ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Jacobsen, Jannes; Fuchs, Lukas Marian

Article — Published Version

Can We Compare Conceptions of Democracy in Cross-Linguistic and Cross-National Research? : Evidence from a Random Sample of Refugees in Germany

Social Indicators Research

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Jacobsen, Jannes; Fuchs, Lukas Marian (2020) : Can We Compare Conceptions of Democracy in Cross-Linguistic and Cross-National Research? : Evidence from a Random Sample of Refugees in Germany, Social Indicators Research, ISSN 0303-8300, Springer, Berlin, Vol. 151, Iss. 2, pp. 669-690,
https://doi.org/10.1007/s11205-020-02397-6

This Version is available at:
https://hdl.handle.net/10419/232011

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

http://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Social Indicators Research (2020) 151:669–690
https://doi.org/10.1007/s11205-020-02397-6

ORIGINAL RESEARCH

Can We Compare Conceptions of Democracy in Cross-Linguistic and Cross-National Research? Evidence from a Random Sample of Refugees in Germany

Jannes Jacobsen[1] · **Lukas Marian Fuchs**[2]

Accepted: 25 May 2020 / Published online: 5 June 2020
© The Author(s) 2020

Abstract
This study addresses the heated academic and public debate on the compatibility and comparability of refugees' and host societies' democratic values. Comparative values research has long capitalized on global similarities and differences in support for Western democratic values. We argue that such cross-cultural comparisons of culturally diverse groups are challenged by (1) different conceptions of democracy determined by different experiences with democratic systems and (2) bias introduced by linguistic differences and translation processes. In order to analyze whether the conception of democracy is comparable between different nationalities and languages, we test data from the German IAB-BAMF-SOEP Survey of Refugees and the world values survey (WVS) for measurement invariance using multi-group confirmatory factor analysis (MGCFA). Applying strict and conservative criteria for measurement invariance and fit indices, our results suggest that the applied democracy scales are problematic for comparing conceptions of democracy between refugees and Germans and across languages. However if more lenient criteria regarding partial invariance and fit indices are considered acceptable, mean comparisons could be carried out between language groups and between groups of refugees.

Keywords Democracy · Measurement invariance · Migration · Refugee · Integration · Values

1 Introduction

With the unprecedented influx of refugees to Europe between 2015 and 2018, an intense public debate arose in Germany over how to accommodate and integrate the new arrivals. One crucial aspect of this debate was the fear that the newcomers do not share fundamental values of the host society (Banulescu-Bogdan and Benton 2017). Studies

✉ Jannes Jacobsen
jjacobsen@diw.de

[1] Socio-Economic Panel (SOEP) at German Institute for Economic Research (DIW), Berlin, Germany

[2] Freie Universität Berlin, Berlin, Germany

began to address the question of how to convey German values to the newly arriving refugees and how to measure their agreement with these values (Banulescu-Bogdan and Benton 2017; Müller-Hilmer and Gagné 2018; SVR 2019). From a political and sociological perspective, this question is important for various reasons. Value consensus—a group's collective agreement with certain fundamental ethics and ideals (Parsons 1968; Wan et al. 2007)—is said to enhance social cohesion by promoting cooperation and simultaneously preventing conflict (Partridge 1971). Following this line of argumentation, it should be possible to predict conflict or cooperation between immigrants and the receiving society by assessing whether the two groups share the same values.

In the public and political debate in Germany and Europe, liberal democratic values are often conflated with European or national values and have been studied extensively. In this research, they are described as the foundation of stable liberal democracies because they mirror the aspiration to support and be actively involved in political processes, the central arena of societal participation (Diamond and Linz 1989; Shin 2007). The use of inferential statistics in analyzing and comparing democratic values has long been a focus of interest in comparative values studies producing a vast body of literature comparing democratic values between countries or cultural entities: the results served as proxies for the democratic condition of a state and the chances that a country will become (or remain) democratic (Diamond and Plattner 2015; Linz and Stepan 1996). Outside academia, results from such general population surveys tend to be used to fuel those concerns pertaining to refugees and immigrants as threats to European or "Western" values. Such assumptions are rash for several reasons: First, comparisons from general population surveys, cannot be extrapolated to highly selective samples of migrants and refugees and are thus unfit to draw conclusions on wether or not refugees challenge European democratic values. Second and more importantly, more recent research strongly warrants a more cautious approach to comparing complex concepts like values and attitudes, indicating that especially attitudes toward democracy in particular are at risk of not being comparable across political cultures (e.g., Ariely and Davidov 2011). If values, democratic or else, are not comparable between cultures, it would be almost impossible to analyze whether people share the same values and thus also whether immigrants threaten certain values of the host society. Numerous recent studies critisize comparative values research for often ignoring or failing to establish necessary levels of measurement equivalence (Davidov et al. 2014), a precondition for any study involving cross-cultural comparison of democratic values (Canache 2012).

Focussing on the influx of refugees into Germany as a case study this article expands the current academic debate on measurement equivalence, demonstrating that even in a study that is conducted in the same cultural and historical context (Germany) careful assessment of whether conceptions of democracy are comparable amongst refugee respondents from diverse backgrounds is mandatory. In a number of publications on this topic, such tests are lacking (e.g., Brücker et al. 2016 use a German panel study on refugees and pool it with the WVS; Buber-Ennser et al. 2016 do the same in Austria), rendering the empirical and statistical comparisons of conceptions of democracy between respondents of different origins and their conclusions about their democratic values flawed. We thus contribute to and advance the current research in two aspects: First, because migrants and refugees self-select nonrandomly to the receiving country, we assess whether cross-cultural and cross-national value consensus is actually challenged within the receiving country. Second, we pose and discuss this important question in light of the heated public debate on value compatibility in many refugee receiving countries since 2015. Many voices in Germany

explicitely claim that refugees do not share the fundamental values of western societies, however, reliable and valid empirical evidence on this matter is still missing to date.

We develop the theoretical argument that even in surveys that are conducted in a single national context, two main factors may pose a challenge to comparability. The first of these is experience with democratic systems. Different countries of origin are here considered indicators for people's past experience with democratic systems. Because there is no clear benchmark definition of democracy, different conceptions of democracy may exist in different populations. The definition of what constitutes a democracy is therefore likely to vary among asylum seekers from diverse and often quite undemocratic backgrounds (a similar argument is made by Ariely and Davidov 2011). Second, from the standpoint of survey methodology, the respondent's language poses additional challenges to comparability (Zavala-Rojas and Saris 2017). This is due both to the fact that words and semantics often do not translate directly from one language to another (Bratton and Mattes 2001), and to the bias introduced by translation itself (Behr et al. 2018; Comanaru and d'Ardenne 2018; Goerman et al. 2018).

We explore both of the aforementioned challenges by analyzing data from the German IAB-BAMF-SOEP Survey of Refugees a unique sample of refugees in Germany and later pooling it with the German world value survey (WVS). We start with an overview of previous comparative research on democratic conceptions and values, illustrating that much of the interpretation to date has been biased by insufficient assessment of measurement invariance. In a second step, we revisit the ongoing academic debate on measurement equivalence and bring forward two theoretical arguments for why we consider measurement invariance imperative when studying a sample of nationally and culturally diverse refugees. In the empirical section, by testing measurement invariance (for an example of the method see Saris et al. 2018), we indicate that democracy remains an ambivalent concept amongst individuals from different political cultures and with different mother tongues and that cross-cultural and cross-linguistic comparisons are likely to be problematic. Hence, in the sensitive context of refugee and migration research a comparison of democratic values between different political cultures using quantitative methods has to be supported by careful considerations of measurement invariance.

2 Lessons Learned from Comparative Values Research

Cultural values are among the most prominent areas of sociological research, not least for measuring possible cultural diversification in the wake of transnational migration. Values are known to assume the role of mediators between individual conceptions of the desirable and undesirable, on the one hand (Marini 2000; see also Kluckhohn 1951), and societal demands, on the other hand (Grube et al. 1994). They thus govern societal cooperation by defining ideal modes of interaction and coexistence as well as determining the conditions for conflict settlement. Agreement to values that foster cooperation and prevent conflict in particular is described as the foundation of liberal civil democracies (Inglehart and Welzel 2005). Democratic values here are of specific prominence as expressions of an individual's satisfaction with society and at the same time, as determinants of the stability of democratic institutions and systems (Inglehart 2000). Empirical research on cross-cultural agreement to democratic and societal values has produced diverse and in part conflicting results to date (for an overview see also Gabriel 2020).

In times of high transnational migration, studies on global differences in values, particularly those providing comparisons between "Western" and "Non-Western" societies, are prominent: Many studies map the globally changing support for the democratic values items in the world value survey (WVS) from a historical perspective (e.g. Welzel 2013). The main argument is that democratic enhancements can be causally explained by socio-economic development and increasing prosperity (Inglehart and Baker 2000; Inglehart and Welzel 2005, 2009). While this narrative seemingly works on an overall scale, it accounts primarily for developments in Western countries. Looking closer, its explanatory power seems limited in explaining the strongly diverging effect sizes of economic development on democratic or equality values between "Western" and "Non-Western" societies.

A competing body of studies, though similar in authorship, thus capitalizes on the cross-cultural differences in democratization, secularism, and gender equality (Inglehart and Welzel 2009; Norris and Inglehart 2012). These studies argue that there is in fact no globally increasing support for democracy and instead accentuate differences in democratic support between what they categorize as liberal, secular Western countries, and clerical, patriarchal non-Western societies, i.e. between refugee sending and receiving societies (Alexander and Welzel 2011; Tausch and Heshmati 2003; Welzel 2013). Inglehart and Norris (2003), for example, concluded that "Muslims and their Western counterparts" desire democracy equally, but at the same time, that Muslims do not share Western egalitarian and equality values. Instead of investigating how values surveys find simultaneously that Muslim and Western societies differ little in their desire for democracy but differ strongly in their conceptions thereof, they translated these findings into a generally pessimist outlook for democracy in Muslim countries. In the same vein, Alexander and Welzel (2011) argued that Muslim support for patriarchal values is robust across time as well as geographic space, irrespective of democratic advancements, vaguely blaming religious and cultural factors but not empirical approaches to democracy research.

These already inconclusive findings, however, have often been more or less directly conferred onto migrants from Non-Western countries, refugees in particular, to foresee cultural clashes and value conflicts (Tausch 2016). Such conclusions are hasty for numerous reasons: First, migrants and refugees are usually a highly selective group compared to those who stay behind and should not be considered representative for their countries of origin (Belot and Hatton 2012; Docquier et al. 2018; Wimmer and Soehl 2014). Second, comparative values studies on a global scale tend to overestimate value homogeneity within countries (Schwartz and Sagie 2000). Within-country variations in democratic values are in fact often stronger than the aggregate differences between countries (Silver and Dowley 2000; Fischer and Schwartz 2011). Finally and most far-reaching, democratic values and conceptions of democracy in refugee countries of origin are likely to differ from those in Western Europe. Using the Arab Barometer, Kostenko et al. (2016) for example demonstrated that democratic values in Arab countries are not linked to gender equality (Kostenko et al. 2016; Rizzo et al. 2007). Meanwhile, Vlas and Gherghina (2012), contest claims about Muslim patriarchy by showing that democratic and equality values are not linked to religion but rather to living in a patriarchal society.

Comparative values studies—despite longstanding and extensive research—in some regards produced controversial and in parts inconclusive results. We argue that comparative values studies have often suffered from an empirical bias and lack of rigor assessments of comparability of value conceptions. This can be particularly harmful where this

engendered hasty conclusions concerning potentially salient areas, such as refugee accommodation. In recent years, however, a growing body of research addressing this issue emerged. The following section discusses these recent developments.

2.1 Testing Perceptions and Attitudes Towards Democracy for Comparability: Recent Findings

From the perspective of empirical and survey-based social science research, before comparing conceptions of democracy or democratic values it is crucial to ask whether the underlying concept of democracy is comparable, meaning that people actually think about the same concept when hearing the term democracy. Only if this is the case comparisons are unbiased. Yet, two seemingly opposing camps are involved in an ongoing academic debate as to how measurement invariance should be assessed—strict proponents of testing constructs' internal measurement invariance on the one hand (e.g. Ariely and Davidov 2011) and those championing constructs' external and aggregate validity (e.g. Welzel and Inglehart 2016).

The majority of studies assessing the measurement invariance of democratic values comes to the conclusion that there are major differences, both cross-culturally and cross-nationally. Using the Arab Barometer, Tessler et al. (2012) used a novel approach to estimate differences in perceptions of democracy. They asked respondents from Algeria, Jordan, Lebanon, Palestine, Yemen, Egypt, and Tunisia for their understanding of democracy. The response options were "free elections", "freedom of speech", "low economic inequality", and "basic necessities for all". Their results reveal that none of the potential outcomes are mentioned by more than thirty percent of respondents, which gives a strong indication that these populations hold diverse understandings of democracy. Likewise, Ariely and Davidov (2011), using WVS data and confirmatory factor analysis, question that concepts such as "democracy-autocracy preference" (DAP) and "democratic-performance evaluation" (DPE) are comparable cross-nationally. For the DAP they find that although the understanding of the items might be similar, comparing means is problematic. At the same time they find that the DPE means are comparable across a large set of countries. Meanwhile, Behr et al. (2014) assessed ISSP data to demonstrate that the "civil disobedience" item as part of the "rights in a democracy" is understood different in the United States and Canada in contrast to European countries Denmark, Germany, Hungary, and Spain. Using the Latin Barometer and survey data from Romania, Canache et al. (2001) showed that the measurement of the well-known satisfaction with democracy (SWD) concept is not reliable cross-nationally (Canache et al. 2001; Linde and Ekman 2003).

Opposed to this more strict and technical approach, a counter movement led by Welzel and Inglehart (2016) argues that measurement invariance tests have fetishized a construct's internal vildity without regards to potential external validation. Welzel and Inglehart (2016) indicate that amid careful theoretical considerations democratic values might nevertheless be comparable cross-culturally. And indeed, Ariely and Davidov in a later study (2012) reported that the "attitudes towards government intervention" scale of the ISSP is comparable between the United States, Britain, West Germany, and Sweden. Other studies tried to find methodological solutions to the challenge of measurement equivalence in democracy research. Schedler and Sarsfield (Schedler and Sarsfield 2007) proposed the use of cluster analysis to study different conceptions of democracy. Using the Mexican 2003 National Survey on Political Culture, they showed that although there is general support for democracy, people can be divided in different groups reflecting deeper understandings

of democracy. In a recent study, Ulbricht (2018), while showing that the understanding of democracy indeed varies around the world and that support for representative democracy has been substantially overestimated in previous research, maps out an innovative analytical hierarchy process that allowed him to assess different conceptions of democracy (Ulbricht 2018).

In light of the inconclusive ongoing debate and the contradictory findings concerning democratic values' comparability, we argue that for research on delicate topics, such as refugee's value conceptions, testing for measurement invariance needs to be a precondition. Otherwise, the debate on the contestation of Germany's social cohesion and democratic condition as a receiving country is prone to be misguided by faulty data. Based on previous research on value conceptions, we identify two aspects that are likely to hamper comparability: political culture and language.

2.2 Democracy: A Cross-Culturally Ambiguous Concept

Bueno (2012) as well as Ariely and Davidov (2011) argue that the absence of comparability between different countries is the result of different political cultures (Ariely and Davidov 2011; Bueno 2012). If people have different experiences with democracy from one country to another, their perceptions of democracy must be diverse as well. Understanding that conceptions of democracy are strongly shaped by the cultural and historical context points to the "paradox of democracy" (Alvarez and Welzel 2014), the idea that support for democracy in a given country does not reflect the actual democratic state of that country. Support for democracy is argued to be linked primarily to the cognitive understanding of democracy and knowledge of institutional functioning (Miller et al. 1997). The relation between people's awareness of and support for democracy is, however, not linear but instead mediated and influenced by individual biographical experiences with democracy (Cho 2014). These experiences, and in turn also the understanding of democracy, are determined first and foremost by the cultural context and civic educational system in the country of origin (Finkel and Smith 2011). This gives rise to our first hypothesis:

Hypothesis 1 Conceptions of democracy are not comparable between refugees from different countries and the local German population.

2.3 Linguistic Challenges in Cross-Cultural Democracy Research

Linguistic and cross-cultural research shows that concepts which are referred to by the same name can still vary between languages, cultures, and states (Behr et al. 2018). Thus, in addition to the aforementioned difficulties in comparing perceptions of democracy between countries, there is a second dimension challenging comparability. As most of the articles cited above use multi-lingual survey data, the aspect of questionnaire language becomes a crucial one. Translating questionnaires entails a serious risk of bias: conveying a specific meaning from one language to another is not always straightforward (Finkel and Smith 2011) and can trigger a change in attitudes (Zavala-Rojas 2018). Some languages, for instance, have various words for a given concept, whereas others have only one. Words for democracy have entered some languages (e.g., in Africa) only very recently (Bratton and Mattes 2001). Furthermore, a given language can have different dialects, and people who speak the same language often use different expressions in their various dialects. A prominent example is Arabic. Although standard Arabic exists as a language, most people

speak regional dialects. Thus, the formal or official language does not necessarily represent a respondent's mother tongue (Comanaru and d'Ardenne 2018). If this causes respondents to understand questions differently, the measurement would no longer be comparable. Some concepts or terms also have different meanings in a given language, or in some cases translations do not exist and a term can only be described instead of being translated. The language itself incorporates the meaning of a term. Thus, this meaning can vary between languages and impede comparability (Davidov and Beuckelaer 2010). This leads to our second hypothesis:

Hypothesis 2 Percepetions of democracy are not comparable across languages.

In sum, we built the theoretical argument that different political culture and language hamper the comparability of perceptions of democracy across culturally distinct samples. We therefore assume that respondents' experience with democracy and respondents' language pose a challenge to measurement invariance. These challenges are especially important when analyzing and comparing refugees, who, rather than constituting a homogenous group, are characterized by immense cultural and linguistic diversity and a variety of backgrounds (Dustmann et al. 2017).

3 Methods

In order to test whether the conceptions of democracy are comparable between different nationalities and languages, we test for measurement invariance and conduct multi-group confirmatory factor analysis (MGCFA) with a bottom-up stepwise procedure. This is a commonly accepted method (Medina et al. 2009; Saris et al., 2018; Vandenberg and Lance 2000). Measurement invariance assures that mean differences in latent variables between groups are not due to different factor loadings or intercepts and thus meaningful comparisons can be carried out.

Generally speaking, the relationship between democratic values (ξ) and the manifest variables of conceptions of democracy as responses (y) can be described as a function of:

$$y = \tau + \lambda\xi + \delta \tag{1}$$

In this case the intercepts τ and slopes λ are assumed to be equal across people with e.g. different nationalities. In order to test whether the concept of democracy (DEM) is actually comparable, the equation needs to be estimated separately for each manifest variable that measures democratic values (G = {1,2,…,k}) by means of:

$$y_1 = \tau_1 + \lambda_{11} DEM + \delta_1 \tag{2}$$

$$y_2 = \tau_2 + \lambda_{21} DEM + \delta_2 \tag{3}$$

$$y_k = \tau_k + \lambda_{k1} DEM + \delta_k \tag{4}$$

Further, we assume that:

$$Covariance\left(DEM, \delta_i\right) = 0, \quad for\ all\ i \tag{5}$$

$$Covariance\left(\delta_i, \delta_j\right) = 0, \quad for\ all\ i \neq j \tag{6}$$

We handle missing data by employing full information maximum likelihood estimation (Schafer and Graham 2002).

In a first step, we assess whether the latent construct exists in all sub-groups separately but with similar configuration (configural invariance). In order to do so, the factor loadings need to be adequate in all groups. Additionally, the fit indices should not indicate a bad model fit (CFA $\geq$ 0.95; RMSEA $\leq$ 0.05).[1] In the next two steps, we restrict the confirmatory model increasingly and test for metric and scalar invariance. At first, we restrict the factor loadings to be equal across groups (metric invariance), and second, we also restrict the intercepts to be equal across groups (scalar invariance). Between those two steps, the fit indices need to be assessed. The restrictions are commonly confirmed as adequate using the comparative fit index (CFI). However, if the comparative fit index (CFI) is substantially lower than 0.95 or drops by more than 0.01, the procedure needs to be stopped (Chen 2007; Cheung and Rensvold 2002).[2] In this case, the literature proposes testing the initial step again, but instead of restricting all parameters for all variables, estimating parameters for one factor freely (the variable should be determined by considering modification indices, not displayed as tables). If the assumption then holds, we might speak of partial measurement invariance. How many parameters can be estimated freely is the subject of an intense debate in the literature dealing with measurement invariance. Steenkamp and Baumgartner (1998) summarize the debate and argue that due to the nature of a confirmatory model, estimating parameters freely should be treated with caution in order to avoid applying excessive researcher's degrees of freedom. However, they also indicate that under some circumstances, restricting parameters for two variables only can be sufficient. If models rely on only few groups with quite different sample sizes and an overall only medium total sample—like in our case (compared to other studies applying CFA, e.g., Ariely and Davidov 2011; Alemán and Woods 2016)—literature warrants a more cautious approach (McNeish et al. 2017). Moreover, Chen (2007) indicates that smaller samples have a higher chance of producing acceptable confirmatory models. This should be kept in mind when examining the fit indices. We therefore choose a conservative strategy and argue that at least half of the parameters should be fixed in order to make sure that the latent constructs are robust to differences in slopes and intercepts between groups, while also discussing how more liberal cut-off criteria would influence the results in our limitations section.

The models for the different manifest variables (1,2,…,k) are tested with the same restrictions simultaneously for all groups using the lavaan package implemented in R (Jöreskog 1971; Muthen and Satorra, 1995; Rosseel, 2019).

[1] There is not a commonly defined cut-off criterion. We assume that a factor loading is inadequate when, compared to other items, its variance is explained to a lesser degree by the latent variable. We additionally rely on the fit indices in the event that some factor loadings appeared to be substantially smaller than others.

[2] In regard to the stepwise procedure there is no clear cut-off criterion defined in the literature. Most studies however use a CFI between 0.90 and 0.95 (Hu and Bentler 1999; Marsh 2004). Therefore, in order to determine invariance, the deterioration and the absolute CFI have to be taken into account, equally. Moreover, simulation studies suggest, that models based on medium sized factor loadings should be treated more strictly (McNeish et al. 2017). As we will present further down, many of our factor loadings are around 0.5, 0.6 and some even around 0.4, suggesting the application of strict and conservative thresholds.

Table 1 Documentation of manifest variables

Manifest variable	Dataset
Do you think that the following things are what should happen in a democracy or not?	
The government taxes the rich and supports the poor	IAB-BAMF-SOEP/WVS
Religious leaders ultimately determine the interpretation of laws	IAB-BAMF-SOEP
The people choose their government in free elections	IAB-BAMF-SOEP/WVS
Civil rights protect the people from government oppression	IAB-BAMF-SOEP/WVS
Minorities are protected	IAB-BAMF-SOEP
Women have the same rights as men	IAB-BAMF-SOEP/WVS

Table 2 Country of origin, number of respondents and source dataset

Country of origin	Source dataset	Observation (%)
Syria	IAB-BAMF-SOEP	2229 (38.8)
Afghanistan	IAB-BAMF-SOEP	573 (10.0)
Iraq	IAB-BAMF-SOEP	594 (10.3)
Eritrea	IAB-BAMF-SOEP	302 (5.3)
Germany	WVS	2046 (35.6)
Total		*5744 (100)*

3.1 Data

In order to have a dataset consisting of a sufficient number of Germans and recent immigrants, we pool two datasets that both employed a set of the same variables regarding conceptions of democracy: the 2016 wave of the IAB-BAMF-SOEP survey of refugees and the 2014 wave of the World Value Survey. The IAB-BAMF-SOEP survey of refugees is a random sample of refugees and asylum seekers who arrived in Germany between 2013 and 2016 (Kühne et al. 2019). The world value survey (WVS) is a global survey on public opinions and covers around 80 percent of the world population. Separate random samples are drawn for each participating country (Inglehart et al. 2014). Both surveys employ the same four questions asking about conceptions of democracy (see Table 1). The only difference is that the WVS relies on a ten-point scale (from 1 "should definitely not happen in a democracy" to 10 "should definitely happen in a democracy") and the SOEP on an 11-point scale (0–10). In order to harmonize the scales, we split the middle category of the IAB-BAMF-SOEP scale randomly between the neighboring steps. However, as different response scales (even if harmonized) mitigate the power of measurement invariance tests, the assessment of fit indices should be more lenient when comparing refugees (IAB-BAMF-SOEP Survey) and Germans (WVS) (see also limitations section for a discussion). Measurement invariance tests between refugees only and between language groups are not affected by this, as for this part of the analysis we rely on the IAB-BAMF-SOEP data only.

In order to test whether conceptions of democracy are comparable between refugees and the German population, we use the four largest national groups in the IAB-BAMF-SOEP Survey of Refugees. Excluding all other countries from the refugee sample is necessary because they are not represented by a sufficient number of respondents. German respondents are identified in the WVS and integrated to the refugee survey (see Table 2).

Table 3 Mother tongue, survey language, number of respondents

Mother tongue	Observation (%)	Survey language (%)	Observation
Albanian	121 (3.2)	English	728 (16.5)
Arabic	2062 (53.8)	Arabic	2952 (67.0)
Dari/Farsi	533 (13.9)	Farsi/Dari	571 (13.0)
Kurmanji	779 (20.3)	Kurmanji	152 (3.5)
Pashto	70 (1.8)		
Somali	54 (1.4)		
Tigrinya	217 (5.7)		
Total	*3836 (100)*		*4403 (100)*

Table 4 Confirmatory factor analyses without groups—refugees only

	Model 1	Model 2	Model 3
Variable	Loading (SE)		
The government taxes the rich and supports the poor	0.40 (0.02)	0.41 (0.02)	–
Religious leaders ultimately determine the interpretation of laws	0.21 (0.02)	–	–
The people choose their government in free elections	0.63 (0.02)	0.63 (0.02)	0.63 (0.02)
Civil rights protect the people from government oppression	0.70 (0.02)	0.71 (0.02)	0.70 (0.02)
Minorities are protected	0.80 (0.02)	0.79 (0.02)	0.80 (0.02)
Women have the same rights as men	0.64 (0.02)	0.64 (0.02)	0.63 (0.02)
Fit indices	*CFI = 0.95* *RMSEA = 0.08*	*CFI = 0.96* *RMSEA = 0.09*	*CFI = 0.96* *RMSEA = 0.14*

We test for measurement invariance twice: Once for refugees only, and once for refugees and Germans.

To estimate whether perceptions of democracy are comparable across languages, we rely solely on the refugee data because the WVS has only small in-country variance in languages. In the IAB-BAMF-SOEP Survey of Refugees, the target population is, first, multi-linguistic and, second, respondents are offered translated field instruments. For a first model, we group respondents by reported mother tongue (excluding languages used by very small numbers of respondents). In a second model, we use groups defined according to the language chosen by respondents to complete the questionnaire (see Table 3). Respondents could choose between German, English, Farsi/Dari, Pashto, Urdu, Arabic, and Kurmanji (Jacobsen 2018). Due to low usage as a survey language, Pashto and Urdu are omitted from all estimations. Table 3 displays the distribution of mother tongues and the choice of survey language.

4 Results

We started by testing for cross-national measurement invariance within the refugee population. In order to show that the latent construct actually exists in the data, we first conduct a non-grouped confirmatory factor analysis (CFA). Table 4 displays the factor loadings of

Table 5 Fit indices for each step of measurement invariance for country of origin—refugees only

	Full invariance	Partial invariance 1	Partial invariance 2
	Comparative fit index/root mean squared error of approximation		
Configural	0.96/0.09	0.96/0.09	0.96/0.09
Metric	0.94/0.09	0.94/0.10	0.96/0.09
Scalar	0.92/0.09	0.92/0.09	0.95/0.07

Table 6 Confirmatory factor analyses without groups – refugees and germans

	(1)	(2)
Variable	Loading (SE)	
The government taxes the rich and supports the poor	0.42 (0.02)	–
The people choose their government in free elections	0.69 (0.02)	0.73 (0.02)
Civil rights protect the people from government oppression	0.57 (0.02)	0.53 (0.02)
Women have the same rights as men	0.56 (0.02)	0.55 (0.02)
Fit Indices	*CFI = 0.98* *RMSEA = 0.07*	*CFI = /* *RMSEA = /*

a CFA within the IAB-BAMF-SOEP Survey of Refugees. As indicated by their low loadings, for two manifest variables, it is at least questionable whether they are explained by the latent construct: "The government taxes the rich and supports the poor" and "religious leaders ultimately determine the interpretation of laws". Both factor loadings are substantially lower than the others.

Therefore, in a second and third model, both variables are excluded stepwise. The fit indices indicate that model 2 has the best model fit. Therefore, the remaining five variables of the second model will be the basis for further tests of measurement invariance.

Table 5 displays the fit indices for the stepwise procedure. They indicate that configural invariance (CFI = 0.96) is given, whereas metric and scalar invariance are not because the CFI drops substantially (by more than 0.01). This conclusion is supported by the size of RMSEA. Individual factor loadings for each group are displayed in Table 10 in the appendix. Looking at the factor loadings for all countries of origin separately reveals that for the Afghan and Syrian population, the item "the government taxes the rich and supports the poor" has a substantially lower factor loading than for other countries, indicating that these populations have a different understanding of this aspect of what constitutes a democracy.

Testing for partial metric and partial scalar invariance by setting parameters for one variable free (determined by the modification indcies and the expected parameter change; "civil rights protect the people from government oppression") did not improve the model. When setting additional parameters for one more item free ("the people elect the government in free elections"), the fit indices show improved model fit and less deterioration of the CFI. This might point to the conclusion that means across groups could be compared meaningfully if setting parameters for all but three items free is considered adequate. However, as we observe an absence of strict measurement invariance, we conclude that between-group comparisons regarding country of origin of refugees is likely to be problematic.

Additionally, we test whether conceptions of democracy are comparable between refugees and the German population. As displayed in Table 1 we rely on slightly different variables, because not all variables are measured in both data sources. Again, in a first step we test whether the latent construct actually exists in the data. As Table 6 indicates, the fit

Table 7 Fit indices for each step of measurement invariance for country of origin—refugees and germans

	Full invariance	Partial invariance 1	Partial invariance 2
	Comparative fit index/root mean squared error of approximation		
Configural Invariance	0.99/0.05	0.99/0.05	0.99/0.05
Metric Invariance	0.96/0.07	0.97/0.06	0.99/0.05
Scalar Invariance	0.80/0.12	0.80/0.13	0.81/0.13

indices reflect reasonable model fit. Excluding the variable with the lowest factor loading as a robustness check would lead to a just identified model, thus fit indices are not estimable. We therefore decide to proceed with model 1.

The use of multi-group confirmatory factor analysis (MGCFA) reveals that for some countries, the factor loadings are quite small, indicating that configural invariance might be difficult to achieve (see Table 11 in the appendix). However, the fit indices indicate good model fit. Therefore, we proceed with the measurement invariance test. Table 7 indicates that configural as well as metric invariance are achieved (even though the CFI drops by more than 0.01 it is still relatively high in absolute terms)—but not scalar invariance. In order to test for partial scalar invariance, we do not restrict the parameters for "the government taxes the rich and supports the poor". Nevertheless, the CFI does not improve substantially, and the deterioration of the CFI remains the same. Setting an additional parameter free ("the people choose their government in free elections") does not change these results and the CFI remains substantially below any recommend size discussed in the data section. Additionally, the RMSEA supports the conclusion that mean comparisons are problematic.

Both findings together—that within national groups of refugees only partial measurement invariance is achieved, and between national groups of refugees and Germans only metric invariance—leads us to the following conclusion: Conceptions of democracy are most likely not comparable between refugees from different countries or between refugees and the German population. Thus, we can consider Hypothesis 1 to be confirmed.

4.1 Cross-Linguistic Comparisons of Conceptions of Democracy

In a second step, we replicate the previous analyses. However, instead of grouping over country of origin, we use language groups. As a robustness check, we use two different strategies. First, we test for measurement invariance between mother tongues, and second, we group by the language used in the survey.

Testing the second hypothesis regarding comparability between different languages reveals a similar picture. Again, in a first step, we estimate whether the latent construct exists in all groups separately. Again, for some groups, single factor loadings are somewhat too small (see Table 12 in the appendix). However, the CFI for the mother tongue indicates that the latent construct exists in all groups.

When testing for metric invariance, however, we find neither full nor partial metric invariance between different mother tongues, as shown in Table 8 (for partial invariance 1, we set parameters free for the item "minorities are protected"). The same can be seen in Table 9 for the survey language (for partial invariance 1, we set parameters free for the item "women have the same rights as men"; factor loadings within groups are displayed in Table 13 in the appendix). Thus, all models for languages (mother tongue and survey language) indicate that strict measurement invariance is not given. Setting additional parameters free ("civil

Table 8 Fit indices for each step of measurement invariance for mother tongue—refugees only

	Full invariance	Partial invariance 1	Partial invariance 2
	Comparative fit index/root mean squared error of approximation		
Configural Invariance	0.96/0.09	0.96/0.09	0.96/0.09
Metric Invariance	0.93/0.10	0.93/0.10	0.94/0.10
Scalar Invariance	0.91/0.09	0.92/0.09	0.92/0.09

Table 9 Fit indices for each step of measurement invariance for survey language—refugees only

	Full invariance	Partial invariance 1	Partial invariance 2
	Comparative fit index/root mean squared error of approximation		
Configural Invariance	0.96/0.09	0.96/0.09	0.96/0.09
Metric Invariance	0.94/0.09	0.94/0.10	0.95/0.09
Scalar Invariance	0.92/0.09	0.93/0.09	0.95/0.08

rights protect the people from government oppression" for mother tongue; "minorities are protected" for survey language) in each model does not change the conclusion for the mother tongue, although the CFIs slightly improve—the model regarding survey language in Table 9 indicates that mean comparisons based on partial measurement invariance could be valid if setting almost half of all parameters free is satisfactory. However, the overall picture regarding language does not indicate robustness regarding group mean comparisons as strict invariance is not given and the deterioration of the CFI for partial scalar invariance based on the mother tongue is too large. Thus, we accept Hypothesis 2, which states that perceptions of democracy are most likely not comparable across languages.

5 Discussion

5.1 Can We Compare Perceptions of Democracy in Cross-Linguistic and Cross-National Research?

In this paper, we examine whether conceptions of democracy are comparable cross-culturally and cross-linguistically in a nationally and culturally diverse sample of refugees and asylum seekers in Germany. Adding to previous research based primarily on between-country comparisons, we show that conceptions of democracy are also problematic to compare cross-culturally or cross-linguistically within the same societal context and within the same survey. The instruments at hand do not allow for strictly reliable conclusions concerning respondents' democratic values or their underlying conceptions of democracy.

Our results support previous research showing that the democracy scales in the WVS are not adequate to compare conceptions of democracy in cross-national and cross-cultural samples (e.g. Alemán and Woods 2016). Furthermore, we provide new insights showing that such conceptions can also be problematic across mother tongues or survey languages. As previous research suggests, the main reason for the incomparability between different countries of origin are the different political cultures in which respondents are brought up and socialized, which engender different concepts of democracy. Thus, the very reason why the comparison

of conceptions of democracy draws so much attention is the same reason why such comparisons should be treated with caution. Caution is warranted particularly when studies fail to test for measurement invariance and assume that perceptions of democracy are in fact comparable across groups of interest. If there is no uniform conception of democracy, as is the case among refugees and Germans, it is equally misguided to treat democratic values as universally valid constructs. Our research suggests that the longstanding measure of democratic values in the WVS is not suited to measuring these values. Hence, one reason for the notorious problems comparative values studies have faced in explaining the divergent effects of economic development on the liberalization of values between Western and non-Western countries could lie in the fact that the WVS items measure only the Western liberal understanding of democracy, treating it as the only possible option.

Regarding the likely non-comparability of conceptions of democracy across different languages, we argue that languages evolve historically and that translations therefore sometimes incorporate different meanings. Additionally, language is embedded within a cultural frame, which is connected with different images and connotations for the same concept. Therefore, it is likely that questions are interpreted differently across languages (Bond and Yang 1982; Davidov and Beuckelaer 2010; Luna et al. 2008). This conclusion is further supported by the finding that mean comparisons are especially problematic between mother tongues and to a lesser degree between survey languages—as some respondents have to rely on their second language to answer the questionnaire.

However, we would not argue that conceptions of democracy are not comparable per se. First, we see that although almost all of the tests suggest an absence of strict measurement invariance, the CFI in many cases is just below the threshold or even partial measurement invariance is achieved. This is a hint that the items might only be mildly problematic. We assume that the non-comparability in our data is due to the fact that the questions used in the WVS survey (which have served as a model for many other studies such as the employed refugee study in this paper) seem to reflect a Western understanding of liberal democracy. Therefore, we wonder whether our findings would hold when replicating this study with another, broader, definition of democracy (see e.g. Gabriel 2020 who presents a different approach to estimate the understanding of democracy using WVS data). Moreover, as demonstrated by previous international research, support for democracy does not necessarily reflect the democratic state of a country. Thus, support for some items may be driven by a desire for an abstract idea of democracy rather than a critical understanding of the concept of democracy, a relationship that has not been sufficiently researched.

5.2 Outlook: Invalid Comparisons Have Political Implications

Following our findings, we conclude that many of the concerns that were raised during the ongoing heated and emotional public debate over the comparability of cultural values between refugees and Germans were hasty. Wrong or premature conclusions on refugees and migrants in turn have real political implications. Following this, we find a number of important implications for future research amongst refugees and migrants and for comparative values studies. As scholars before us have already emphasized, when comparing different latent constructs, these constructs need to be tested for measurement invariance in order to show that a comparison is valid. Depending on the research question at hand, invariance should be estimated for different groups.

Regarding survey quality, we would argue that tests of measurement invariance may be a useful tool when conducting pretests for surveys. Our example shows that the manifest variables from the WVS do not represent the same latent construct in the refugee data.

Thorough pretests can avoid such misspecifications and might lead to the development of new and more appropriate items for comparative values research.

Additionally, papers addressing political values as a marker of integration should ensure that they base their analyses on a latent construct that actually reflects the intended subject equally in all groups under investigation. Thus, besides the implications of our study for the quality of such analyses, there is also a political dimension: If the latent constructs are not comparable and scholars find substantial differences in value conceptions (e.g., some foreign nationals show lower support for democratic values), this can create a negative narrative based on flawed analyses. Existing democratic attitude items should therefore not be used as a sole basis for conclusions about whether the consensus over democratic values in Germany is in jeopardy. Caution is imperative when talking about value consensus, national values, or presumed disruptions in these values due to migration when considering that previous studies found that value consensus is no defining feature of democracies per se (Schwartz and Sagie 2000). Due to the strong political and societal implications, we propose that measurement invariance tests dealing with such delicate topics should be strict and upfront in their evaluation criteria in order to impede a normatively biased interpretation of the models.

5.3 Limitations

Some limitations of this study should be noted. First, it is unfortunate that we had to rely on two datasets to compare refugees and the German population. This might introduce some error resulting from different modes of data collection, the different institutes conducting the fieldwork, and different incentive strategies. Additionally, fieldwork for the WVS took place two years prior to the IAB-BAMF-SOEP survey of refugees. However, we assume that this does not bias our analyses because it is unlikely that such an important concept as democratic values changes substantially within three years. Furthermore, the cross-country comparison between refugees and Germans relies on a harmonization of scales. While the harmonization procedure was straightforward (randomly splitting the middle category in the refugee survey), it still would have been better if the answers had been collected using the same scale in the first place. The results of the cross-country comparisons should be viewed in light of this harmonization of scales. Differences in slopes or intercepts could also be due to the different scales, respondents use to express their opinion. When scales are not strictly equal, differences in slopes and intercepts can be an expression of this rather than expressing differences in the latent construct. Applying more lenient criteria could therefore be justified. However, the CFI for scalar invariance is substantially lower than 0.90, indicating not acceptable model fit (CFI = 0.80).

Finally, some tests for (partial) scalar invariance were just below the threshold of the CFI (0.95). From a critical stance one might argue that conceptions of democracy are therefore, in contrary to our conclusion, indeed comparable. However, in order to minimize the researcher's degree of freedom, we refrain from altering the way of interpretation aposterioi. Additionally, beyond being below our cut-off criterion, the deterioration of the CFI in those cases is beyond the recommended criteria as well. Moreover, as our results are in line with previous research on this matter (e.g. Alemán and Woods, 2016) and in line with other studies who used the same cut-off criterion (e.g. Hu and Bentler, 1999; Chen, 2007), we are confident that the used threshold does not pose a problem. Yet, we should note that freeing parameters for all variables but two (in spite of the recommendations in the literature) for the refugee only models would lead to partial measurement invariance for between country and between language comparisons (not displayed as a table).

Therefore, we conclude that different interpretations of partial measurement invariance or the use of more liberal cut off criteria may engender other conclusions regarding the comparability of conception of democracy in immigration societies, however, only if such more lenient interpretations are accompanied by very careful theoretical and contextual arguments. On a different note, we suppose that conceptions of democracy might align over time and what was determined incomparable in this article might be comparable in the future when refugees have lived in Germany longer and their German language proficiency has improved.

Although our study faced some obstacles, it clearly provides new insights for comparative value research. We strongly suggest that future cross-cultural, cross-country, and cross-linguistic comparative research on values be carried out with caution, and that it be backed up by an assessment of measurement invariance—even when the target population lives in the same country.

Acknowledgements Open Access funding provided by Projekt DEAL.

Funding Federal Ministry for Education and Research, Germany.

Availability of Data and Material https://doi.org/10.5684/soep.v34 and https://www.worldvaluessurvey.org/wvs.jsp.

Code availability statement Code available upon request.

Compliance with Ethical Standards

Conflicts of interest The authors declare that they have no conflict of interest

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix

See Tables 10, 11, 12, 13

Table 10 Grouped confirmatory factor analysis by country of origin—refugees only

Variable	Loading (SE)			
	Syria	Iraq	Afghanistan	Eritrea
The government taxes the rich and supports the poor	0.28 (0.03)	0.52 (0.04)	0.51 (0.05)	0.39 (0.08)
The people choose their government in free elections	0.74 (0.02)	0.74 (0.04)	0.43 (0.05)	0.66 (0.08)
Civil rights protect the people from government oppression	0.66 (0.02)	0.80 (0.04)	0.62 (0.06)	0.47 (0.08)
Minorities are protected	0.75 (0.02)	0.87 (0.04)	0.77 (0.06)	0.51 (0.08)
Women have the same rights as men	0.52 (0.02)	0.74 (0.04)	0.71 (0.05)	0.76 (0.08)
Fit Indices	*CFI = 0.96* *RMSEA = 0.09*			

Table 11 Grouped confirmatory factor analysis by country of origin—refugees and germans

Variable	Loading (SE)				
	Syria	Iraq	Germany	Afghanistan	Eritrea
The government taxes the rich and supports the poor	0.26 (0.03)	0.50 (0.05)	0.35 (0.03)	0.53 (0.06)	0.26 (0.08)
The people choose their government in free elections	0.86 (0.03)	0.81 (0.04)	0.64 (0.03)	0.61 (0.06)	0.73 (0.08)
Civil rights protect the people from government oppression	0.58 (0.03)	0.78 (0.04)	0.47 (0.03)	0.56 (0.06)	0.36 (0.08)
Women have the same rights as men	0.46 (0.03)	0.71 (0.04)	0.57 (0.03)	0.58 (0.06)	0.92 (0.09)
Fit Indices	*CFI = 0.99* *RMSEA = 0.05*				

Table 12 Grouped confirmatory factor analysis by mother tongue—refugees only

Variable	Loading (SE)						
	Albanian	Arabic	Pashto	Tigrinya	Dari/Farsi	Kurmanji	Somali
The government taxes the rich and supports the poor	0.20 (0.11)	0.32 (0.02)	0.30 (0.14)	0.51 (0.10)	0.49 (0.05)	0.38 (0.04)	0.48 (0.17)
The people choose their government in free elections	0.70 (0.10)	0.81 (0.02)	0.39 (0.14)	0.49 (0.11)	0.45 (0.06)	0.58 (0.04)	0.69 (0.15)
Civil rights protect the people from government oppression	0.76 (0.10)	0.73 (0.02)	0.95 (0.13)	0.63 (0.11)	0.57 (0.06)	0.63 (0.04)	0.44 (0.16)
Minorities are protected	0.73 (0.10)	0.77 (0.02)	0.76 (0.16)	0.63 (0.11)	0.80 (0.06)	0.79 (0.05)	0.65 (0.16)
Women have the same rights as men	0.69 (0.10)	0.56 (0.02)	0.54 (0.14)	0.41 (0.10)	0.74 (0.05)	0.61 (0.04)	1.00 (0.12)
Fit Indices	*CFI = 0.96* *RMSEA = 0.09*						

Table 13 Grouped confirmatory factor analysis by survey language—refugees only

Variable	Loading (SE)			
	English	Arabic	Farsi/Dari	Kurmanji
The government taxes the rich and supports the poor	0.37 (0.05)	0.30 (0.02)	0.47 (0.05)	0.64 (0.10)
The people choose their government in free elections	0.65 (0.05)	0.79 (0.02)	0.44 (0.05)	0.36 (0.11)
Civil rights protect the people from government oppression	0.70 (0.05)	0.69 (0.02)	0.59 (0.06)	0.86 (0.10)
Minorities are protected	0.72 (0.05)	0.79 (0.02)	0.79 (0.06)	0.85 (0.10)
Women have the same rights as men	0.60 (0.05)	0.61 (0.02)	0.72 (0.05)	0.43 (0.10)
Fit Indices	*CFI = 0.96* *RMSEA = 0.09*			

References

Alexander, A. C., & Welzel, C. (2011). Islam and patriarchy: how robust is muslim support for patriarchal values? *World Value Research*. https://doi.org/10.2139/ssrn.2390634.

Alemán, J., & Woods, D. (2016). Value orientation from the world values survey: how comparable are they cross-nationally? *Comparative Political Studies, 49*(8), 1039–1067. https://doi.org/10.1177/0010414015600458.

Alvarez, A. M., & Welzel, C. (2014). Enlightening people. In R. J. Dalton & C. Welzel (Eds.), *The civic culture transformed: from allegiant to assertive citizens*. Cambridge: Cambridge University Press.

Ariely, G., & Davidov, E. (2011). Can we rate public support for democracy in a comparable way? Cross-national equivalence of democratic attitudes in the world value survey. *Social Indicators Research, 104*(2), 271–286. https://doi.org/10.1007/s11205-010-9693-5.

Ariely, G., & Davidov, E. (2012). Asessment of measurement equivalence with cross-national and longitudinal surveys in political science. *European Political Science, 11*, 263–377. https://doi.org/10.1057/eps.2011.11.

Banulescu-Bogdan, N., & Benton, M. (2017). In Search of common values amid large-scale immigrant integration pressures. Retrieved from https://www.migrationpolicy.org/research/search-common-values-amid-large-scale-immigrant-integration-pressures

Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2014). Item comparability in cross-national surveys: Results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience. *Quality & Quantity, 48*(1), 127–148.

Behr, D., Brzoska, P., & Schoua-Glusberg, A. (2018). Linguistic and cultura aspects in migrant surveys: introduction and overview. In D. Behr (Ed.), *Surveying the migrant population: consideration of linguistic and cultural issues* (pp. 5–12). Köln: GESIS.

Belot, M. V. K., & Hatton, T. J. (2012). Immigrant selection in the OECD. *Scandinavian Journal of Economics, 114*(4), 1105–1138. https://doi.org/10.1111/j.1467-9442.2012.01721.x.

Bond, M. H., & Yang, K.-S. (1982). ethnic affirmation versus cross-cultural accommodation: the variable impact of questionnaire language on Chinese bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology, 13*(2), 169–185. https://doi.org/10.1177/0022002182013002003.

Bratton, M., & Mattes, R. (2001). Support for democracy in Africa: intrinsic or instrumental? *British Journal of Political Science, 31*(3), 447–474. https://doi.org/10.1017/S0007123401000175.

Brücker, H., Rother, N., Schupp, J., von Gostomski, C. B., Böhm, A., Fendel, T., et al. (2016). Forced migration, arrival in Germany, and first steps toward integration. *DIW Economic Bulletin, 6*(48), 541–556.

Buber-Ennser, I., Kohlenberger, J., Rengs, B., Al Zalak, Z., Goujon, A., Striessnig, E., et al. (2016). Human capital, values, and attitudes of persons seeking refuge in Austria in 2015. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0163481.

Bueno, O. (2012). Perception and conception: shaping human minds. *Biosemiotics, 6*(3), 323–336. https://doi.org/10.1007/s12304-013-9170-z.

Canache, D. (2012). Citizens' conceptualizations of democracy: structural complexity, substantive content, and political significance. *Comparative Political Studies, 45*(9), 1132–1158. https://doi.org/10.1177/0010414011434009.

Canache, D., Mondak, H. J., & Seligson, M. A. (2001). Meaning and measurement in cross-national research on satisfaction with democracy. *Public Opinion Quartely, 65*(4), 506–528. https://doi.org/10.1086/323576.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modelling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Cho, Y. (2014). To know democracy is to love it: a cross-national analysis of democratic understanding and political support for democracy. *Political Research Quarterly, 67*(3), 478–488. https://doi.org/10.1177/1065912914532721.

Comanaru, R., & d'Ardenne, J. (2018). The development of a research programme to translate and test the personal well-being questions in Sylheti and Urdu. In D. Behr (Ed.), *Surveying the migrant population: consideration of linguistic and cultural issues* (pp. 13–28). Köln: GESIS.

Davidov, E., & Beuckelaer, A. D. (2010). How harmful are survey translations? A test with Schwartz's human values instrument. *International Journal of Public Opinion Research, 22*(4), 485–510. https://doi.org/10.1093/ijpor/edq030.

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

Diamond, L., & Linz, J. J. (1989). Introduction: politics, society, and democracy in Latin America. In L. Diamond, J. J. Linz, & S. M. Lipset (Eds.), *Democracy in developing countries: Latin America* (pp. 1–58). London: Adamantine Press Limited.

Diamond, L., & Plattner, M. F. (2015). *Democracy in DECLINE?*. Baltimore: John Hopkins University Press.

Docquier, F., Tansel, A., & Turati, R. (2018). Do emigrants self-select along cultural traits? Evidence from the MENA countries. FERDI working papers (220).

Dustmann, C., Fasani, F., Frattini, T., Minale, L., & Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic Policy, 32*(91), 497–550.

Finkel, S. E., & Smith, A. E. (2011). Civic education, political discussion, and the social transmission of democratic knowledge and values in a new democracy: Kenya 2002. *American Journal of Political Science, 55*(2), 417–435.

Fischer, R., & Schwartz, S. (2011). Whence differences in value priorities?: Individual, cultural, or artifactual sources. *Journal of Cross-Cultural Psychology, 42*(7), 1127–1144. https://doi.org/10.1177/0022022110381429.

Gabriel, O. W. (2020). Einstellungen zur Demokratie. In T. Faas, O. W. Gabriel, & J. Maier (Eds.), *Politikwissenschaftliche Einstellungs- und Verhaltensforschung* (pp. 230–247). Baden-Baden: Nomos.

Goerman, P., Meyers, M., & Trejo, Y. G. (2018). The place of expert review in translation and questionnaire evaluation for hard-to-count populations in national surveys. In D. Behr (Ed.), *Surveying the migrant population: consideration of linguistic and cultural issues* (pp. 29–42). Köln: GESIS.

Grube, J. W., Mayton, D. M., & Ball-Rokeach, S. J. (1994). Inducing change in values, attitudes, and behaviors: belief system theory and the method of value self-confrontation. *Journal of Social Issues, 50*(4), 153–173. https://doi.org/10.1111/j.1540-4560.1994.tb01202.x.

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for fit Indixes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Inglehart, R. (2000). Globalization and postmodern values. *The Washington Quarterly, 23*(1), 215–228.

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review, 65*(1), 19–51. https://doi.org/10.2307/2657288.

Inglehart, R., & Norris, P. (2003). The True Clash of Civilizations. *Forein Policy, 135*, 62–70. https://doi.org/10.2307/3183594.

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: the human development sequence*. Cambridge: Cambridge University Press.

Inglehart, R., & Welzel, C. (2009). How development leads to democracy: What we know about modernization. *Foreign Affairs, 88*(2), 33–48.

Inglehart, R., Haerpfer, C., Moreno A., Welzel, C., Kizilova, K., Diez-Medrano, J., et al. (Eds.). (2014). *World values survey: Round six*. Country-Pooled Datafile Version: http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp. Madrid: JD Systems Institute.

Jacobsen, J. (2018). Language barriers during the fieldwork of the IAB-BAMF-SOEP survey of refugees in Germany. In D. Behr (Ed.), *Surveying the migrant population: consideration of linguistic and cultural issues* (pp. 75–84). Köln: GESIS.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409–426. https://doi.org/10.1007/BF02291366.

Kluckhohn, C. (1951). Values and value-orientations in the theory of action. In T. Parsons, E. A. Shils, E. C. Tolman, G. W. Allport, C. Kluckhohn, H. A. Murry, R. R. Sears, R. D. Sheldon, & S. A. Stouffer (Eds.), *Toward a general theory of action* (pp. 388–433). Cambridge: Harvard University Press.

Kostenko, V. V., Kuzmuchev, P. A., & Ponarin, E. D. (2016). Attitudes towards gender equality and perception of democracy in the Arab world. *Democratization, 23*(5), 862–891. https://doi.org/10.1080/13510347.2015.1039994.

Kühne, S., Jacobsen, J., & Kroh, M. (2019). Sampling in times of highimmigration: the survey process of the IAB-BAMF-SOEP survey of refugees. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2019-00005.

Linde, J., & Ekman, J. (2003). Satisfaction with democracy: a note on a frequently used indicator in comparative politics. *European Journal of Political Research, 42*(3), 391–408. https://doi.org/10.1111/1475-6765.00089.

Linz, J. J., & Stepan, A. (1996). *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore: John Hopkins University Press.

Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One individual, two identities: frame switching among biculturals. *Journal of Consumer Research, 35*(2), 279–293. https://doi.org/10.1086/586914.

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–342. https://doi.org/10.1207/s15328007sem1103_2.

Marini, M. M. (2000). Social values and norms. *Encyclopedia of Sociology, 4*, 2828–2840.

McNeish, D., An, J., & Hancock, G. R. (2017). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment, 100*(1), 43–52. https://doi.org/10.1080/00223891.2017.1281286.

Medina, T. R., Smith, S. N., & Long, J. S. (2009). Measurement models matter: implicit assumptions and cross-national research. *International Journal of Public Opinion Research, 21*(3), 333–361. https://doi.org/10.1093/ijpor/edp037.

Miller, A. H., Hesli, V. L., & Reisinger, W. M. (1997). Conceptions of democracy among mass and elite in post-Soviet societies. *British Journal of Political Science, 27*(2), 157–190. https://doi.org/10.1017/S0007123497000100.

Müller-Hilmer, R., & Gagné, J. (2018). Was verbindent, was trennt die Deutschen?—Werte und konfliktlinien in der deutschen wählerschaft im Jahr 2017. Retrieved from https://www.boeckler.de/pdf/p_fofoe_report_002_2018.pdf

Muthen, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267–316.

Norris, P., & Inglehart, R. (2012). *Sacred and Secular: religion and politics world wide*. Cambridge: Cambridge University Press.

Parsons, T. (1968). *The structure of social action*. New York: Free Press.

Partridge, P. (1971). *Consent and consensus*. New York: Praeger.

Rizzo, H., Abdel-Latif, A.-H., & Meyer, K. (2007). The relationship between gender equality and democracy: a comparison of Arab versus non-Arab Musilm societies. *Sociology, 41*(6), 1151–1170. https://doi.org/10.1177/0038038507082320.

Rosseel, Y. (2019). The lavaan tutorial. Retrieved from: https://lavaan.ugent.be/tutorial/tutorial.pdf

Saris, W. E., Pirralha, A., & Zavala-Rojas, D. (2018). Testing the comparability of different types of social indicators across groups. *Social Indicators Research, 135*(3), 927–939. https://doi.org/10.1007/s11205-016-1506-z.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Schedler, A., & Sarsfield, R. (2007). Democrats with adjectives: linking direct and indirect measures of democratic support. *European Journal of Political Research, 46*(5), 637–659. https://doi.org/10.1111/j.1475-6765.2007.00708.x.

Schwartz, S. H., & Sagie, G. (2000). Value consensus and importance: a cross-national study. *Journal of Cross-Cultural Psychology, 31*(4), 465–497. https://doi.org/10.1177/0022022100031004003.

Shin, D. C. (2007). Democratization: perspectives from global citizenries. In R. J. Dalton & H.-D. Klingemann (Eds.), *The Oxford handbook of political behavior* (pp. 259–282). Oxford: Oxford University Press.

Silver, B. D., & Dowley, K. M. (2000). Measuring political culture in multiethnic societies: reaggregating the world values survey. *Comparative Political Studies, 33*(4), 517–550. https://doi.org/10.1177/0010414000033004004.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). assessing measruement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–107. https://doi.org/10.1086/209528.

SVR. (2019). Andere länder, andere sitten? Welche kulturellen unterschiede flüchtlinge wahrnehmen—und wie sie damit umgehen. Retrieved from Berlin:

Tausch, A. (2016). Muslim immigration continues to divide Europe: A quantitative analysis of European social survey data. *Middle East Review of International Affairs, 20*(2), 14.

Tausch, A., & Heshmati, A. (2003). Islamism and gender relations in the Muslim world as reflected in the recent world value survey data. *Society and Economy, 38*(4), 427–453. https://doi.org/10.1556/204.2016.38.4.1.

Tessler, M., Jamal, A., & Robbins, M. (2012). New findings on Arabs and democracy. *Journal of Democracy, 23*(4), 89–103.

Ulbricht, T. (2018). Perceptions and coneceptions of democracy: applying thick conecepts of democracy to reassess desires for democracy. *Comparative Political Studies, 51*(11), 1387–1440. https://doi.org/10.1177/0010414018758751.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestion, practices, and recommandations for organizational research. *Origanizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002.

Vlas, N., & Gherghina, S. (2012). Where does religion meet democracy? A comparative analysis of attitudes in Europe. *International Political Science Review, 33*(3), 336–351. https://doi.org/10.1177/0192512111415983.

Wan, C., Chiu, C.-Y., Tam, K.-P., Lee, S.-L., & Lau, I. Y.-M. (2007). Perceived cultural importance and actual self-importance of values in cultural identification. *Journal of Personality and Social Psychology, 92*(2), 337–354. https://doi.org/10.1037/0022-3514.92.2.337.

Welzel, C. (2013). *Freedom rising*. Cambridge: Cambridge University Press.

Welzel, C., & Inglehart, R. F. (2016). Misconceptions of measurement equivalence: Time for a paradigm shift. *Comparative Political Studies, 49*(8), 1068–1094.

Wimmer, A., & Soehl, T. (2014). Blocked acculturation: cultural heterodoxy among europe's immigrants. *American Journal of Sociology, 120*(1), 146–186. https://doi.org/10.1086/677207.

Zavala-Rojas, D. (2018). exploring language effects in crosscultural survey research: Does the language of administration affect answers about politics? *Methods, Data, Analyses, 12*(1), 127–150. https://doi.org/10.12758/mda.2017.16.

Zavala-Rojas, D., & Saris, W. E. (2017). Measurement invariance in multilingual survey research: the role of the language of the questionnaire. *Social Indicators Research, 140*, 2485–2510. https://doi.org/10.1007/s11205-017-1787-x.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.