Muthén Asparouhov, 2012) or alignment (Asparouhov Muthén, 2014; Muthén Asparouhov, 2018) have been proposed and can be classified as approximate approaches (Byrne van de Vijver, 2017). Several studies already compared the results of MGCFA with alignment. Most studies found an acceptable degree of noninvariance for a large number of countries when previous exact scalar MI tests failed (e.g., Lomazzi, 2018; Munck et al., 2018). Furthermore, in a recent simulation study (Pokropek et al., 2019) it was shown that given a few large invariant parameters, alignment was best in recovering latent means compared with exact partial measurement invariance tests and Bayesian estimation (BSEM). However, neither the exact nor the liberal approaches can provide reasons for missing comparability if measurement invariance tests fail.

1.2 Web Probing as a Tool to Reveal the Reasons for Missing Comparability

This study contributes to recent developments in the field of MI testing by supplementing the results of multigroup confirmatory factor analysis and alignment with qualitative insights from web probing (WP). We argue that the qualitative input of WP provides valuable insights into the comparability assessment. We will first introduce the methods of MGCFA, alignment, and WP as well as the substantive example of citizen evaluation of patriotism (Ariely, 2018). This item battery has been newly added in the 2013 ISSP Module on National Identity and its cross-national comparability has not been assessed, yet. This will be followed by a presentation of the results of the different approaches and a discussion of the results' implications.

2 Quantitative Measurement Invariance

2.1 Multigroup Confirmatory Factor Analysis

MI tests using MGCFA (Jöreskog, 1971) are the predominant approach to assess the cross-national comparability of constructs (Davidov et al., 2014). If configural invari-

ance is established, the latent concept can be meaningfully discussed with respect to all countries. If metric invariance is supported, it is possible to explore cross-national structural relationships, such as regression coefficients. Finally, achieving scalar invariance is a precondition for comparing mean values across countries (Meredith, 1993; Millsap, 2011; Steenkamp Baumgartner, 1998; Vandenberg Lance, 2000). The different levels of MI are usually assessed with chi-square difference tests. However, Monte-Carlo simulations (Chen, 2007) showed that chi-square difference tests too often reject the invariance hypotheses given the true model. Therefore, the model fit of the baseline model (configural) is assessed with goodness-of-fit (GOF) indices, for example the root mean square error of approximation (RMSEA) (Browne Cudeck, 1992) and the comparative fit index (CFI) (Bentler, 1990). RMSEA values below 0.06 and CFI values of at least 0.95 indicate a good model fit and RMSEA values below 0.08 an acceptable model fit. The metric and scalar invariance tests are assessed with the difference in the RMSEA and CFI values of the test levels, ΔRMSEA and ΔCFI, with changes exceeding 0.015 for RMSEA and 0.01 for CFI indicating problematic values (Chen, 2007). In addition, researchers can use modification indices (MIs) (Steenkamp Baumgartner, 1998), the power of the modification index test, and the expected parameter change (EPC) (Van der Veld Saris, 2018) for identifying problematic parameters.

Achieving metric and scalar invariance is challenging using MGCFA. One strategy within the MGCFA approach is the idea of partial metric and partial scalar invariance which is based on a relaxation of parameters based on modification indices to improve model fit. A country comparison is already feasible if the construct is measured with a minimum of two items with equal factor loadings or intercepts (Byrne, Shavelson, Muthén, 1989; Steenkamp Baumgartner, 1998; Pokropek et al., 2019).

2.2 Alignment

An alternative strategy to deal with a lack of scalar MI is to use BSEM (Muthén Asparouhov, 2012; van de Schoot et al. 2013) or alignment (Asparouhov Muthén, 2014). Alignment can be conducted using either Maximum Likelihood or Bayesian estimation. In this study, we focus on alignment using Maximum Likelihood estimation since it is particularly interesting for applied researchers due to its ease of implementation and the sufficient efficiency of the Maximum Likelihood estimation (Kim, Cao, Wang, Nguyen, 2017).

The alignment procedure estimates factor means, variances, and loadings without constraining loadings and intercepts to be equal across groups (Byrne van de Vijver,

2017). The baseline model in alignment is based on the configural model and has the factor means set to 0 and the factor variance set to 1. The second step of alignment consists in an optimization of the configural model through a simplicity function which is similar to a factor rotation in an exploratory factor analysis. Factor means and variances are freely estimated. During this procedure, factor loadings and intercept parameters with the same likelihood estimation as in the configural model are searched for every group factor mean and variance parameter. The goal of the process is to find values for factor means and factor variances that minimize the total amount of noninvariance with few large noninvariant parameters and many approximately invariant parameters (Asparouhov Muthén, 2014; Byrne van de Vijver, 2017). Based on simulation results, Asparouhov and Muthén (2014) suggest a maximum of about 25 percent non-invariant parameters as a cut-off for trustworthy estimates in the alignment estimates. Recent studies even propose a maximum of 29 percent of non-invariant intercepts (Flake McCoach, 2018).

Alignment performs well with many but also with few groups. Asparouhov and Muthén (2014) demonstrated via a Monte-Carlo Simulation that alignment worked well with two and three groups. This was confirmed by Finch (2016) as well as for two groups by DeMars (2020) and for two and four groups by Pokropek et al. (2020). Additionally, Muthén and Asparouhov (2018) found that both small groups and low numbers of indicators per construct are adequate if sample sizes are large.

The advantage of alignment is that the latent means and loadings are estimated while taking into account real differences in loadings and intercepts. This leads to more trustworthy mean values than MGCFA (Cieciuch et al., 2014). Since the optimization process is automated, testing is simplified (Asparouhov Muthén, 2014; Byrne van de Vijver, 2017; Munck et al., 2018) and it prevents "wrong" model respecification through the researcher (Byrne van de Vijver, 2017).

MI tests using MGCFA or approximate approaches like alignment allow researchers working with secondary data to assess the equivalence of data and help to locate incomparable items and groups (Meitinger, 2017). However, it is important to note that MGCFA as well as alignment look at the statistical properties of the data. Statistical properties are a necessary but not a sufficient condition for comparability. Measurement consists of statistical features but also a substantive understanding of the concept or relevant key terms. It might be possible that statistical equivalence is confirmed, but data are still incomparable because respondents differ in their associations (concept and item bias).

As van de Vijver points out: "Sources of BIAS can be easily overlooked in standard equivalence tests based on confirmatory factor analysis, thereby reaching overly lib-eral conclusions about equivalence. Thus construct inequivalence cannot be identified in deductive equivalence testing (i.e. testing in which only data from a target instrument are available, as is the case in confirmatory factor analysis)" (van de Vijver, 2018, p.32). Therefore, quantitative approaches need to be combined with qualitative insights. In addition, it might be possible that MGCFA and alignment differ in their ability to capture these nuances related to concept and item bias.

3 Web Probing and Mixed Methods Approaches

Several qualitative approaches exist that help to assess the cross-national comparability of questions. For example, in cognitive interviewing respondents answer survey questions and during the interview produce "additional verbal information about survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that its author intends" (Beatty Willis, 2007, p. 287). Respondents receive follow-up questions called "probes" to retrieve additional information regarding the validity of items. Several probe types exist, such as comprehension probes that encourage respondents to define how they understand a specific term or expression (Prüfer Rexroth, 2005; Willis, 2005). To assess the comparability of questions, cross-cultural cognitive interviewing (CCCI) can be conducted across different countries or cultural groups (e.g., Willis, 2015).

3.1 Web Probing

Web probing is "the implementation of probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions" (Behr et al., 2017, p. 1). When implemented in cross-national surveys, it is also a valuable tool to assess the comparability of respondents' associations but also to reveal the reasons for incomparability. The implementation in web surveys provides large sample sizes which allows for an evaluation of the prevalence of problems or themes as well as an analysis of response patterns or specific subpopulations. Additionally, it has a broader geographical coverage than traditional cognitive interviewing (Behr et al., 2017; Edgar, Murphy, Keating, 2016; Meitinger Behr, 2016; Meitinger, Braun, Behr, 2018). WP is particularly useful for detecting cases of construct bias (construct is not identical across countries) and item bias (e.g., bias due to ambiguous source items, poor item translation, inapplicability of item contents; van de Vijver Poortinga, 1997; He van de Vijver, 2012). However, the qualitative nature of the probe responses limits the analysis to a small number of questions and countries

since the data analysis is rather extensive (potential translation of responses, development of coding schema, coding of responses) (Behr, 2015; Meitinger, 2017).

3.2 Mixed Methods Approaches

The awareness in comparative survey research is increasing that the complexity of creating and assessing cross-national data should be tackled with multiple methodologies (e.g., Smith, 2010). Although mixed methods approaches in cross-national research are still scarce (Van de Vijver Chasiotis, 2010), previous research already combined cognitive interviewing and exploratory as well as confirmatory factor analysis (CFA) (Latcheva, 2011), cognitive interviewing, multidimensional scaling, and CFA (Efremova et al., 2017), differential item functioning analysis and cognitive interviewing (Benítez Padilla, 2014), as well as MGCFA and WP (Meitinger, 2017). The unique contribution of this article is the combination of WP with alignment and MGCFA for invariance testing in a large cross-national survey (ISSP 2013 National Identity module) and its application to items measuring citizen evaluations of patriotism.

4 Substantive Application: Citizen Evaluation of Patriotism

There has been a long discussion concerning the definition and operationalization of the concepts of nationalism, patriotism, and national identity in political science (Huddy, 2016; Mylonas Tudor, 2021), sociology (Bonikowski, 2016; Mußotter, 2021), and social psychology (Roccas Berlin, 2016). To capture the different aspects of national identity, a special module on national identity was launched in the International Social Survey Program (ISSP) in 1995 and (partially) replicated in 2003 as well as 2013 (ISSP Research Group, 2015). The module contains items to measure concepts such as nationalism and patriotism (Höllinger Hadler, 2012), which are two concepts many researchers studying national identity distinguish (Feshbach Sakano, 1997; Huddy Khatib, 2007; Kosterman Feshbach, 1989; Sidanius, Feshbach, Levin, Pratto, 1997).

Some researchers perceive patriotism in a very general sense as "degree of love and pride in one's nation" (Kosterman Feshbach, 1989, p. 271). Other researchers – in the tradition of Habermas' (1992) constitutional patriotism – additionally mention a 'critical loyalty,' questioning and criticism of current group practices that are driven by a "desire for positive change" (Schatz et al., 1999, p. 153) or a notion of conditional support that evaluates whether the nation is sufficiently working according to humanistic and democratic principles.

In contrast, definitions of nationalism refer to national superiority and dominance (Kosterman Feshbach, 1989) and a "rigid and unflexible attachment to the country, characterized by unquestioning positive evaluation, staunch allegiance, and intolerance of criticism" (Schatz et al., 1999, p. 151). It is important to mention, that various definitions and empirical operationalizations of patriotism and nationalism exist and that these concepts are very much in dispute (see for example, Bitschnau Mußotter, 2022; Hanson O'Dwyer, 2019; Mußotter, 2021; Satherly et al., 2019).

Several empirical studies already examined the consequences of nationalism and patriotism. Nationalism has been described as "dark side of national attachments" (Huddy, 2016, p. 10) since it correlates with a negative assessment of outgroups, anti-Semitism, derogation of foreigners, racism, and a high tendency for social dominance (De Figueiredo Elkins, 2003; Sidanius et al., 1997; see also Huddy, 2016). Findings regarding consequences of patriotism are inconsistent and inconclusive (Ariely, 2018). On the one hand, empirical research found that patriotism can nurture a joint sense of identity and support toward fellow citizens that increases cohesion in a society (Brown, 1999), leads to citizen support for paying taxes (Gangl et al., 2016), political involvement (Huddy Khatib, 2007), and a reduced rejection of outgroups and anti-Semitism (Blank Schmidt, 2003). On the other hand, previous research revealed negative consequences of patriotism, such as beliefs of "democratic superiority" over other nations (Bar-Tal, 1997) and found that not patriotism in itself but the support of democratic values reduces outgroup derogation (Bar-Tal, 1997; Cohrs et al., 2004; Wagner et al., 2012).

While there is extensive research regarding patriotism, nationalism, and their empirical consequences, there is less research on how citizicens evaluate patriotic feelings and their consequences (Ariely, 2020). Since "the question of how people themselves evaluate patriotism has largely been overlooked" (Ariely 2018, p. 436), the International Social Program wanted to address this research gap by including a new item battery on "Perceived Consequences of Patriotic Feelings" in the 2013 ISSP module on National Identity. The item battery asks respondents about the potential effect of strong patriotic feelings in their country on different issues (e.g., intolerance, feeling of unity) and contains items adressing positive and negative consequences.

So far, the cross-national comparability of the ISSP item battery on "Citizen Evaluation of Patriotism" has not been tested but studies already explored cross-national variations in these measures. In his study with 29 countries, Ariely (2018) showed that patriotism is viewed more positively than negatively in nearly all countries. Two countries are exceptions from this general pattern. Only in Spain and Sweden, respondents frequently associated patriotism with negative consequences. In addition, Spain was the country

were respondents least likely perceived patriotism as something positive.

In his study on the interplay of regional identification, nationalism, and patriotism, Dirksmeier (2022) illustrates that the distinction between patriotism and citizen's evaluation of patriotism matters. He found that positive evaluations of patriotism have more nationalistic connotations than patriotism. Respondents that regarded patriotism as something positive, showed also preferences for low immigration, xenophobia, and nationalist attitudes. Whether the data are indeed cross-nationally comparable is an open research question.

5 Research Objectives

The research objectives for this study are twofold:
The methodological contribution is:

1. an illustration of an innovative mixed methods approach: The combination of MI tests using MGCFA and an approximate approach like alignment with WP
2. a contribution to the current discussion of exact vs. approximate MI:
Alignment is more liberal. But does this reflect that the finding that unbiased equal factor loadings and latent means could be estimated for the different countries in alignment can be interpreted as an indication that the construct is unproblematic in cross-national data analysis? We will show later that this is not the case according to the results of WP.

The substantive contribution is:

1. an assessment of the cross-national comparability of the ISSP item battery "Citizen Evaluation of Patriotism"
2. an assessment whether the term "patriotic feelings" is equally understood by respondents in different countries.

6 Data Methods

6.1 Tested Item Battery

The item battery on "Citizen Evaluation of Patriotism" in the 2013 ISSP module on National Identity (ISSP Research Group, 2015) serves as our substantive example. Contrary to previous measures of patriotism, this newly introduced item battery explicitly focusses on the respondents' perceptions of patriotism and asks the respondents for their evaluation of patriotisms' positive and negative consequences (Ariely, 2018). It asked respondents' level of agreement that strong patriotic feelings in their country a) strengthen (COUNTRY's) place in the world (Pat_a), b) lead to intolerance
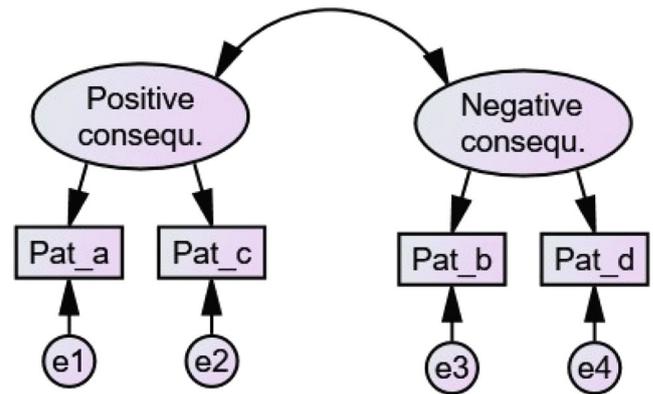

Fig. 1

Expected factor structure of ISSP item battery "Citizen Evaluation of Patriotism"

in (COUNTRY) (Pat_b), c) are needed for (COUNTRY) to remain united (Pat_c), and d) lead to negative attitudes towards immigrants in (COUNTRY) (Pat_d). Respondents had to choose their answers on a five-point scale ranging from *agree strongly* to *disagree strongly*. According to the research group that developed this item battery, it should consist of two scales with item a) and c) addressing the positive consequences of patriotism and item b) and d) tapping the negative consequences of patriotism (ISSP National ID report, 2015). The path diagram for the postulated measurement model has a two-dimensional factor structure (see Fig. 1).

We built on the research of Ariely (2018) who found a two-dimensional structure in this item battery when conducting an exploratory factor analysis.

Regarding this item battery, we wondered whether respondents understand the expression "patriotic feelings," whether respondents' associations differ across countries, and whether this reduces the cross-national comparability. We had the following concerns: The expression "patriotic feelings" is not part of everyday language in most countries; therefore, the respondents might not grasp the intended meaning of this expression. The term might not measure patriotism as intended but something distinct, such as na-

Fig. 2

Screenshot of Comprehension Probe in Web Survey

Table 1

Unconstrained Model: Unstandardized and Standardized Factor Loadings, Standard Errors, and Factor Correlations

	Germany		Great Britain		U.S.		Spain		Mexico	
	Loading	Err.	Loading	Err.	Loading	Err.	Loading	Err.	Loading	Err.
Unstandardized solution										
Dimension I										
Pat_a	1		1		1		1		1	
Pat_c	1.05	0.12	0.86	0.12	0.74	0.11	1.15	0.08	0.80	0.17
Dimension II										
Pat_b	1		1		1		1		1	
Pat_d	1.33	0.18	1.14	0.17	1.31	0.20	1.17	0.10	1.02	0.23
Standardized solution										
Dimension I										
Pat_a	0.73	0.04	0.87	0.06	0.85	0.06	0.79	0.03	0.89	0.09
Pat_c	0.80	0.05	0.77	0.05	0.64	0.05	0.90	0.03	0.69	0.08
Dimension II										
Pat_b	0.63	0.04	0.74	0.06	0.61	0.05	0.71	0.03	0.71	0.08
Pat_d	0.82	0.05	0.83	0.06	0.79	0.06	0.87	0.03	0.74	0.08
Factor correlation	0.34	0.04	0.29	0.04	0.33	0.04	0.45	0.03	–0.21	0.04

Table 2

Results: Exact Measurement Invariance Tests for the ISSP 2013 Data Set

	Chi^2	Df	CFI	Δ	RMSEA	Δ
All countries						
Configural model	1.059	5	1.000		0.000	
Metric model	14.921	13	1.000	0.000	0.011	0.011
Scalar model	88.002	21	0.987	0.013	0.052	0.041
Model without Spain						
Configural model	0.828	4	1.000		0.000	
Metric model	6.443	10	1.000	0.000	0.000	0.000
Scalar model	41.285	16	0.993	0.007	0.037	0.037

tionalism. Additionally, the term might be affected by social desirability effects. Previous research on other elements of national identity found that the answer selection can be driven by the degree in which it is acceptable to express national pride (Fleiß, Höllinger, Kuzmics, 2009; Meitinger, 2018). Similar effects might appear for "patriotic feelings." To address both research questions (MI and meaning of "patriotic feelings"), we analyzed this item battery with MGCFA, alignment, and WP.

6.2 Data for MGCFA Alignment

We assessed the MI of the item battery on the basis of the 2013 ISSP data set of the National Identity module (ISSP Research Group, 2015). For our quantitative analysis, we selected the five countries that we also assessed with WP:

Germany ($N = 1717$), Great Britain ($N = 904$), the U.S. ($N = 1274$), Mexico ($N = 1062$), and Spain ($N = 1225$). We tested the exact MI and alignment with Mplus Version 8.4 (Muthén Muthén, 2015) using Maximum Likelihood estimation.

6.3 Data for Web Probing

We replicated questions from the 2013 ISSP module on National Identity with a web survey with 2685 respondents from nonprobability online access panels in May 2014. The respondents from Germany, Great Britain, the U.S., Mexico, and Spain were selected with quotas for age (18–30, 31–50, and 51–65), gender, and education (low and high). Respondents received on a separate screen a comprehension probe that asked what they associate with the phrase

Table 3

Results Alignment for the ISSP 2013 Data Set

	Factor Loadings	Intercepts
Dimension I		
Pat_a	1 2 3 4 5	1 2 3 4 5
Pat_c	1 2 3 4 5	1 2 3 4 5
Dimension II		
Pat_b	1 2 3 4 5	1 (2) 3 4 5
Pat_d	1 2 3 4 5	1 2 3 4 5

Noninvariant parameters are parenthesized. Number labels: 1 = Mexico; 2 = Spain; 3 = U.S.; 4 = GB; 5 = Germany

"strong patriotic feelings" (see Fig. 2). We developed a coding scheme based on the probe answers and one researcher and two student assistants coded all 2685 responses after having been trained accordingly. Multiple coding applied for all probe answers. A randomly selected sample of 20 percent of the responses was coded twice to calculate the intercoder reliability which was satisfactory (Holsti's coefficient: 86 percent). The coding team discussed and corrected all instances of deviating coding.

To ensure that the results of our web survey and the ISSP data are comparable, we compared the response distribution of both data sets. Similar mean scores in the ISSP and our web survey are a prerequisite for using the web survey data to suggest explanations for the ISSP data. The comparison shows that the patterns found in the ISSP are nearly reproduced in the web survey (see Table ESM1 in the Appendix).

Table 4

Latent Mean Comparisons for the ISSP 2013 Data Set

Ranking	Country	Mean	Sig. Groups[a]
Dimension I (Positive Consequences)			
1	USA (3)	0.061	4 5 2
2	Mexico (1)	0.000	4 5 2
3	Great Britain (4)	–0.338	5 2
4	Germany (5)	–0.787	2
5	Spain (2)	–1.243	
Dimension I (Negative Consequences)			
1	Mexico (1)	0.000	3 4 2
2	Germany (5)	–0.091	2
3	USA (3)	–0.122	2
4	Great Britain (4)	–0.130	2
5	Spain (2)	–0.362	

Parenthesized values represent country-assigned number within data.
[a]Groups with significantly smaller factor means

6.4 Country Selection

The reported WP results were collected in a Web survey that replicated multiple constructs from the ISSP Module on National Identity (e.g., patriotism, nationalism, national pride, and perceived consequences of patriotism). The country selection for this Web survey was based on the goal to include a diverse set of countries regarding national identity and the possibility to reveal potentially problematic issues. Germany was selected because its national identity is closely intertwined with its history (e.g., the Nazi regime) which led to a pride taboo (Miller-Idriss, 2009) and the "war guilt effect" (Smith Kim, 2006). Great Britain is a multicultural and multinational state. Postwar migration from its previous Empire (McCrone, 2002) and current immigration have created an ethnic pluralization of British society (Tilley Heath, 2007). In addition, the rise of Scottish and Welsh nationalism has increased the importance of "territorial identities" (Bechhofer McCrone, 2010). Territorial identities exist in Spain, too (e.g., Catalonia and the Basque Country; Medrano Gutiérrez, 2001). Following Franco's authoritarian regime, Spain's constitution only dates back to 1978 and it is therefore a relatively young democracy (Muñoz, 2009). The U.S. is an influential military, economic, and cultural superpower (Hutcheson et al., 2004), which is reflected in extremely high national pride (Smith Kim, 2006). The U.S. also has a high level of immigration-related diversity (Schildkraut, 2014). In contrast, Mexico is not a country of immigration but emigration (Theiss-Morse Wals, 2014). Given the discrepancy between the economic power of the U.S. and Mexico, the U.S. serves as the "predominant other" for Mexican national identity, with pride in its past Indian civilization and the Mexican revolution serving as other important features constituting Mexican national pride (Morris, 1999).

7 Quantitative Results

We start with presenting the quantitative results of the exact tests and the alignment approach. We were particularly interested whether MGCFA and alignment come to similar conclusions regarding the comparability of the constructs and whether the two approaches provide indications which countries might be particularly problematic regarding nonequivalence.

Table 5

Percentage of Respondents Mentioning Code

	Germany %	Great Britain %	U.S. %	Mexico %	Spain %
Aspects of National Identity					
Identification	24	12	9	14	17
National pride	17	19	14	11	7
Love for country	8	9	11	20	6
Nationalistic attitudes	3	4	6	3	3
Irrelevance of such emotions	5	1	1	3	3
Specific symbols					
Flag	2	4	9	13	6
National anthem	1	2	1	9	1
Other symbols	3	4	3	14	3
Specific groups					
Army veterans	0	2	10	2	2
Queen royal family	0	10	0	0	0
The people of the country	1	1	1	9	2
Other specific groups	2	4	1	5	2
Values					
Democratic values	1	1	8	8	2
Commitment support	14	13	13	16	10
Defend fight for the country	5	4	8	14	10
Other values	2	1	1	5	1
Characteristics of country					
Culture tradition	2	5	2	9	4
History ancestors	5	6	3	17	6
Other characteristics	1	1	0	2	2
Political orientations					
Political orientations	2	6	2	5	3
Negative Associations					
Negative statements	3	5	4	4	10
Fascism (incl. Franquismo)	0	0	0	0	11
Extremism, radicalism, fanaticism	1	1	1	1	18
Nationalism	3	5	2	1	11
Racism	1	8	2	1	10
Non-Response					
Problems with question wording	10	2	2	2	2
Unintelligible answers	10	8	13	2	4
Refusal of question	1	1	2	0	0
Don't know	5	6	4	2	3
Nothing	13	3	5	2	2
Rest	3	4	8	5	5
N	553	535	531	533	533

7.1 Results Exact Approach: MGCFA of the ISSP Data Set

First, we need to verify that the postulated measurement model exists in each country. The factor loadings on each dimension are sufficiently high, with Great Britain and the U.S. having the most similar factor loadings, whereas Spain and Mexico are most distinct (see Table 1). The factor correlation in each country is weak which supports a two-factor

solution. In contrast to the other countries, the factors in Mexico are negatively correlated.

We continue with the results of the exact MI test (see Table 2). The baseline configural model (equal factor structure) has a very good fit (CFI: 1.00; RMSEA: 0.00) which indicates that the constructs "negative consequences of patriotism" and "positive consequences of patriotism" have the same measurement model in all countries. We also find metric measurement invariance (ΔCFI=0.000 and ΔRMSEA=0.011). Unfortunately, scalar MI tests failed (ΔCFI=0.013 and ΔRMSEA=0.041). Therefore, the intercepts of the items of the latent constructs differ across countries and latent mean comparisons would be biased.

To better understand which countries are particularly problematic for cross-national comparisons, we ran several four-country models while dropping one country at a time and detected that Spain is the country that threatens the MI the most. Although scalar MI still failed based on ΔRMSEA (0.037), all goodness-of-fit indicators show better values than for any other four-country solution.

Modification indices and expected parameter changes are a further important source of information when testing for exact MI. For the metric model, Spain was the country with most of the highest modification indices (4) and their respective expected parameter changes (over 0.10). All modification indices flagged parameters measuring positive consequences of patriotism (Pat_a and Pat_c). This finding mirrors our results regarding the overall assessment of model fit that flagged Spain as the most deviating country in our study. For the scalar model, high modification indices in combination with expected parameter changes above 0.10 for the intercepts were found twice in Mexico, one time in Spain and the U.S., and none in Germany and Great Britain.

To summarize the results of the exact MI testing: The exact approach using MGCFA indicates that comparability can only be found on the configural and metric level and MGCFA flags Spain as the most deviating country.

7.2 Results Alignment

The alignment results differ from the exact approach (see Table 3). All factor loadings were invariant. Only one of the intercepts turned out to be non-invariant in the ISSP data set which is well below the recommended cut-off value of 25 percent of deviant parameters (Asparouhov Muthén, 2014). According to the alignment results, an estimation of unbiased equal factor loadings and latent means was possible for all countries.

Table 4 contains the latent means and their significant differences in the five countries. The numbers in the fourth column refer to the numbers given to the countries in parentheses in the second column. The U.S. respondents per-

ceived the strongest positive consequences of strong patriotic feelings. The differences are nonsignificant compared with Mexico, but the latent means of Germany, Spain, and Great Britain are significantly lower. Mexico has the highest value on the dimension of negative consequences of patriotic feelings (see Smith Kim, 2006) and Spain has the lowest value which is significantly lower than the latent means of all other countries.

7.3 Results Web Probing

We evaluated with a comprehension probe whether respondents have similar associations with regard to the expression "patriotic feelings." Based on the probe responses, we developed a coding schema.

7.3.1 Coding schema

Respondents in the five countries had a large variety of associations in mind when they responded to this comprehension probe. Many respondents associated different *aspects of national identity*, particularly identification with the country, national pride, love for the country, and nationalistic attitudes. In contrast, some respondents also mentioned the irrelevance of such emotions. Respondents also referred to *specific symbols*, such as flags, national anthems, and specific groups, such as the army and veterans, the Queen and the royal family as well as the people of the country. Also, respondents connect *specific values* with patriotic feelings, such as democratic values, the need to commit to and support as well as defend and fight for the country. Further associations were *specific characteristics of the country*, such as the country's culture and traditions as well as its history and ancestors, and specific political orientations. Several respondents had *negative associations*: respondents wrote very general negative statements but were also thinking more specifically about fascism, extremism, nationalism, and racism. The coding schema also controlled for *respondents' problems with the expression* and different types of *probe nonresponse*, such as nonsense and unintelligible answers, refusals, don't know as well as "nothing" responses. Finally, any specific symbols, groups, values, and country characteristics that were not mentioned by more than five percent of respondents in any country were coded in *specific other categories*. Any responses that did not classify for any substantive or nonresponse code were coded in the *rest* category. Table ESM2 in the online Appendix provides an overview of the different codes, code descriptions, and examples for each code.

7.3.2 Results qualitative analysis

Respondents had various associations (see Table 5). The majority of respondents in each country mentioned specific aspects of national identity. Many German (24 percent) and Spanish (17 percent) referred to identification with the country. Around one-fifth of the British respondents defined patriotic feelings as national pride. Mexican respondents (20 percent) most frequently thought about love for the country. In each of the countries, few respondents thought about nationalistic attitudes or the irrelevance of such emotions.

With regard to specific symbols, several U.S. and Mexican respondents mentioned flags (9 and 13 percent respectively). Also, the national anthem seems to be more relevant in Mexico (9 percent) than in the other countries.

The specific groups that respondents have in mind when answering this probe seem to be country-specific. U.S. respondents associate most often the army and veterans (10 percent), whereas British respondents primarily think of the Queen and the royal family (10 percent), and Mexicans of their fellow citizens (9 percent).

Respondents in all countries also referred to specific values, especially the need to commit to and support one's country which was mentioned by at least 10 percent of respondents in each country. Mexican and U.S. respondents also associated democratic values with this expression (8 percent each) and respondents from both countries and Spain pointed to the need to defend and fight for the country. Mexican respondents mentioned most often specific characteristics of the country, such as culture and tradition as well as history and ancestors.

The most striking result of this probe is the large number of negative associations of Spaniards. Overall, they most frequently provided general negative statements (10 percent) and thought most often of fascism (11 percent), extremism (18 percent), nationalism (11 percent) as well as racism (10 percent). Spaniards seem to associate the term "patriotic feelings" with the Franco regime. The term "patriotic feelings" reminded one respondent "of the typical and rancid Spanish nationalism, heir of the Francoism [...]. Decisions on the future of a country should not be the result of conviction, but of reflection. (*own translation*[1], male, 39 years)

Another respondent explained the origin of this association. When reading the term "patriotic feelings," he thinks of

"the legacy of Francoism, which was taught to us in

schools well into the 1970s. The myth of one, great and free, the supposed superiority over other nationalities, feeling proud of one's own while despising that of others, the glorification of one's roots, etc. (*own translation*[2], male, 51 years)

The association with the Franco regime seems again to trigger various negative connotations. For example, this Spaniard associates the following with the term "patriotic feelings":

"phrases like: everything for the homeland, which remind me of Franco's times, of disdain for immigrants, of the marginalized, of machismo, and of the brutality of the armed forces (*own translation*[3], female, 27 years)

The association with the Franco regime also seems to trigger feelings of shame, as one of the Spanish respondents mentions:

"I think that the United States is very patriotic, so is France, so is Germany, and I think that we, after the dictatorship, have national shame for associating the flag, for example, with the dictatorship, without having anything to do with it." (*own translation*[4], female, 58 years)

Surprisingly, only a few German respondents had negative associations with regard to "patriotic feelings." Given previous research on the perception of national pride in Germany and the existence of a "national pride taboo" (Miller-Idriss, 2009), the near absence of negative associations is unexpected. However, far more German respondents indicated problems with the wording (10 percent) than respondents from the other countries (all 2 percent). In a similar vein, Germans provided more frequently "nothing" responses (13 percent). Both categories might be indications that Germany respondents still have issues with this expression but do not show this with negative associations like the Spanish respondents.

The reason for the various and unexpected associations in the different countries may be located in the rendering

---

[1] Spanish original: "Al típico y rancio nacionalismo español, heredero del franquismo [...]. Las decisiones sobre el futuro de un país no deberán ser fruto de la convicción, sino de la reflexión.".

[2] Spanish original: "La herencia del franquismo, lo que nos enseñaron en las escuelas hasta bien entrados los años 70. El mito de una, grande y libre, la supuesta superioridad sobre otras nacionalidades, el sentirse orgullosos de los propio despreciando lo ajeno, la exaltación de las raíces, etc.".

[3] Spanish original: "frases como: todo por la patria, que me resuenan a tiempos franquistas, al desprecio a los inmigrantes, a los marginados, al machismo, y a la brutalidad de las fuerzas armadas".

[4] Spanish original: "Pienso que los EE UU son muy patrióticos, Francia también, Alemania también y creo que nosotros después de la dictadura, tenemos vergüenza nacional por asociar la bandera por ejemplo a la dictadura, sin tener nada que ver.".

of "patriotic (feelings)" and their different connotations in the five countries. The German questionnaire uses "patriotisch" which is a foreign word and, as such, less likely to become associated with Germany's Nazi past. Therefore, in Germany, the term assumes a neutral to positive meaning, as appears to be the case in Great Britain and the U.S. In Spain, on the contrary, "patrióticos" is reminiscent of "patria," one of the core concepts of the Franco regime. As a consequence of this particular linguistic reason and its associations, the dictatorial and nationalistic past in Spain and Germany has entirely different consequences for the response behavior.

A possible alternative explanation, that the critical stance to patriotism is restricted to those parts of Spain in which a large minority imagines themselves as a proper nation in more or less pronounced opposition to the Spanish state, e.g., to Catalonia and the Basque Country, can be largely excluded on the basis of our data: In Catalonia and the Autonomous Region of Madrid, roughly half of the respondents mention one of the subcodes of the fascism category. This seems to be largely independent of the fact that, for instance in Catalonia and the Basque country, regional identity is very strong compared to national identity.

In Mexico, there is at least folkloric evidence that the term patriotic is perceived by many in the context of the vague concept of "Mexicanidad" which combines pride in the creation of a new nation by mestizaje of the indigenous population with the Spanish and immigrants from other European countries and love for countryside, history and so on. According to Vizcaíno Guerra (2004, p. 143) the modern nationalism in Mexico "corresponde con la apertura al mundo, apertura económica y también política, la tolerancia y el reconocimiento de las minorías culturales [corresponds with openness to the world, economic and also political openness, tolerance and recognition of cultural minorities]." These minorities are associated with indigenous people in Mexico to a high degree (about 20 percent in Mexico compared to 2 percent in the U.S.) when respondents are asked for their associations of the term "minorities". In this case, nationalism has a very benign form and the positive stance of respondents to the concept of patriotism does not come as a surprise.

8 Discussion

This study provides an intriguing contribution to recent developments in methodological studies on exact versus approximate MI. In our study, we find a similar pattern as in previous studies using exact and approximate approaches of MI testing (e.g., Lomazzi, 2018; Munck et al., 2018): we could find configural and metric invariance when applying the exact approach of MI testing (MGCFA). In MGCFA,

Spain was flagged as the country that reduced the comparability the most in our country sample, indicating that Spanish respondents somehow differ from respondents in the other four countries on how they use the latent scales. In the alignment estimation, only one intercept was flagged as noninvariant which is well below the recommended cut-off value of 25 percent noninvariant parameters (Asparouhov Muthén, 2014). The alignment analysis indicated that unbiased equal factor loadings and latent means could be estimated for all countries.

However, the qualitative WP results reveal that the alignment results might not capture all issues of comparability (i.e. biases) in our country sample. Spanish respondents tend to have negative associations with the term "patriotic feelings." The term triggers associations with the Franco regime with many Spaniards and influences their response behavior.

Interestingly, Spain was also the country that reduced the cross-national comparability the most in the exact MI tests. Therefore, the results mirror the WP findings. In contrast, the alignment procedure did not detect this issue.

These findings provide some indication that the alignment procedure might not detect necessarily all issues of cross-national comparability. It seems that alignment provides information on the statistical MI but might miss aspects of item and construct bias. Interestingly, MGCFA did detect these nuances. The current finding that MGCFA mirrors qualitative findings in WP was already found in previous studies (Meitinger, 2017). Therefore, we recommend using a mixed methods approach to assess the cross-national comparability of crucial measures.

8.1 Limitations of Study

For this study, we collected the qualitative data with a nonrandom online access panel because cross-national probability-based web panels did not exist at the time of data collection for all of our countries. However, both data sets showed very similar response distributions when we compared the means in the ISSP data set and our web survey.

8.2 Future Research

The current study provided an intriguing case study on the relationship between MGCFA, alignment, and WP. It gives a first indication that MGCFA and alignment might capture different types of biases. Future studies should disentangle more clearly what both methods can indeed capture and to which degree both methods should be complemented with qualitative insights.

References

Ariely, G. (2018). Evaluations of patriotism across countries, groups, and policy domains. *Journal of Ethnic and Migration Studies*, *44*, 462–481.

Ariely, G. (2020). How people view patriotism: the evidences from cross-national surveys. In *Handbook of Patriotism* (pp. 633–650).

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

Bar-Tal, D. (1997). The monopolization of patriotism. In D. Bar-Tal & E. Staub (Eds.), *Patriotism in the lives of individuals and nations* (pp. 246–270). Chicago: Nelson-Hall.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: the practice of cognitive interviewing. *Public Opinion Quarterly*, *71*(2), 287–311.

Bechhofer, F., & McCrone, D. (2010). Choosing national identity. *Sociological Research Online*, *15*(3), 13–25.

Behr, D. (2015). Translating answers to open-ended survey questions in cross-cultural research: a case study on the interplay between translation, coding, and analysis. *Field Methods*, *27*(3), 284–299.

Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2017). *"Web probing – implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions."* GESIS – Survey Guidelines. GESIS – Leibniz-Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_023.

Benítez, I., & Padilla, J. (2014). Analysis of nonequivalent assessments across different linguistic groups using a mixed methods approach: understanding the causes of differential item functioning by cognitive interviewing. *Journal of Mixed Methods Research*, *8*, 52–68.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bitschnau, M., & Mußotter, M. (2022). (National) pride and (conceptual) prejudice: critical remarks on the distinction between patriotism and nationalism. *Journal of Political Ideologies*, *22*(3), 1–15.

Blank, T., & Schmidt, P. (2003). National identity in a united Germany: Nationalism or patriotism? An empirical test with representative data. *Political psychology*, *24*(2), 289–312.

Bonikowski, B. (2016). Nationalism in settled times. *Annual Review of Sociology*, *42*, 427–449.

Brown, D. (1999). Are there good and bad nationalisms? *Nations and Nationalism*, *5*, 281–302.

Browne, M., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods Research*, *21*, 230–258.

Byrne, B., & van de Vijver, F. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: a paradigmatic cross-cultural application. *Psicothema*, *29*, 539–551.

Byrne, B., Shavelson, R., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Cieciuch, J., Davidov, E., Vecchione, M., Beierlein, C., & Schwartz, S. H. (2014). The cross-national invariance properties of a new scale to measure 19 basic human values: A test across eight countries. *Journal of Cross-Cultural Psychology*, *45*, 764–776.

Cohrs, J. C., Dimitrova, D., Kalchevska, T., Kleinke, S., Tomova, I., Vasileva, M., & Moschner, B. (2004). Ist patriotischer Nationalstolz wünschenswert? *Zeitschrift für Sozialpsychologie*, *35*(4), 201–215.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75.

DeMars, C. E. (2020). Alignment as an alternative to anchor purification in DIF analyses. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 56–72.

Dirksmeier, P. (2022). The relationship between patriotism and regional identification: a cross-country analysis. *The Annals of Regional Science*, 1–20.

Edgar, J., Murphy, J., & Keating, M. (2016). Comparing traditional and crowdsourcing methods for pretesting survey questions. *Sage Open*, *6*(4).

Efremova, M., Panyusheva, T., Schmidt, P., & Zercher, F. (2017). Mixed methods in value research: an analysis of the validity of the Russian version of the Schwartz Value Survey (SVS) Using Cognitive Interviews, Multidimensional Scaling (MDS), and

Confirmatory Factor Analysis (CFA). *ASK. Research and Methods*, *26*, 3–30.

Feshbach, S., & Sakano, N. (1997). The structure and correlates of attitudes toward one's nation in samples of United States and Japanese college students: A comparative study. In D. Bar-Tal & E. Staub (Eds.), *Patriotism: In the lives of individuals and nations* (pp. 91–107). Nelson-Hall Publishers.

de Figueiredo, R., & Elkins, Z. (2003). Are patriots bigots? An inquiry into the vices of in-group pride. *American Journal of Political Science*, *47*, 171–188.

Finch, W. H. (2016). Detection of differential item functioning for more than two groups: a Monte Carlo comparison of methods. *Applied Measurement in Education*, *29*(1), 30–45.

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

Fleiß, J., Höllinger, F., & Kuzmics, H. (2009). Nationalstolz zwischen Patriotismus und Nationalismus? *Berliner Journal für Soziologie*, *19*, 409–434.

Fox, G. J. A., & Verhagen, J. (2018). Random item effects modeling for cross-national survey data. In E. Davidov, P. Schmidt, J. Billiet & B. Meuleman (Eds.), *Cross-cultural analysis: methods and applications* (pp. 529–550). New York: Routledge.

Gangl, K., Torgler, B., & Kirchler, E. (2016). Patriotism's impact on cooperation with the state: an experimental study on tax compliance. *Political Psychology*, *37*(6), 867–881.

Habermas, J. (1992). Citizenship and national identity: some reflections on the future of Europe. *Praxis International*, *12*(1), 1–19.

Hanson, K., & O'Dwyer, E. (2019). Patriotism and nationalism, left and right: AQ-methodology study of American national identity. *Political Psychology*, *40*(4), 777–795.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*, 1–18.

Höllinger, F., & Hadler, M. (2012). *Crossing borders, shifting boundaries. National and transnational identities in Europe and beyond*. Frankfurt New York: Campus.

Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: techniques and applications*. New York: Routledge.

Huddy, L. (2016). Unifying national identity research: interdisciplinary perspectives. In J. Grimm, L. Huddy, P. Schmidt & J. Seethaler (Eds.), *Dynamics of na-tional identity. Media and societal factors of what we are* (pp. 22–43). New York: Routledge.

Huddy, L., & Khatib, N. (2007). American patriotism, national identity, and political involvement. *American Journal of Political Science*, *51*(1), 63–77.

Hutcheson, J., Domk, D., Billeaudeaux, A., & Garland, P. (2004). US national identity, political elites, and a patriotic press following September 11. *Political Communication*, *21*(1), 27–50.

ISSP National ID report (2015). A compilation of documents used to develop the National Identity III Questionnaire for the International Social Survey Program in 2014. https://dbk.gesis.org/dbksearch/download.asp?db=Eid=56372. Accessed 17 May 2018.

ISSP Research Group (2015). *International Social Survey Programme: National Identity III— ISSP 2013. ZA5950 Data file version 1.0.0.* Cologne: GESIS Data Archive.

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

Kim, E., Cao, C., Wang, Y., & Nguyen, D. (2017). Measurement invariance testing with many groups: a comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 524–544.

Kosterman, R., & Feshbach, S. (1989). Toward a measure of patriotic and nationalistic attitudes. *Political Psychology*, *10*, 257–274.

Latcheva, R. (2011). Cognitive interviewing and factor-analytic techniques: a mixed method approach to validity of survey items measuring national identity. *Quality Quantity*, *45*, 1175–1199.

Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, Data, Analyses*, *12*, 77–103.

McCrone, D. (2002). Who do you say you are? Making sense of national identities in modern Britain. *Ethnicities*, *2*(3), 301–320.

Medrano, J. D., & Gutiérrez, P. (2001). Nested identities: national and European identity in Spain. *Ethnic and Racial Studies*, *24*(5), 753–778.

Meitinger, K. (2017). Necessary but insufficient: why measurement invariance tests need online probing as a complementary tool. *Public Opinion Quarterly*, *81*(2), 447–472.

Meitinger, K. (2018). What does the general national pride item measure? Insights from web probing. *International Journal of Comparative Sociology*, *59*(5-6), 428–450.

Meitinger, K., & Behr, D. (2016). Comparing cognitive interviewing and online probing: do they find similar results? *Field Methods*, *28*(4), 363–380.

Meitinger, K., Braun, M., & Behr, D. (2018). Sequence matters in web probing: the impact of the order of probes on response quality, motivation of respondents, and answer content. *Survey Research Methods*, *12*(2), 103–120.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Miller-Idriss, C. (2009). *Blood and culture*. Duke University Press.

Millsap, R. (2011). *Statistical approaches to measurement invariance*. New York: Taylor Francis.

Morris, S. D. (1999). Reforming the nation: Mexican nationalism in context. *Journal of Latin American Studies*, *31*(02), 363–397.

Munck, I., Barber, C., & Torney-Purta, J. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICSS. *Sociological Methods Research*, *47*(4), 687–728.

Muñoz, J. (2009). From national-catholicism to democratic patriotism? Democratization and reconstruction of national pride: the case of Spain (1981–2000). *Ethnic and Racial Studies*, *32*(4), 616–639.

Mußotter, M. (2021). We do not measure what we aim to measure: testing Three Measurement Models for Nationalism and Patriotism. *Quality Quantity*, 1–21.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335.

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, *5*, 978.

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many group: alignment and random effects. *Sociological Methods and Research*, *47*, 637–664.

Muthen, L.-K., & Muthen, B. O. (2015). Mplus User's Guide. Los Angeles: Muthen Muthen.

Mylonas, H., & Tudor, M. (2021). Nationalism: What we know and what we still need to know. *Annual Review of Political Science*, *24*, 109–132.

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

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

Prüfer, P., & Rexroth, M. (2005). *Kognitive Interviews*. ZUMA How-to-Reihe, Nr. 15. Mannheim: ZUMA.

Roccas, S., & Berlin, A. (2016). Identification with groups and national identity: applying multidimensional models of group identification to national identification. In J. Grimm, L. Huddy, P. Schmidt & J. Seethaler (Eds.), *Dynamics of national identity. Media and societal factors of what we are* (pp. 22–43). New York: Routledge.

Satherly, N., Yogeeswaran, K., Osborne, D., & Sibley, C. (2019). Differentiating between pure patriots and nationalistic patriots: A model of national attachment profiles and their socio-political attitudes. *International Journal of Intercultural Relations*, *72*(1), 13–24.

Schatz, R., Staub, E., & Lavine, H. (1999). On the varieties of national attachment: blind versus constructive patriotism. *Political Psychology*, *20*, 151–174.

Schildkraut, D. J. (2014). Boundaries of American identity: evolving understandings of 'us. *Annual Review of Political Science*, *17*, 441–460.

Sidanius, J., Feshbach, S., Levin, S., & Pratto, F. (1997). The interface between ethnic and national attachment: ethnic pluralism or ethnic dominance? *Public Opinion Quarterly*, *61*, 102–133.

Smith, T. (2010). The globalization of survey research. In J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg, P. Mohler, B. Pennell & T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 475–484). Hoboken: Wiley-Blackwell.

Smith, T., & Kim, S. (2006). National pride in comparative perspective: 1995/96 and 2003/04. *International Journal of Public Opinion Research*, *18*, 127–136.

Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

Theiss-Morse, E., & Wals, S. (2014). *Con La Vara que Midas... National Identity and Attitudes toward Immigration and Emigration in Mexico*. Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC, August 28–31, 2014.

Tilley, J., & Heath, A. (2007). The decline of British national pride. *The British Journal of Sociology*, *58*(4), 661–678.

Van der Veld, W., & Saris, W. (2018). Measurement equivalence testing 2.0. In E. Davidov, P. Schmidt, J. Billiet & B. Meuleman (Eds.), *Cross-cultural analysis: methods and applications* (pp. 281–316). New York: Routledge.

Van de Vijver, F., & Poortinga, Y. (1997). Towards an integrated analysis of bias in cross-cultural assessment.

European Journal of Psychological Assessment, *13*, 29–37.

Van de Vijver, F. J., & Chasiotis, A. (2010). Making methods meet: mixed designs. In J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg, P. Mohler, B. Pennell & T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 455–473). Hoboken: Wiley-Blackwell.

Van de Vijfer, F. J. R. (2018). Capturing Bias in Structural Equation Modeling, In E. Davidov, P. Schmidt, J. Billiet & B. Meuleman (Eds.), *Cross-cultural analysis: methods and applications* (pp. 3–44). New York: Routledge.

Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (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.

Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, , 4–70.

Vizcaino Guerra, F. (2004). *El nacionalismo mexicano en los tiempos de la globalización y el multiculturalismo*. UNAM.

Wagner, U., Becker, J. C., Christ, O., Pettigrew, T. F., Schmidt, P. (2012). A longitudinal test of the relation between German nationalism, patriotism, and outgroup derogation. *European Sociological Review*, *28*(3), 319–332.

Willis, G. B. (2005). *Cognitive interviewing: a tool for improving questionnaire design*. London: SAGE.

Willis, G. B. (2015). The practice of cross-cultural cognitive interviewing. *Public Opinion Quarterly*, *79*(S1), 359–395.