Open access • Journal Article • DOI:10.1007/S11205-010-9693-5

Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey — Source link

Gal Ariely, Eldad Davidov

Institutions: University of Haifa, University of Zurich

Published on: 01 Nov 2011 - Social Indicators Research (Springer Netherlands)

Topics: Equivalence (measure theory) and World Values Survey

Can we rate public support for democracy in a comparable way? Cross-national equivalence of democratic attitudes in the World Value Survey

Ariely, G; Davidov, E

Can we rate public support for democracy in a comparable way?

Cross-national equivalence of democratic attitudes in the World Value Survey

Dr. Gal Ariely (corresponding author)
University of Haifa
School of Political Sciences,
Haifa, Israel, 31905

Prof. Dr. Eldad Davidov
University of Zürich
Institute of Sociology
Andreasstr. 15 CH-8050 Zürich
Switzerland

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in **Social Indicators Research**, *following peer review.*

The final publication is available since September 3rd 2010 at www.springerlink.com, under URL:

http://www.springerlink.com/content/640n0pqq80353403/

or under

DOI: 10.1007/s11205-010-9693-5

Can we rate public support for democracy in a comparable way? Cross-national equivalence of democratic attitudes in the World Value Survey

Abstract

In this study we examine the cross-cultural equivalence of two scales that measure attitudes toward democracy across 36 countries in the World Value Survey (WVS) 2000. We examine the equivalence of these scales in order to explore if we can meaningfully compare democratic attitudes across countries. Multiple group confirmatory factor analyses (MGCFA) is applied to answer this question. The analyses indicate that the scales may be compared but only to a certain extent and not across all the countries. We close this article by discussing the implications of the findings.

Key words: attitudes toward democracy; Measurement invariance; multiple-group confirmatory factor analysis (MGCFA); World Value Survey (WVS).

1. Introduction

Scholars have long assumed that a democratic system's stability depends upon its legitimacy – and therefore also upon the extent to which the public subscribes to democratic attitudes (Diamond 1999). Due to the importance of these attitudes in the legitimization of democratic regimes, it is not surprising that, for over five decades, a substantial portion of the empirical literature has been devoted to measuring mass public attitudes toward democracy in different countries. The proliferation of cross-national research surveys like the World Value Survey (WVS) and various regional barometers has extended scholarly abilities to explore democratic attitudes (Heath, Fisher and Smith 2005; Kittilson 2007; Shin 2007). Comparative survey research enables scholars to test cross-national variation in attitudes toward democracy and the extent to which such differences may be explained. These studies have expanded our understanding of democratic values and democratization alike (e.g., Norris 1999). Beyond that, the comparison of mass support for democratic values across various cultures enables scholars to examine important questions like the relationship between democratic attitudes and cultural values or religious identities (e.g., Dalton and Ong 2005; Inglehart and Norris 2003).

Paradoxically, the very differences in culture that give such cross-national studies their value also threaten the achievement of equivalence of the scales that are used to measure the concept of attitudes toward democracy across different countries (Smith 2003). Assuming equivalence of scales designed to measure concepts across countries may be misleading, since differences between groups only reflect *true* differences if the measures are equivalent (Billiet 2003). Measurement invariance is conceptually defined as "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute"

(Horn and McArdle 1992: 117; Wu, Li and Zumbo 2007). Differences in scales means or in relationships (regression coefficients, covariances) between scales and other theoretical constructs of interest may derive from systematic biases of responses across countries or from variant understandings of the question items rather than from 'true' differences across groups (Vandenberg and Lance 2000; Van de Vijver 2003). Although current cross-national surveys take great care in item selection, translation, and other procedures in order to increase the probability of comparability of concepts across countries (Jowell et al. 2007), these procedures cannot guarantee invariance, which requires statistical testing. Guaranteeing that the measurement of relevant constructs is invariant consequently constitutes a central concern when applying theories and instruments across different contexts of measurements like countries (Billiet 2003; Cheung and Rensvold 2002; Harkness, Van de Vijver and Mohler 2003).

Despite the extensive examination of democratic attitudes in comparative survey research, their equivalence has seldom been investigated. In this paper we examine cross-country equivalence of two operationalizations for attitudes toward democracy that have been used in many studies employing the WVS. These operationalizations measure the Democracy-Autocracy Preference (DAP) scale and the Democratic Performance Evolution scale (DPE). We examine the equivalence of these two scales in order to find out if attitudes toward democracy (as measured in the WVS) can be meaningfully compared across all the countries in the WVS 2000. Thus, we will answer the question whether it is meaningful to compare the means of these scales and their correlates across countries participating in the WVS

2. Challenges of the comparability of attitudes toward democracy in cross-national surveys

Various cross-national surveys like the New Democracies Barometer, the LatinoBarometer, the AfroBarometer and the World Values Survey include questions intended to provide researchers with a comprehensive measurement of democracy as perceived by the public. The challenging task of constructing a valid assessment of attitudes toward democracy has led to the introduction of several scales such as the "democracy as an ideal form of government" scale (Klingemann 1999)[1], or the "realistic measures of democracy" scale (Mishler and Rose 2001) [2], just to name two. The importance of measuring attitudes toward democracy has created an extensive discussion in the literature concerning the measurement of such attitudes and their employment in various studies (e.g., Canache, Mondak and Seligson 2001; Linde and Ekman 2003; Mishler and Rose 2001; Schedler and Sarsfield 2007). However, discussion in the literature regarding the operationalization of attitudes toward democracy has not paid sufficient attention to the challenges of cross-cultural comparisons. It seems that, as Inglehart (2003) argued, there is a tendency to replicate the use of items that are considered well designed and effective. However, these items rest mainly on their face validity and have not been subject to any test for their comparability across different contexts.

The comparability of cross-national surveys is challenged by various methodological problems like translation and differences in survey practice that affect the sampling and coverage (Curtice 2007; Heath et al. 2005). This challenge increases as one expands the number of countries that are included in the survey. To date, the

[1] The two scale items include rating of the importance of having a democratic system and whether it is better than any other form of government (WVS 1995-97).
[2] The scale was based on a respondent's quality assessment (good vs. bad) of the political system as it was in [reference to previous regime], as it is today, and as it is expected to be tomorrow? (WVS 1995-97)

WVS is the only academic (nearly) global public opinion survey that covers over 80% of the world population (Norris 2009). While the WVS collects data in face-to-face interviews using standardized sampling methods, standardization is still limited (for review see Curtice 2007; Heath et al. 2005). This limitation underlines the importance of testing for invariance of the scales used across countries.

The fact that this issue has been overlooked is evident in the literature that discusses measurements of attitudes toward democracy. While there is continuous debate about the contents of democratic attitudes and the ways they should be measured, the impact of cross-national variations such as differential scale use, translation, or dissimilar understandings of the questions is not addressed. However, such differences may have substantial consequences that will result in biased estimated of means and regression coefficients. For example, several studies have criticized the commonly used "satisfaction with democracy" single item measure for its vagueness and misinterpretations. It has been argued that it is not clear whether this scale measures support for democratic values or support for the regime (e.g., Canach et al. 2001; Lagos 2003; Linde and Ekman 2003). However, these studies have ignored fundamental methodological limitations of single item scales. The underlying assumption that there is a one to one relationship between the single item and the theoretical construct and that it is measured without error (which is the implicit assumption when a scale is measured by only one item, see, e.g. Brown 2006 or Bollen 1989) is doubtful, especially with a multifaceted concept like democracy. Only multiple indicator scales allow controlling for random and non-random measurement errors and test for the convergent and discriminate validity of the scale, as well as testing for cross-cultural invariance of meaning and scale use (Billiet 2003; Bollen 1989; Brown 2006; Saris and Gallhofer 2007). Put differently, while single

item scales that measure attitudes toward democracy are frequently used, the cross-national comparability of such scales cannot be considered unless additional items are added to the scale. Consequently, we have chosen to analyze in this study two multiple indicator scales that are available at the WVS and have been used in numerous studies.

2.1 **Measuring attitudes toward democracy in the WVS**

The WVS contains various items that capture different aspects of democratic attitudes like the satisfaction with democracy performance in the country, as well as questions inquiring about different characteristics of democracy. Among these items, different rounds of the WVS contain two sets of items that were applied in various studies to measure two scales of democratic attitudes: the "democracy-autocracy preference" (DAP) scale and, what we have labeled the "democratic performance evaluation" (DPE) scale.

The first set of items is based on the question: "I am going to describe various types of political systems and ask what you think about each as a way of governing this country. For each, would you say it is a very good, fairly good, fairly bad, or a very bad way of governing this country?": 1) Having a strong leader who does not have to bother with parliament and elections. 2) Having experts, not governments, make decisions according to what they think is best for the country 3) Having the army rule 4) Having a democratic political system. This scale examines people's preferences of various types of political systems for their countries. It taps support for democratic systems versus support for autocracies, like the rule of a strong leader or of the army. It is based on the assumption that people can discriminate between their current regime and conceivable options. In that sense, this scale grasps the support people give to a democratic regime vis-a-vis autocracy alternatives (Diamond 1999).

As such, these items were used in various studies to create what Inglehart and Welzel (2005) labeled as the DAP scale.

The second set of items is based on the question: "I'm going to read off some things that people sometimes say about a democratic political system. Could you please tell me if you agree strongly, agree, disagree or disagree strongly, after I read each one of them?": 1) In democracy, the economic system runs badly 2) Democracies are indecisive and have too much quibbling 3) Democracies aren't good at maintaining order 4) Democracy may have problems but it's better than any other form of government. Three items reflect respondent's evaluation of democratic performance in different spheres while one item echoes Winston Churchill's notable expression that democracy is "the worst form of government except all those other forms that have been tried from time to time". Therefore, we have labeled these four items as the democratic performance evaluation (DPE) scale.

The items of the DAP scale were included in three waves of the WVS (1995, 2000, 2005) and those of the DPE scale in the 1995 and 2000 WVS waves. Consequently, these two sets of items have been used in scores of studies (e.g. Dalton 1999; Dalton and Ong 2005 Klingemann 1999; Hofmann 2004; Inglehart and Welzel 2005; Guerin, Pétry and Crete 2004; Dixon 2008; Esmer 2002; Haerpfer 2008; Pettersson 2008; Tusicisny 2007; Wang, Dalton and Shin 2006; Welzel 2007; Tessler, Moaddel and Inglehart 2006). It is beyond the scope of this paper to introduce all of these studies. By and large, this research could be divided between studies that focus on civic culture while exploring the relations between democratic attitudes and other spheres of interest, and studies that compare the support for democratic attitudes between countries and cultures.

An example of a study that uses these items to explore civic culture in a comparative way is Klingemann's (1999) work which maps geographically patterns of democratic attitudes as part of his analyses of political support across all the WVS countries. Similar wide ranging analyses have been conducted by Inglehart and Welzel (2005). They have used these items in order to examine which cultural values predict support of democracy. More focused examinations have been conducted across specific countries like Haerpfer's (2008) investigation of changes in public support for alternatives to democracy in the post-Soviet countries and the factors that shape them.

Many studies have used the WVS to examine cross-cultural differences in social attitudes and specifically in democratic attitudes. Dalton and Ong (2005), for instance, used this scale to observe differences in the level of support for democracy in six East Asian countries in comparison to four Western countries. They found that despite a general tendency to support democratic values in all the ten countries they observed, the scale mean was higher in the advanced industrial democracies. Beyond those differences in the scale means, they also observed how supportive democratic attitudes could be explained by non-authoritarian orientations.

Recently, the examination of cross-cultural differences in democratic attitudes has received increasing attention. Samuel Huntington's idea of the clash of civilizations, which received exceptional prominence even beyond academic circles after September 11th, has led scholars to examine differences in democratic attitudes across different cultures, especially between Muslims and non-Muslims (e.g., Esmar 2008; Hofmann 2004; Inglehart and Norris 2003; Rizzo, Abdel-Latif and Meyer 2007; Rowley and Smith 2009; Tessler 2002; Toros 2010). By and large, these studies indicated that significant differences in democratic attitudes between Muslims and

non-Muslims could not be found. Specifically, according to these studies, a systematic comparison of mass attitudes indicated that Huntington's idea of a clash of civilizations was in doubt. As Inglehart and Norris claimed in a Foreign Policy article: "Any claim of a "clash of civilizations" based on fundamentally different political goals held by Western and Muslim societies represents an oversimplification of the evidence. Support for the goal of democracy is surprisingly widespread among Muslim publics, even among those living in authoritarian societies" (2003: 68).

In sum, such studies emphasize the advantages in using cross-national survey data to address important questions regarding cross-cultural differences in democracy and development, or democratic values and cultures. Nevertheless, the statistical equivalence of tools which these studies used in order to draw conclusions was not examined. Without such equivalence, the conclusions of these studies are also doubtful. Given the widespread use of the WVS as a source for cross-national comparison of democratic attitudes, one can presume that these scales will be used in future studies as well. Therefore, to examine the comparability of the DAP and DPE scales, we evaluate measurement characteristics (convergent validity, see e.g. Campbell and Fiske 1959) and equivalence across all WVS 2000 countries.

3 Method

3.1 Data

The WVS collected data regarding various dimensions of attitudes toward democracy, covering countries with widely divergent histories, cultures, and political conditions, from established democracies to non-democratic countries like China and Iran. We have analyzed all the countries that included the scales items as part of the 2000

survey since the WVS 2000 is the most recent survey that contains both scales.[3] Our sample contains 36 countries (with number of respondents in parentheses): USA (1200), Canada (1931), Spain (1209), Serbia (1200), Moldova (1008), Montenegro (1060), Albania (1000), Macedonia (1055), Bosnia (1200), India (2002), Pakistan (2000), Bangladesh (1500), Philippines (1200), Japan (1362), China (1000), South Korea (1200), Vietnam (1000), Indonesia (1004), Mexico (1535), Argentina (1280), Puerto Rico (720), Chile (1200), Peru (1501), Venezuela (1200), Kyrgyz (1043), South Africa (3000), Uganda (1200), Tanzania (1171), Nigeria (2022), Zimbabwe (1002), Turkey (3041), Morocco (2264), Jordan (1223), Iraq (2325), Iran (2532) and Algeria (1282).

3.2 Measurement model

Table 1 summarizes the items we used in order to measure the DAP scale and the DPE scale. Figures 1a and 1b provide a schematic representation of the measurement models used in this study. The large circles represent the hypothesized measured latent constructs, DAP and DPE. The rectangles represent measured items. The small circles (e1 through e4) represent measurement errors.

Table 1 about here

Figures 1a/1b about here

3.3 Analytical strategy

Confirmatory factor analysis (CFA) was used in order to assess the measurement model in each of the 36 countries in the sample, as well as to examine the comparability of the scale across countries.[4] CFA provides estimates of the relations

[3] For further details, check the website of the World Values Survey Association. http://www.worldvaluessurvey.org/

[4] Several studies have used multiple group CFA (MGCFA) to assess measurement invariance of scales in such cross-national surveys as the European Social Survey (ESS) and the International Social Survey Program (ISSP) (e.g., for human values, Davidov 2008; Davidov, Schmidt and Schwartz 2008; for national identity, Davidov 2009; for trust, Reeskens and Hooghe 2008).

between observed indicators and the hypothesized latent construct (factor), and provides fit indices that report whether the hypothesized structure of associations between a latent construct and its proposed indicators fits the data. This information is used to determine whether a hypothesized latent construct underlies a scale. Assuming the existence of a latent variable means that observed variables are only correlated to the extent that they share an underlying concept (Brown 2006). This framework allows researchers to test empirically for measurement invariance across groups when the factor is compared across groups using multiple group confirmatory factor analysis (MGCFA) (for further details, see, e.g., Bollen 1989; for an application, see, e.g., Davidov et al. 2008).

Steenkamp and Baumgartner (1998) and Vandenberg and Lance (2000) offer step-by-step guidelines to facilitate testing for different levels of measurement invariance in cross-national studies using a MGCFA approach. They propose assessing three hierarchical levels of measurement invariance: configural, metric, and scalar. Configural invariance, the lowest level of invariance, requires that the items in the measuring instrument display the same configuration of loadings in each nation or cultural unit. Metric invariance reflects a higher level of invariance and is required to guarantee a similar understanding of the concept. Metric invariance is necessary to allow a meaningful comparison of a construct's correlates (covariances, unstandardized regression coefficients) across countries. For comparing the mean of the construct across countries, a higher level of invariance is required - scalar invariance. Scalar invariance guarantees that cross-national differences in the means of the observed items are the consequence of differences in the means of their corresponding constructs and not due to differences in factor loadings or indicator intercepts (Steenkamp and Baumgartner 1998). In the following, we first present

single-country CFAs of DAP and DPE from the WVS data followed by a discussion of the invariance tests.

4. Results

4.1 DAP scale

4.1.1 *Single-country analyses*

In line with Byrne's (2001) assertion regarding the importance of conducting single-group analyses prior to multiple-group comparisons, we began with 36 separate CFAs for each country. First, we computed 36 Pearson product-moment (unstandardized) covariance matrices, one for each country, as input for estimating the CFAs.[5] For the estimation we employed the Amos 16.0 software package (Arbuckle 1995 - 2007). Table 2 displays the global fit measures, p value of close fit (Pclose), root mean square error of approximation (RMSEA), and comparative fit index (CFI), and model modifications for the DAP scale in the single-country analyses.[6]

- Table 2 around here -

Looking at Table 2 it is evident that in nearly all countries, one modification was needed to achieve a better fit to the data in the model. For several countries, a covariance between the errors of V165 and V167 had to be released, indicating that the two questions (preferring experts and having a democratic political system) are more strongly related to each other than to the other indicators tapping the construct. In other countries, the covariance between the errors of V166 and V167 (having the army rule and having a democratic political system) had to be released. The requirement to allow two indicators in the scale to relate to each other more strongly may reflect variations in the importance of each element in the scale for the

[5] In the current analyses, we firstly ran the models using pairwise deletion to deal with the problem of missing values (Schafer and Graham 2002). As a final test, we have also analyzed the models using the Full Information Maximum Likelihood (FIML) procedure which is preferred when portions larger than 5% of the data are missing (Schafer and Graham 2002).

[6] For cut off criteria for global fit measures, see Hu and Bentler 1999 and Marsh, Hau and Wen 2004.

measurement of democracy preference. Even after these modifications we find that, based on the global fit measure CFI reported in Table 2 and the standardized factor loadings reported in Table 3, the measurement model is not acceptable in many countries[7].

The content of V167 that asked explicitly about democracy is different from the three other items that did not explicitly mention the word "democracy". This may have resulted in a weak standardized factor loading for this indicator in many countries. As Schedler and Sarsfield (2007) have argued, when survey questions explicitly use the term "democracy", interviewees might be influenced by the idealization of democracy, and this in turn might bring about interviewer effects and socially desirable responses. People pay overt lip service to democracy all over the world but this does not necessarily indicate the depth of democratic values (Inglehart, 2003) which may be better tapped by the other three questions of the scale. People might answer that they support democracy and, at the same time, they support a strong leader or expert rule, as well. Thus, V167 was dropped from further analyses. Furthermore, nine countries where other item/s did not load substantially on the DAP concept were also dropped from further analyses of invariance.

Table 3 about here

4.1.2. *Multiple-Group CFAs and Testing for Invariance*

In order to test for invariance of the scale we conducted a MGCFA across 27 countries,[8] inspecting differences in the chi-square and other global fit measures between the models. The results of the invariance tests are summarized in Table 4.

[7] Brown (2006) suggests that standardized factor loadings lower than 0.3-0.4 should indicate that the question is inadequate to measure the latent construct.

[8] Although scholars have argued that MGCFA may be an inappropriate tool for testing for invariance of Likert scales (see, e.g., Lubke and Muthén 2004), studies have demonstrated that it works well even when data are not continuous or normally distributed but ordinal (De Beuckelaer 2005; Welkenhuysen-Gybels and Billiet 2002; Welkenhuysen-Gybels 2004).

Based on the model fit measures (RMSEA = .017, Pclose = 1.00, CFI = .928), we do not reject the metric invariance model (4a). Metric invariance, however, still does not allow for mean comparison of the DAP concept across countries. To compare means meaningfully, a higher level of invariance is necessary, full or partial scalar invariance.

- Table 4 around here -

The second row in Table 4 reports the fit indices of the full scalar invariance model (Model 4b). Based on the fit measures (RMSEA = .045, Pclose = 1.00, CFI = .000), we reject the model. The deterioration in the global fit measures is way beyond the recommended criteria (Chen 2007). However, we can still fall back to partial scalar invariance, when full scalar invariance is rejected. The partial scalar invariance model (4c in Table 4) requires that at least two indicators have equal factor loadings and intercepts across countries (Steenkamp and Baumgartner 1998). So we release the equality constraint on the factor loading and intercept of item V166, as it has displayed the most severe violations of invariance across countries. Examining the model fit measures reveals that in spite of releasing these constraints, the model cannot be accepted by the data (RMSEA = .040, Pclose = 1.00, CFI = .554).

In summary, while the meaning of the constructs as measured by the three indicators may be regarded as similar in these countries as metric invariance could be established, and relationships between DAP and other theoretical constructs of interest may be compared meaningfully across these 27 countries, comparing means is still problematic.

4.2. DPE scale

We will now present the results of the analyses for the DPE scale. In order to avoid repeating the procedure of the analyses, we will present the findings in a shorter

format.[9] We started with single-country CFAs across all countries. In ten countries we have found that V172 (democracy better than other forms of government) is insignificant or negatively loaded on the construct. Like in the DAP scale, mentioning democracy as a form of governance without referring to its characteristics (such as its economic system or order) explicitly taps another dimension than the other questions in the scale. Thus, we have excluded this item and re-analyzed the CFAs in the single countries with the other three indicators (V169 – economic systems in democracies, V170 – democracies are indecisive, and V171 – democracies are not good at maintaining order). We have found that except for Iran, the three items were positively and significantly loaded on the construct across 35 countries and thus, that the model worked well for this set of countries.

We conducted a MGCFA for the three items across all the countries (except Iran). The results of the invariance tests are summarized in Table 5. Full metric invariance was accepted by the data (RMSEA = .010, Pclose = 1.00, CFI = .990). However, based on the global fit measures, full scalar invariance was rejected by the data (CFI = .909, RMSEA = .021, Pclose = 1.00). The deterioration in the global fit measures is beyond the criteria recommended by different authors (Chen, 2007). However, we can still fall back to partial scalar invariance, when full scalar invariance is rejected. The partial scalar model global fit measures (CFI = .983, RMSEA = .013, Pclose = 1.00) reported in the third row of Table 5 are satisfactory. The difference in CFI between the partial scalar and metric invariance models (ΔCFI = .007) lay below the recommended cut-off criteria (Chen, 2007). Therefore, we consequently accepted the partial scalar invariance model.

Table 5 around here

[9] Model outputs, factor loadings and global fit measures may be available from the first author upon request.

To conclude then, the three items DPE is a better scale then the three item DAP for cross cultural comparisons since it can be compared across 35 countries and not only across 27 countries. Probably, that is because of the larger differences in the content of the DAP items in comparison to the DPE items. Furthermore, the DPE scale, was found to demonstrate partial scalar invariance. So, comparing the scales means across all countries in the sample is also possible

5. Conclusion

Citizens' perceptions regarding democracy are considered a key issue in the comparative study of democracy, and this has been measured in numerous cross-cultural surveys (Klingemann 1999; Shin 2007). Despite disagreements regarding the definition of democracy and the ways democratic attitudes are operationalized, the employment of various scales measuring attitudes toward democracy appears likely to continue in comparative research. In view of the increasing use of cross-national surveys in the study of democracy and democratization, the issues of measurement and measurement equivalence are highly relevant. The proliferation of cross-national surveys emphasizes the need to pay more attention to issues of equivalence (Adcock and Collier 2001; Heath et al. 2005; King et al. 2004). After all, in perhaps no other subfield of social science are research issues of methodology and measurement open to challenge and criticism as when they are applied in cross-national settings (Johnson 1998). Consequently, recent cross-national survey projects have put great effort into increasing cross-national invariance by applying high standards of data collection, response rates, and translation procedures. However, the application of such high standards in the data collection procedures is not enough to guarantee that measurement scales are *valid* and *invariant* across countries. It is consequently crucial to establish the equivalence of scales across different contexts statistically.

Our study demonstrates how scales may be subjected to such tests in order to verify that attitudes toward democracy can be meaningfully compared across different contexts of measurement. In the present analysis we used two scales that are part of the WVS to measure attitudes toward democracy. We employed data from the 2000 wave of the WVS and tested the scale's comparability in a cross-national perspective across 36 countries using MGCFA, one of the proposed methods to conduct tests of a scale's comparability (for an overview of several other methods to conduct the test, see De Beuckelaer 2005; Davidov, Schmidt and Billiet in press; for another alternative, see, e.g., Steenbergen 2000).

In order to examine if the scales indicators really measure the same construct across countries, we conducted firstly separate CFAs for each of the countries. The single country analyses revealed that for both scales, the four items did not measure the same construct across countries, i.e., convergent validity was not established in each country and as a result configural invariance could not be established for the full set of countries. In the next step, after excluding from each scale the item that caused model misspecification, we examined if the remaining indicators were valid and invariant across the remaining countries and if it was meaningful to compare the means of this scale across them. Testing for invariance of the DAP scale across the 27 countries where it was valid demonstrated that only full metric invariance could be established. The DPE scale guaranteed partial scalar invariance across 35 countries. Therefore, correlates (unstandardized regression coefficients, covariances) between these scales and other theoretical constructs of interest may be compared across countries. The means of the DPE scale may be compared across 35 countries. However, it may be problematic to compare the means of the DAP scale across countries.

Our results call for deeper attention to the ways democratic attitudes are being measured and compared across different contexts. While cross-national survey projects like the WVS are an important source for analyzing mass attitudes toward democracy, overlooking the issue of measurement problems in general and measurement invariance in particular might cast doubt on theoretical implications that are based on invalid measurement (Knoppen and Saris 2009). For instance, studies that use these items in order to put Huntington's idea of the clash of civilizations to an empirical investigation presume that these survey items are comparable across states, cultural or religious groups. However, our analyses indicate that further investigation of such assumptions is required.

The evidence in this paper does not imply that the current WVS data cannot continue to facilitate cross-national research on public opinion toward democracy; in fact, it is just the contrary. Our study shows that *correlates* of the two scales with other theoretical constructs of interest and means of one of the scales may be compared across most countries meaningfully. In order to remove such doubts from the cross-national study of democratic attitudes we would like to offer two directions for students of democratic attitudes. The first is based on conducting similar analyses of data, as we have demonstrated in this paper. The second is aimed at future development of scales to measure democratic attitudes in cross-national contexts.

With respect to the first proposal, the current study shows that researchers who wish to apply the WVS scales of attitudes toward democracy have to consider them with an awareness of their potential limits. In this study we examined the DAP/DPE scales across all of the WVS 2000 countries. Perhaps researchers who particularly study different subsets of countries (e.g., post-Communist countries) may well find that these particular countries demonstrate higher levels of invariance of the scale

because of their contextual similarities. As we have shown, conducting MGCFA analyses is an efficient way to examine measurement invariance and this procedure can be applied by researchers even during the process of the selection of items which measure their theoretical concepts.

With respect to the second proposal, the fact that some of the items that we have examined failed to achieve the highest level of invariance across all the countries in the sample does not mean that fully comparable public conceptions of democracy do not exist in reality. It only means that the invariance of the items that we have examined is quite limited. Heath and Martin (1997) argue that there is a resource tradeoff between the ability of a survey project to invest in preliminary methodological work in scale construction and its ability to collect large representative samples. While that is a problem in any national survey, it is even a bigger problem in cross-national survey projects that need more resources and are required to handle more challenges, one of which is scale comparability. Nevertheless, such reasons should not prevent students of democratic attitudes to develop better scales that will tap public perceptions across different contexts in a more comparable way.

It is also worth noting that MGCFA is also subject to criticism. First of all, there is no consensuses about the cut-off criteria for the models goodness of fit measures (Marsh et al. 2004) and some scholars offer alternative ways to evaluate models than those we have used here (e.g., Saris, Sartora and Van Der Veld 2009). Furthermore, Steenbergen (2000) argues that CFA in general and MGCFA in particular requires relatively large sample sizes, and proposes instead to use the item similarity index. Alternative strategies of analyses might find that higher levels of invariance can be established. Second, it is important to note that in addition to statistical tests like the

MGCFA, there are various strategies to establish comparability (Johnson 1998). For example, Dalton and his colleagues analyzed open-ended questions about the meaning of democracy that were used in various regional barometers (Dalton, Shin and Jou 2007). Open-ended questions might be perceived as a more rigorous test of democratic attitudes since the respondents are required to define democracy in their own words. However, such tests are not complete without complementary standardized questions and statistical tests of comparability. Furthermore, one could try to find out *why* certain countries demonstrate invariance of the scale and others do not by including contextual variables in the analysis and investigating whether they are responsible for the noninvariance.

Establishing measurement invariance is not a goal in itself. Nonetheless, without establishing invariance, it is more difficult to conduct meaningful comparisons of attitudes toward democracy based on accessible cross-national survey data. The current study demonstrates how equivalence may be examined in comparative survey research using MGCFA.

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review,* 95(3), 529-546.

Arbuckle, J., L. (1995 - 2007). *Amos 16.0 user's guide.* Chicago, IL: SPSS.

Billiet, J. (2003). Cross-cultural equivalence with structural equation modeling. In Harkness, J. A. Van de Vijver, F. J. R., & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 247-264). New York: John Wiley.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. London: Lawrence Erlbaum Associates.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethods matrix. *Psychological Bulletin,* 56, 81-105.

Canache, D., Mondak, J. J., & Seligson, M. A. (2001). Meaning and measurement in cross-national research on satisfaction with democracy. *Public Opinion Quarterly, 65*(4), 506-528.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling,* 14(3), 464-504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling,* 9(2), 233-255.

Curtice, J. (2007). Comparative opinion surveys. In J. R. Dalton, & H. Klingemann (Eds.), *Oxford handbook of political behavior* (pp. 898-909). Oxford: Oxford University Press.

Dalton, R. J. (1999). Political support in advanced industrial democracies. In P. Norris (Ed.), *Critical citizens: Global support for democratic governance* (pp. 57-77). Oxford: Oxford University Press.

Dalton, J. R., & Ong, N. T. (2005). Authority orientations and democratic attitudes: A test of the 'Asian values' hypothesis. *Japanese Journal of Political Science,* 6(2), 211-231.

Dalton, R., Shin, D. C., & Jou, W. (2007). Understanding democracy: Data from unlikely places. *Journal of Democracy,* 18(4), 142-156.

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods,* 2(1), 33-46.

Davidov, E. (2009). Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. *Political Analysis,* 17(1), 64-82.

Davidov, E., Schmidt, P., & Billiet, J. (Eds.), (In press). *Cross-cultural analysis: Methods and applications*. New York: Taylor and Francis.

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

De Beuckelaer, A. (2005). *Measurement invariance issues in international management research.* Unpublished dissertation, Limburgs University Centrum, Centrum, Limburg, the Netherlands.

Diamond, L. J. (1999). *Developing democracy: Toward consolidation*. Baltimore: Johns Hopkins University Press.

Dixon, J. C. (2008). A clash of civilizations? examining liberal-democratic values in Turkey and the European Union. *British Journal of Sociology,* 59(4), 681-708.

Esmer, Y. (2002). Is there an Islamic civilization? *Comparative Sociology,* 1(3-4), 265-298.

Guerin, D., Pétry, F., & Crete, J. (2004). Tolerance, protest and democratic transition: Survey evidence from 13 Post-Communist countries. *European Journal of Political Research,* 43(3), 371-395.

Haerpfer, C. W. (2008). Support for democracy and autocracy in Russia and the Commonwealth of independent states, 1992-2002. *International Political Science Review,* 29(4), 411-431.

Harkness, J. A., Van de Vijver, Fons, J.R., & Mohler, P. P. (2003). *Cross-cultural survey methods* Hoboken, New Jersey: Wiley-Interscience.

Heath, A., Fisher, S., & Smith, S. (2005). The globalization of public opinion research. *Annual Review of Political Science,* 8, 297-333.

Heath, A., & Martin, J. (1997). Why are there so few formal measuring instruments in social and political research. In L. E. Fyberg, et al. (Eds.), *Survey measurement and process quality* (pp. 71-86). New York: Wiley.

Hofmann, S. R. (2004). Islam and democracy: Micro-level indications of compatibility. *Comparative Political Studies, 37*(6), 652-676.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research,* 18(3), 117-144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6(4), 1-55.

Inglehart, R. (2003). How solid is mass support for Democracy—and how can we measure it? *Political Science and Politics, 36*(1), 51-57.

Inglehart, R., & Norris, P. (2003). The true clash of civilizations. *Foreign Policy,* 135, 63-70.

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

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA Nachrichten Spezial,* 31-40.

Jowell, R., Roberts, C., Fitzgerald, R., & Gillian, E. (Eds.), (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. London: Sage Publications.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review,* 97(4), 567-583.

Kittilson, M., Caul. (2007). Research resources in comparative political behavior. In J. R. Dalton, & H. Klingemann (Eds.), *Oxford handbook of political behavior* (pp. 865-895). Oxford: Oxford University Press.

Klingemann, H. (1999). Mapping political support in the 1990s: A global analysis. In P. Norris (Ed.), *Critical citizens: Global support for democratic Governance* (pp. 151-189). Oxford: Oxford University Press.

Knoppen, D., & Saris, W. E. (2009). Do we have to combine values in the Schwartz' human values scale? A comment on the Davidov studies. *Survey Research Methods,* 3(2), 91-103.

Lagos, M. (2003). Support for and satisfaction with democracy. *International Journal of Public Opinion Research,* 15(4), 471-487.

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.

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling,* 11(4), 514-534.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers

in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling,* 11(3), 320-341.

Mishler, W., & Rose, R. (2001). Political support for incomplete democracies: Realist vs. idealist theories and measures. *International Political Science Review,* 22(4), 303-320.

Norris, P. (Ed.), (1999). *Critical citizens: Global support for democratic government*. Oxford: Oxford University Press.

Norris, P. (2009). The globalization of comparative public opinion research. In N. Robinson, & T. Landman (Eds.), *Handbook of comparative politics* (pp. 522-539) London: Sage.

Pettersson, T. (2008). Different patterns of compartmentalisation among Muslims and Christians? In J. Haynes (Ed.), *Routldge handbook of religion & politics* (pp. 246-270). New York: Routledge.

Reeskens, T., & Hooghe, M. (2008). Cross-cultural measurement equivalence of generalized trust. evidence from the European Social Survey (2002 and 2004). *Social Indicators Research,* 85(3), 515-532.

Rizzo, H., Abdel-Latif, A., & Meyer, K. (2007). The relationship between gender equality and democracy: A comparison of Arab versus non-Arab Muslim societies. *Sociology,* 41(6), 1151-1170.

Rowley, C. K., & Smith, N. (2009). Islam's democracy paradox: Muslims claim to like democracy, so why do they have so little? *Public Choice,* 139(3), 273-299.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New Jersey: Wiley.

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling,* 16(4), 561-582.

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.

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

Smith, T. W. (2003). Developing comparable questions in cross-national surveys. In J. A. Harkness, Van de Vijver, Fons, J.R. & P. Mohler P.H. (Eds.), *Cross-cultural survey methods* (pp. 69-91). New York: John Wiley & Sons.

Steenbergen, M. R. (2000). Item similarity in scale analysis. *Political Analysis,* 8(3), 261-283.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research,* 25(1), 78-107.

Tessler, M. (2002). Islam and democracy in the Middle East: The impact of religious orientations on attitudes toward democracy in four Arab countries. *Comparative Politics,* 34(3), 337-354.

Tessler, M., Moaddel, M., & Inglehart, R. (2006). What do Iraqis want? *Journal of Democracy,* 17(1), 38-50.

Toros, E. (2010). The relationship between Islam and democracy in Turkey: Employing political culture as an indicator. *Social Indicators Research,* 95(2), 253-265.

Tusicisny, A. (2007). Security communities and their values: Taking masses seriously. *International Political Science Review,* 28(4), 425-449.

Van de Vijver, Fons, J.R. (2003). Bias and equivalence: Cross-cultural perspective. In J. A. Harkness, Van de Vijver, Fons, J.R. & P. Mohler P.H. (Eds.), *Cross-cultural survey methods* (pp. 143-155). New York: John Wiley & Sons.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods,* 3(1), 4-70.

Wang, Z., Dalton, R., J.D., & Shin, D. C. (2006). Political trust, political performance, and support for democracy. In J. R. Dalton, & D. C. Shin (Eds.), *Citizens, democracy, and markets around the Pacific rim* (pp. 135-154). Oxford: Oxford University Press.

Welkenhuysen-Gybels, J. (2004). The performance of some observed and unobserved conditional invariance techniques for the detection of differential item functioning. *Quality and Quantity,* 38(6), 681-702.

Welkenhuysen-Gybels, J., & Billiet, J. (2002). A comparison of techniques for detecting cross-cultural inequivalence at the item level. *Quality and Quantity,* 36(3), 197-218.

Welzel, C. (2007). Are levels of democracy affected by mass attitudes? testing attainment and sustainment effects on democracy. *International Political Science Review*, 28(4), 397-424.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment Research & Evaluation,* 12(3), 1-26.

Table 1: Items in the study and their scale

Construct	Item	Content	Scale
Democracy-Autocracy Preference (DAP)	V164	Having a strong leader who does not have to bother with parliament and elections.	1 = Very good 2 = Fairly good 3 = Fairly Bad 4 = Very bad
	V165	Having experts, not governments, make decisions according to what they think is best for the country	
	V166	Having the army rule	
	V167	Having a democratic political system	
Democratic Performance Evaluation (DPE)	V169	In democracy, the economic system runs badly	1= Agree strongly 2=Agree 3=Disagree 4= Strongly disagree
	V170	Democracies are indecisive and have too much quibbling	
	V171	Democracies aren't good at maintaining order	
	V172	Democracy may have problems but it's better than any other form of government	

Note: V172 and V167 were coded in the same direction as the other items.

Table 2: Single-Country Confirmatory Factor Analyses: Modifications and Global Fit for the "Democracy-Autocracy Preference" Scale across 36 countries

Country	Modification [A]	Pclose	RMSEA	CFI
1. Albania		.020	.097	.936
2. Algeria		.481	.047	.970
3. Argentina	e 3↔e 4	.768	.000	1.000
4. Bangladesh	e 1↔e 4	.024	.100	.946
5. Bosnia	e 2↔e 4	.043	.097	.942
6. Canada		.470	.040	.989
7. Chile	e 2↔e 4	.002	.136	.972
8. China		.780	.009	.997
9. India	e 3↔e 4	.767	.015	.999
10. Indonesia	e 3↔e 4	.209	.069	.983
11. Iran	e 2 ↔e3	.731	.029	.978
12. Iraq		.860	.028	.990
13. Japan	e 3↔e 4	.665	.025	.997
14. Jordan	e 2↔e 4	.897	.000	1.000
15. Kyrgyz	e 2↔e 4	.893	.000	1.000
16. Macedonia	e 2↔e 4	.360	.052	.993
17. Mexico	e 2↔e 4	.488	.042	.994
18. Moldova	e 2↔e 4	.712	.000	1.000
19. Montenegro	e 2↔e 4	.015	.117	.975
20. Morocco	e 2↔e 4	.914	.000	1.000
21. Nigeria	e 3↔e 4	.061	.082	.973
22. Pakistan		.434	.050	.988
23. Peru	e 2↔e 4	.132	.075	.961
24. Philippines	e 3↔e 4	.413	.048	.995
25. Puerto Rico	e 3↔e 4	.646	.000	1.000
26. Serbia		.350	.055	.985
27. Sought Korea	e 3↔e 4	.844	.000	1.000
28. South Africa		.998	.011	.999
29. Spain		.770	.025	.997
30. Tanzania		.983	.000	1.000
31. Turkey	e 3↔e 4	.016	.086	.974
32. Uganda	e 3↔e 4	.038	.102	.978
33. U.S.A		.973	.000	1.000
34. Venezuela	e 2↔e 4	.000	.199	.785
35. Vietnam	e 1↔e 4	.720	.000	1.000
36. Zimbabwe	e 2↔e 4	.037	.104	.937

Note. Pclose - probability of close fit; RMSEA - Root Mean Square Error of Approximation; CFI - Comparative Fit Index. [A] ↔covariance between errors is freely estimated. Empty cells in the second column indicate that no modification was necessary in the single-country analysis.

Table 3: Unstandardized and Standardized Factor Loadings on Democracy-Autocracy Preference in the Single Country Analyses (Standardized factor loadings in parentheses)

Country	DAP→ V164	DAP → V165	DAP → V166	DAP→ V167
1. Albania	1.00(.75)	-.01(-.01)	.73(.61)	.21(.24)
2. Algeria	1.00(.84)	.10(.11)	.35(.36)	.18(.22)
3. Argentina	1.00(.73)	.69(.52)	.57(.46)	.44(.42)
4. Bangladesh	1.00(.46)	1.16(.45)	1.36(.56)	-.00(-.00)
5. Bosnia	1.00(.72)	.35(.28)	.32(.31)	.22(.23)
6. Canada	1.00(.71)	.75(.52)	.48(.50)	.44(.37)
7. Chile	1.00(.63)	.65(.43)	1.05(.69)	.80(.61)
8. China	1.00(7.62)	.00(.03)	.00(.01)	.00(.00)
9. India	1.00(.81)	.43(.42)	.27(.26)	.03(.04)
10. Indonesia	1.00(.60)	.85(.51)	.07(.06)	.20(.15)
11. Iran	1.00(.067)	-6.41(-.43)	-5.95(-.46)	3.47(.28)
12. Iraq	1.00(.87)	.02(.02)	.37(.38)	.21(.24)
13. Japan	1.00(1.01)	.26(.31)	.10(.22)	.09(.13)
14. Jordan	1.00(.70)	.38(.35)	.69(.49)	.10(.11)
15. Kyrgyz	1.00(.75)	.60(.50)	.39(.30)	.20(.19)
16. Macedonia	1.00(.80)	.65(.61)	.48(.34)	.01(.02)
17. Mexico	1.00(.82)	.50(.45)	.45(.38)	.06(.06)
18. Moldova	1.00(.43)	.67(.36)	.58(.34)	1.00(.49)
19. Montenegro	1.00(.72)	.05(.04)	.77(.67)	.64(.59)
20. Morocco	1.00(.33)	.69(.21)	1.71(.64)	.48(.28)
21. Nigeria	1.00(.95)	.28(.30)	.27(.27)	.08(.13)
22. Pakistan	1.00(.82)	.41(.41)	.12(.16)	.58(.59)
23. Peru	1.00(.92)	.26(.27)	.23(.26)	.02(.03)
24. Philippines	1.00(.78)	.78(.62)	.73(.50)	-.12(-.11)
25. Puerto Rico	1.00(.91)	.54(.50)	.41(.47)	.11(.14)
26. Serbia	1.00(.82)	-.01(-.02)	.58(.59)	.32(.38)
27. Sought Korea	1.00(1.00)	.14(.17)	.24(.35)	.05(.06)
28. South Africa	1.00(.83)	.57(.51)	.34(.36)	.18(.21)
29. Spain	1.00(.74)	.90(.59)	.59(.48)	.31(.27)
30. Tanzania	1.00(.66)	.72(.25)	1.62(.69)	.09(.09)
31. Turkey	1.00(.65)	.85(.64)	.31(.20)	-.02(-.02)
32. Uganda	1.00(.56)	1.32(.76)	.86(.53)	.02(.01)
33. USA	1.00(.72)	.75(.55)	.57(.53)	.29(.25)
34. Venezuela	1.00(.31)	.62(.21)	2.17(.73)	.71(.33)
35. Vietnam	1.00(.62)	1.13(.67)	1.18(.50)	-1.69(-.66)
36. Zimbabwe	1.00(.77)	.24(.18)	.35(.32)	.31(.30)

Table 4: Multiple-Group Confirmatory Factor Analysis: Fit Measures of the Invariance Test for the DAP scale across 27 countries (after excluding V167)

Model	Chi-square	Degrees of freedom	Pclose	RMSEA	CFI
4a. Full metric invariance	643	52	1.00	.018	.928
4b. Full scalar invariance	8,529	104	1.00	.045	.000
4c. Partial scalar invariance [A]	3,704	55	1.00	.040	.554

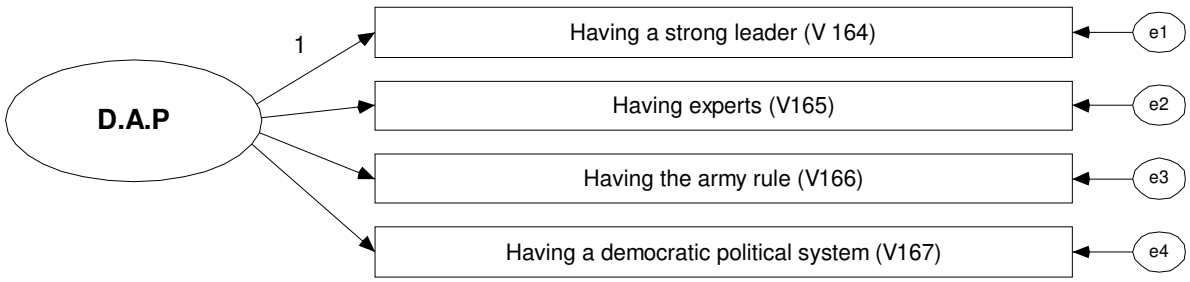
Note: Pclose - probability of close fit; RMSEA - Root Mean Square Error of Approximation; CFI - Comparative Fit Index. [A] After releasing constraints for v166

Table 5: Multiple-Group Confirmatory Factor Analysis: Fit Measures of the Invariance Test for the DPE scale across 35 countries (after excluding V172)

Model	Chi-square	Degrees of freedom	Pclose	RMSEA	CFI
5a. Full metric invariance	389	68	1.00	.010	.990
5b. Full scalar invariance	3144	136	1.00	.028	.909
5c. Partial scalar invariance [A]	635	68	1.00	.028	.983

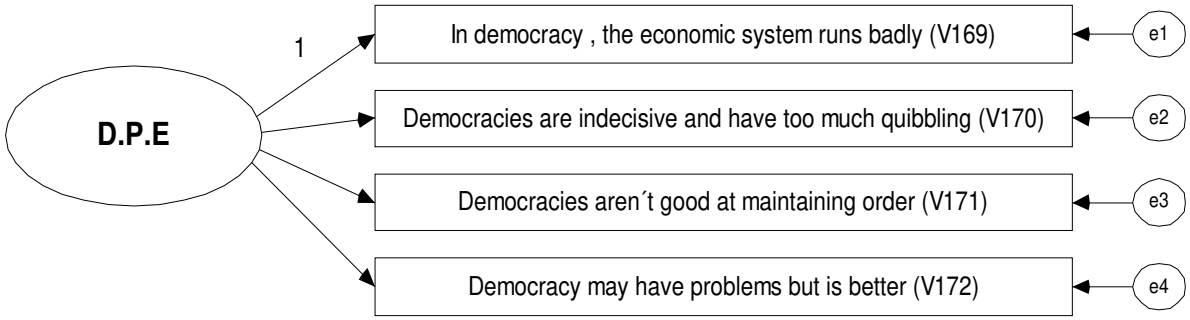
Note: Pclose - probability of close fit; RMSEA - Root Mean Square Error of Approximation; CFI - Comparative Fit Index. [A] After releasing constraints for v170.

Figure 1a: Confirmatory Factor Analysis Model of Democracy-Autocracy Preference (DAP) Scale



Note: V167 is a reversed indicator. V164 is a reference indicator. *Source*: Data are taken from the WVS 2000, *N*=51067.

Figure 1b: The Democratic Performance Evaluation Scale



Note: V172 is a reversed indicator. V169 is a reference indicator. *Source*: Data are taken from the WVS 2000, *N*=51067.