



# Alternative Measures of Political Efficacy: The Quest for Cross-Cultural Invariance With Ordinally Scaled Survey Items

Thomas J. Scotto<sup>1\*</sup>, Carla Xena<sup>2</sup> and Jason Reifler<sup>3</sup>

<sup>1</sup>School of Social and Political Sciences, University of Glasgow, Glasgow, United Kingdom, <sup>2</sup>Institute for Social and Economic Research, University of Essex, Essex, United Kingdom, <sup>3</sup>Department of Politics, University of Exeter, Exeter, United Kingdom

In this paper, we examine the measurement of citizens' beliefs that politicians and political systems are responsive (external efficacy) and that citizens see themselves sufficiently skilled to participate in politics (internal efficacy). This paper demonstrates techniques that allow researchers to establish the cross-context validity of conceptually important ordinal scales. In so doing, we show an alternative set of efficacy indicators to those commonly appearing on cross-national surveys to be more promising from a validity standpoint. Through detailed discussion and application of multi-group analysis for ordinal measures, we demonstrate that a measurement model linking latent internal and external efficacy factors performs well in configural and parameter invariance testing when applied to representative samples of respondents in the United States and Great Britain. With near full invariance achieved, differences in latent variable means are meaningful and British respondents are shown to have lower levels of both forms of efficacy than their American counterparts. We argue that this technique may be particularly valuable for scholars who wish to establish the suitability of ordinal scales for direct comparison across nations or cultures.

**Keywords:** cross-cultural validity, political efficacy, structural equation analyses, latent variable analyses, survey research

## OPEN ACCESS

### Edited by:

Katharina Meitinger,  
Utrecht University, Netherlands

### Reviewed by:

Márton Bene,  
Hungarian Academy of Sciences  
(MTA), Hungary  
Eldad Davidov,  
University of Cologne, Germany

### \*Correspondence:

Thomas J. Scotto  
Thomas.Scotto@glasgow.ac.uk

### Specialty section:

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

**Received:** 08 February 2021

**Accepted:** 21 June 2021

**Published:** 16 July 2021

### Citation:

Scotto TJ, Xena C and Reifler J (2021)  
Alternative Measures of Political  
Efficacy: The Quest for Cross-Cultural  
Invariance With Ordinally Scaled  
Survey Items.  
*Front. Polit. Sci.* 3:665532.  
doi: 10.3389/fpos.2021.665532

## 1 INTRODUCTION

Answering key political and social science questions often requires operationalizing unobserved or latent constructs that are measured by a series of ordinal survey questions. Prominent examples include scales that capture racial animosity otherwise missed by more explicit measures and or concepts like civic engagement and life satisfaction.<sup>1</sup> Across nations, interest lies in comparing and understanding the causes and consequences of terms such as a nation's "level of democracy" that, to the lay-person, should be straightforward to understand and observe but prove to be elusive concepts that are difficult to measure.<sup>2</sup>

<sup>1</sup>For a discussion on symbolic racism see, e.g., Tarman and Sears (2005) and Gomez and Wilson (2006); and for civic engagement and life satisfaction conceptualised as latent variables, see Brehm and Rahn (1997).

<sup>2</sup>For a discussion of democracy as an unobserved latent variable, see Bollen and Jackman (1989); for a trenchant critique concerning the measurement assumptions behind the measurement assumptions commonly made when conducting multivariate research, see Treier and Jackman (2008).

Whether survey items designed in one group or nation are equally valid or “mean the same thing” in other contexts often is overlooked. Coming up with a measure of racism in its modern form is difficult enough if we wish to understand its levels within the United States. Comparing the level of racism manifested by the United States population to what we might find in another country is an incredibly complex undertaking. Can racism scales designed for use in the United States work in other contexts simply by replacing the outgroup referent? How can scholars assess whether scales and latent variables used by social scientists possess cross-cultural validity?

This question of cross-context comparability is particularly important for cross-national comparative work, particularly as multi-nation surveys such as the European Social Survey (ESS), the Comparative Study of Electoral Systems (CSES), and the World Values Survey provide exciting opportunities to examine attitudes and behavior across national contexts. Yet only a small number of concepts on these surveys are subject to cross-cultural validation via Multigroup Confirmatory Factor Analysis (MGCF). Examples include Reeskens and Hooghe, (2008) and Coromina and Peral (2020), who explore the three item ESS battery for political trust, and an exploration of the 2008–2009 ageism battery by Seddig, Maskileyson and Davidov (2020). Davidov et al. (2008) employ MGCF to examine the cross-cultural validity of a battery of basic human values developed by Schwartz (1994). Many of these examples rely on assuming the variables are continuous, which is a concern as Lubke and Muthén (2004) show that treating ordinal indicators as continuous can be problematic in the analysis of multiple groups using structural equation modeling. An exception is Meuleman and Billiet (2012) who treat the scale of the 2002–2003 ESS immigration battery as ordinal in their study of the scale’s cross-national validity.

This paper combines an instructional and substantive aim: We demonstrate the use of MGCF with ordinal indicators to show cross-national validity of less commonly used political efficacy measures. The paper is motivated by the work of Xena (2015) who suggests that the indicators for these concepts fielded as part of the ESS in the early 2000s lack cross-cultural validity. In light of these findings, we examine an alternative set of efficacy indicators developed by Craig, Niemi and their associates (Craig, Niemi and Silver 1990; Niemi, Craig and Mattei 1991). As we intend to write this article partially as an instructional tool, we build on the work of Millsap and Yun-Tein (2004), Temme (2006) and the efforts of Davidov et al. (2018) to describe the process for conducting MGCF with ordinal data. Employing the Mplus software package, our analyses of alternative indicators for internal and external efficacy have almost the same structure across the two groups (in this case, the United States and Great Britain). This analysis leads to two key substantive findings. First, Britons express less political efficacy than their American counterparts. Second, there is higher within country variation in political efficacy for Americans. The paper concludes with suggesting the possibility to abandon the traditional indicators of political efficacy and derivatives thereof on future cross-national surveys and avenues to consider in further developing the alternative indicators employed in this paper.

## 2 POLITICAL EFFICACY: A CONTESTED LITERATURE

The measurement of political efficacy matters because of the concept’s theoretical importance: an efficacious citizenry is more likely to confer legitimacy on political systems and avoid the types of disillusionment that generate civic and participatory decline, or worse outcomes such as illegal political activity or violent protest movements (Easton and Dennis 1967; Finifter 1970; Pateman 1970). However, testing theoretical claims across contexts requires validating that the measures themselves are valid across contexts.

Although early work treats efficacy as a uni-dimensional construct, Lane, (1959: 149) argues that efficacy “combines the image of the self and the images of democratic government” to suggest that two distinct concepts are important. Nonetheless, survey questions remained uni-dimensional. Classic analyses of data from the American National Election Studies (ANES) in the 1950s employ the following four items, asking respondents their levels of disagreement or agreement with the statements (cf. Campbell et al. (1954: 187–188)<sup>3</sup>: 1) I don’t think public officials care what people like me think; 2) Voting is the only way that people like me can have any say about the way the government runs things; 3) People like me don’t have any say about what the government does; and 4) Sometimes government and politics seem so complicated that a person like me can’t really understand what’s going on. Two additional statements appear on the ANES from 1968 to 1980 (Acock and Clarke, 1990): 5) Parties are only interested in people’s votes, not opinions; and 6) Generally speaking, those we elect to Congress in Washington loose touch with the people pretty quickly. The modern cross-national incarnation employed the first wave of the European Social Survey in 2002 utilises items 1, 4, and 5 and adds questions asking respondents (see Xena, 2015): 7) Do you think you could take an active role in a group involved with political issues? and 8) How easy is it to make up your mind about political issues?

Balch (1974) finds that items two and four have a modest correlation with conventional and unconventional participation and are nearly unrelated to attitudes towards political trust. In contrast, items one and three relate better to attitudes towards trust. This analysis further justifies treating efficacy as multi-dimensional: Items two and four are reflective of an individual’s “confidence in his own abilities regardless of political circumstances” and therefore a reflection of internal efficacy. Items one and three correspond to respondents’ beliefs about “the potential responsiveness of individuals” or external efficacy (Balch 1974: 24). As 5 and 6 enter the survey, Miller and Traugott, 1989 argue that item 3 (along with items two and 4) is now reflective of internal efficacy.

Analyzing ANES data from 1972 to 1976, Craig and Maggiotto (1982) question the conceptual validity of the indicators, particularly the idea that item four reflects the internal dimension. They argue that item two also is

<sup>3</sup>ANES studies included additional survey questions to measure efficacy, but these four are those that persistently appear in published research.

problematic because disagreement can be an efficacious response if the individual believes there are other avenues to effective political participation. Acock et al. (1985) argue that indicators are salvageable if researchers: 1) drop item 2, 2) specify items 3 and 4 as internal efficacy indicators, 3) assign items 5 and 6 as external efficacy indicators, and 4) allow item one to load on both latent dimensions. Using data from seven western countries, they find that model fit is adequate across groups, and the dimensions are appropriately associated with external validators. Subsequent research by Acock, Clarke, and colleagues employs these indicators to study the change in efficacy over the course of an election (Clarke and Acock, 1989), differences across levels of government in the Canadian system (Stewart et al., 1992), or in further validation exercises to cope with additional revisions to the ANES battery on the 1984 study (Acock and Clarke, 1990).

Xena's (2015) research on the cross-national validity of a modified form of the traditional efficacy indicators designates items 4, 7, and 8 as reflective of internal efficacy and items one and 5 indicative of external efficacy. Single country CFAs testing fit of the data to this model across 21 European countries reveals less than ideal model fit. Moreover, an examination of the factor loadings or relationships between the latent factors and the designated indicators reveals that the loadings of the pair linked to the external dimension to be reasonably stable in magnitude across countries. However, wide variation in the size of the three items designated to be reflective of internal efficacy mimics the problems Craig and Maggiotto (1982) identify using ANES data. Diagnostic statistics suggest that for some of the countries, certain statements hypothesized to be reflective of internal efficacy actually fit better when a path opens between indicators 4, 7, and/or eight and the latent external efficacy dimension. Results from invariance testing using the Multi-Group Confirmatory Factor Analysis (MGCFA) techniques we describe below lead Xena (2015: 67) to conclude that the indicators of "political efficacy used by the ESS [in] 2002 is not invariant across Europe, as partial invariance, required in the ordinal case to guarantee measurement equivalence is not supported." Thus, comparing mean scores on the latent dimensions across nations is not comparing like-for-like.

The lack of cross-cultural validity for efficacy measures may require scholars to revisit previous findings. For example, closer inspection of Muller's (1970) classic five nation study examining the ability of efficacy to influence political participation reveals that loadings for efficacy indicators (generalised variants of numbers 3–5) vary considerably across nations. Thus, comparison of the latent variable scores across nations and in follow-up multivariate research may be invalid.

In the late 1980s, efforts to replace problematic indicators proceeded in a piecemeal fashion. Craig et al. (1990: 289–290) note that the ad hoc process and lack of consensus results both in a loss of cross-temporal and cross-national validity and "without rigorous prior testing" fails to reassure that substitute indicators are any more valid or reliable. The authors make use of the 1987 ANES Pilot Study as an instrument for revising items pertaining to trust and efficacy. Starting with the premise that efficacy is a multidimensional concept, internal efficacy should have

relationships with campaign participation, political knowledge, and interest that exceeds that of external efficacy. In contrast, political trust should have a higher correlation with external efficacy. Six revised indicators for internal political efficacy emerged with relative ease (see **Section 3** for wording) demonstrating hypothesized associations. Four indicators of external efficacy materialise, but Craig et al. (1990) are more tepid in unabashedly recommending them because they combine the concepts of belief that the regime *and* current political figures are responsive to the political desires of individual. Although some contend regime and incumbent based external efficacy differ, the researchers find that they perform well together as indicators of the single concept of external efficacy. A follow-up study analyzing the performance of the ten indicators as reflections of internal and external efficacy after their placement on the much larger 1988 ANES further validates their performance (Niemi et al., 1991).

The intention for these measures is that they are "comparable across times, places, and populations" (Craig et al., 1990: 296). Morrell (2003) conducts an extensive review and finds the internal efficacy indicators to perform well across a multitude of contexts but remains agnostic as to the utility of the four indicators for external efficacy. Furthermore, while Morrell indicates that researchers employ the revised internal efficacy indicators in a variety of contexts, formal testing of their cross-group and cross-national comparability is absent. The updated external efficacy indicators have yet to receive serious scrutiny. Given this shortcoming, we turn to subjecting the ten revised indicators to empirical tests of cross-cultural validity—multigroup analyses with representative samples drawn from populations of the United States and United Kingdom.

### 3 DATA AND INDICATORS

In late May and early June 2012, an online survey with a primary focus on the measurement of citizen attitudes towards international affairs was fielded to samples of respondents matched to the British and American populations.<sup>4</sup> Indicators of internal and external efficacy are a carbon copy of those of Craig et al. (1990), and are presented below.

#### •Internal Efficacy:

1. I feel that I could do as good of a job in public office as most other people. (PUBOFF)
2. I think I am as well-informed about politics and government as most people. (INFORM)

<sup>4</sup>Fieldwork conducted by YouGov and this wave constitutes the second in a series of the multi-country panel studies. The effective sample size for the British sub-sample is 2,318 and the effective sub-sample size for the American group is 2,295. Funding for the surveys came from ESRC Grant RES-061–25–0405, and support for Scotto and Xena's time on the project comes from the ESRC Research Centre on Micro-Social Change (MISOC) at Essex.

**TABLE 1 |** Distributions of hypothesized internal efficacy indicators.

		n	Strongly agree	Agree	Neither Agr./Dis.	Disagree	Strongly disagree	$\chi^2$
PUBOFF	Unites States	2,253	24.1%	32.9	25.3	11.3	6.4	97.8
	United Kingdom	2,255	14.1%	36.6	27.4	17.1	4.7	( $p = 0.00$ )
INFORM	Unites States	2,266	30.6%	35.3	20.4	8.5	5.1	244.8
	United Kingdom	2,283	12.2%	47.8	26.7	9.2	4.1	( $p = 0.00$ )
NOT-SURE	Unites States	2,265	5.7%	13.1	23.1	25.3	32.8	298.5
	United Kingdom	2,284	4.0%	24.1	27.6	31.3	13.1	( $p = 0.00$ )
UNDER-STAND	Unites States	2,276	32.4%	40.3	19.9	4.6	2.8	278.1
	United Kingdom	2,291	12.7%	49.3	26.2	9.6	2.1	( $p = 0.00$ )
SELF-QUAL	Unites States	2,272	22.5%	29.9	27.8	11.2	8.7	284.9
	United Kingdom	2,294	7.7%	26.1	33.8	23.5	8.8	( $p = 0.00$ )
COM-PLEX	Unites States	2,268	9.0%	20.4	22.6	24.3	23.7	160.2
	United Kingdom	2,292	7.3%	30.6	25.5	25.5	11.1	( $p = 0.00$ )

3. I don't often feel sure of myself when talking with other people about politics and government. (NOTSURE)
4. I feel that I have a pretty good understanding of the important political issues facing our country. (UNDERSTAND)
5. I consider myself well-qualified to participate in politics. (SELFQUAL)
6. Sometimes politics and government seem so complicated that a person like me can't really understand what's going on. (COMPLEX)

•External Efficacy:

1. There are many legal ways for citizens to successfully influence what the government does. (LEGAL)
2. Under our form of government, the people have the final say about how the country is run, no matter who is in office. (FINALSAY)
3. If public officials are not interested in hearing what the people think, there is really no way to make them listen. (MAKELSTN)
4. People like me don't have any say about what the government does. (NOSAY)

Respondents receive the questions (along with others measuring political trust) on two separate grids and the item order rotates. There are six possible responses interviewees can provide to each item: Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree, and Don't Know. "Don't Know" responses are coded as missing and the remainder of the scale is ordinal.<sup>5</sup> Efficacious responses receive higher scores, so the NOTSURE, COMPLEX, MAKELSTN, and NOSAY indicators are reverse coded.

**Table 1** compares the response distributions for the six hypothesized internal efficacy indicators for British and American respondents. For each statement, the dispersion of responses across the five categories differs significantly across the two nations. With the exception of British responses to the

COMPLEX indicator, respondents in both groups are more likely than not to provide an efficacious response. At a glance, Americans appear, on the whole, more likely than their British counterparts to provide responses at the both extreme ends of the scale. As we discuss in the next section, this empirical result underscores the importance of taking into serious consideration the relationship between the hypothesized latent variable and indicator that comes *via* the factor loadings and the thresholds.

**Table 2** tallies the responses, by country, to the statements hypothesized to be reflective of the external efficacy dimension. Again, Americans appear more likely than their British counterparts to provide answers at the endpoints of the scales. For each indicator, answers by country are statistically different from one another (although this is just so for the NOSAY indicator at  $p < 0.05$ ). Results across tables suggest that respondents are more likely exhibit high levels of internal efficacy, with Britons particularly more apt to provide inefficacious answers to the external efficacy indicators.

## 4 MULTI-GROUP CONFIRMATORY FACTOR ANALYSIS WITH ORDINAL DATA

### 4.1 Model Specifics

Parameters from the Multi-Group Confirmatory Factor Analysis (MGCFA) with ordinal data in Mplus are obtained *via* its Weighted Least Squares Estimator (WLSMV) that utilizes the ordered probit link function (Muthén and Muthén 2017). As implemented, respondents' level of agreement or disagreement with the ten indicators is not directly a function of their location on the latent internal or external factors. The answers to the statements we observe instead are hypothesized to come indirectly *via* a continuous and multivariate normally distributed latent response variable,  $y^*$ , which has a theoretical range from  $-\infty$  to  $+\infty$  and is the direct link to the factor loadings and respondent positions on the latent factors. Observed ordinal responses ranging from strongly disagree to strongly agree (or vice versa for the four reverse coded indicators) are obtained via splitting the response variable by four thresholds such that once a respondent reaches a position on the response variable that exceeds a given threshold, they move to a higher response category. Following Temme (2006), the following two equations are estimated, simultaneously:

<sup>5</sup>In the multivariate analyses, the "pairwise present" method of recovering missing observations, as implemented in Mplus for the WLSMV estimator, is employed. Over 90% of each element of the asymptotic variance-covariance matrix for each group is observed.

**TABLE 2 |** Distributions of hypothesized external efficacy indicators.

		n	Strongly agree	Agree	Neither Agr./Dis	Disagree	Strongly disagree	$\chi^2$
LEGAL	Unites States	2,248	16.0%	39.8	25.9	12.9	5.5	308.0 ( $p = 0.00$ )
	United Kingdom	2,183	4.0%	31.0	29.9	26.0	9.1	
FINAL	Unites States	2,256	10.2%	15.8	20.4	27.0	26.6	142.8
SAY	United Kingdom	2,271	2.3%	14.2	21.5	35.7	26.3	( $p = 0.00$ )
MAKE	Unites States	2,261	16.7%	24.1	17.9	24.9	16.4	233.2
LSTN	United Kingdom	2,264	18.5%	39.1	16.3	21.3	4.8	( $p = 0.00$ )
NOSAY	Unites States	2,263	13.9%	22.5	24.5	24.8	14.3	10.3 ( $p = 0.04$ )
	United Kingdom	2,292	18.3%	36.4	21.3	19.9	4.1	

$$y_{ijg}^* = \alpha_{jg} + \sum_{p=1}^2 \lambda_{jpg} \eta_{ipg} + \epsilon_{ijg} \tag{1}$$

$$y_{ijg} = c, \text{ if } \tau_{jcg} < y_{ijg}^* \leq \tau_{j,c+1,g} \tag{2}$$

**Equation 1** states that an individual (*i*) for an indicator (*j*) in a group (*g*) has a position on the response variable,  $y^*$ , which is an additive function of an indicator specific intercept for each group ( $\alpha_{jg}$ ), a respondent’s position on each of the two latent variables,  $\eta_{ipg}$ , multiplied by the indicator specific factor loading for the group, ( $\lambda_{jpg}$ ), plus residual variance specific to the individual, indicator, and group ( $\epsilon_{ijg}$ ).<sup>6</sup> **Eq. 2** states that we observe an individual, *i*, in the United Kingdom or United Kingdom (*g*) make a choice, *c*, on an indicator, *j*, if their position on the continuous response variable falls in between two thresholds. The implications of **Eqs 1, 2** is that the magnitude of each threshold plays a key part in predicting an individual’s response category. As a consequence, thresholds should neither be suppressed nor assumed to be equivalent across groups in Confirmatory Factor Analysis (CFA) with ordinal level indicators.

Due to the joint estimation of the thresholds, intercepts, and factor loadings when data are ordinal, the procedures for MGCFA differ in substantive ways in comparison to procedures employed when the data for the observed indicators are continuous. For the latter, it is possible and often desirable to conduct MGCFA in sequential stages, first testing for equivalence and differences in the factor loadings (metric invariance) and then testing for the equality of item specific intercepts (scalar invariance) across groups (e.g. Meredith 1993; Little 1997). Metric invariance indicates that the underlying latent construct is the same across groups. When established after metric invariance, scalar invariance signals that individuals with the same values on a latent variable across groups will have the same observed value for the indicator across the groups (Hong et al., 2003).

For multi-group analyses with continuous observed data, the factor loadings alone are enough to establish the slopes linking latent variable to indicators. However, as Davidov et al. (2011: 160) note, the resulting observed category responses on ordinal

indicators are, as **Eqs. 1, Eqs 2** imply, “jointly influenced by the factor loadings . . . [and] thresholds . . . [meaning] a distinction between metric and scalar invariance is not substantively meaningful [and] there [should be] only one step in the measurement invariance test, the step that constrains all parameters to be equal.” A difference in the magnitude of the factor loadings across groups is a difference in the strength of the relationship between the latent variable and continuous response variable, while dissimilarities in the thresholds signify that respondents in different groups must reach distinct levels on the response variable to move from one answer category to the next. Erroneously holding thresholds equivalent when they are not could force group differences to artificially appear in the loadings, leading researchers to draw the wrong conclusions about the substantive differences in the indicators. For our data, this distinction is important—recall from **Tables 1, 2** that American respondents appear far more likely than their British counterparts to offer responses at the extreme ends of the ordinal scale; constraining thresholds to be equal might create artificial differences in the factor loadings. As we will see below, much of the observed differences are functions of unequal variances in the latent variables, with additional minor contributions from the unequal group variances of the errors for the ten indicators.

### 4.2 Model Constraints

The equations described above generate a model that is under-identified, meaning restrictions on some of the parameters are necessary before estimation is attempted. As is the case with single group CFA with continuous indicators, the latent variables in a MGCFA with ordinal variables must be scaled either by fixing the loading of one indicator per factor to 1.0 or the variances of the factors to 1.0. As the variance of the Internal and External latent factors is of interest, we choose the former. Much like MGCFA with continuous indicators, identification also necessitates setting the latent variable means for the internal and external efficacy factors,  $\mu_g$ , of a reference group, in our case America, to 0.0 (Byrne 1994).

Additional constraints are necessary when the indicators are ordinal and are dependent partially upon how the researcher wishes to parameterize the model. Mplus offers two options for MGCFA, the so-called “Delta” or “Theta” parameterisations, and the option selected determines some of the model constraints that must be imposed. Under the former, identification is achieved, in part, by setting the variances of the latent variables in the

<sup>6</sup>In our case, the group, *g*, either is the United States or Great Britain and the specific factor, *p*, either is the internal or external efficacy dimension. As we discuss below, for model fitting purposes, we require a third “Methods” factor to be added to the hypothesized model. Thus *p* ranges from one to three in the final model.

reference group all to 1.0. For substantive reasons, we find this empirically untenable and therefore choose the Theta parameterization that requires us, in the first instance, to fix the variance of the residuals for each indicator ( $Var(\epsilon_{ijg})$ ) for both groups to 1.0 Temme (2006).

Following the suggestions of Millsap and Yun-Tein, (2004), we impose the following additional constraints to ensure model identification: 1) We hold lowest category threshold on each indicator ( $\tau_{jg1}$ ) invariant across groups. In the cases where the factor loadings are fixed to 1.0 for identification purposes, we hold the loadings of the bottom two thresholds ( $\tau_{jg1}, \tau_{jg2}$ ) invariant across groups; and 2) As it is of little substantive importance, the intercept for each indicator across all groups,  $\alpha_{jg}$ , is suppressed to zero.

The direction of travel is as follows: First, we take each group separately, and establish the validity of a two factor model of internal and external efficacy in both the United States and United Kingdom. We then combine the data and formally run MGCFA. First, we establish a baseline “Configural Model” where paths between the response variables and indicators ( $\lambda_{jg}$ ) and thresholds ( $\tau_{jgc}$ ) estimated in the separate models are free to vary across groups. At this stage, we fix latent variable means for internal and external efficacy to 0.0 across the groups.<sup>7</sup>

The estimation and validation of the Configural Model produces a baseline  $\chi^2_{WLSMV}$  statistic, which is a reference for difference testing. Comparison models are those where parameters of interest are held to be equal across groups. If the difference in the  $\chi^2_{WLSMV}$  statistics between the baseline and more parsimonious models are statistically indistinguishable, we can state that the latter models fit the data just as well as the former. Not only is a model with equality constraints across the thresholds and factor loadings more parsimonious, but if most of the loadings and thresholds are equal across groups, latent mean and covariance comparisons can be made.

Below we show that a MGCFA model with equality constraints across the thresholds and indicators is tenable after the restriction which fixes the error variances for the observed indicators to 1.0 across groups is relaxed. Testing reveals, in the end, significant differences in the latent mean levels of internal or external efficacy and the relationship between the two in the United States and United Kingdom.

## 5 VALID LATENT MODELS OF POLITICAL EFFICACY

### 5.1 Independent Models of Efficacy–United States and United Kingdom

Confirmatory Factor Analysis (CFA) allows formally testing whether the restrictions placed on the asymptotic covariance matrix by the choice of paths linking the latents to indicators are valid. Separate CFAs for the American and British samples

produce models that fit the data poorly. The  $\chi^2_{WLSMV}$  ( $df = 34$ ) exact fit statistics register 955.7 ( $p = 0.000$ ) for the US sample and 632.3 ( $p = 0.000$ ) for the British group. Close fit statistics also are poor to modest: For the United States, the  $CFI = 0.93$  and the  $RMSEA = 0.11$  ( $p < 0.05 = 0.00$ ), and in the United Kingdom the  $CFI = 0.94$  and the  $RMSEA = 0.09$  ( $p < 0.05 = 0.00$ ).<sup>8</sup>

The source of the poor fit lies with the reverse coded indicators, that is where disagreement with the statement is the more efficacious answer. When a third “methods” factor with freed paths to the reverse coded indicators is added to the model, fit improves substantially (United States:  $\chi^2_{WLSMV} = 301.2$  ( $df = 30$ );  $CFI = 0.98$ ;  $RMSEA = 0.06$  ( $p < 0.05 = 0.00$ ), United Kingdom:  $\chi^2_{WLSMV} = 234.5$  ( $df = 30$ );  $CFI = 0.98$ ;  $RMSEA = 0.05$  ( $p < 0.05 = 0.13$ )).

Despite the vast improvement in model fit by adding a methods factor, substantive misfit remains. In the United Kingdom, PUBOFF, has a negative loading on the internal efficacy factor that is significant but with a magnitude which pales in comparison to the link between the indicator and the external efficacy dimension. Freeing this loading results in approximate fit statistics that suggest a reasonable fit to the data ( $\chi^2_{WLSMV} = 142.5$  ( $df = 29$ );  $CFI = 0.99$ ;  $RMSEA = 0.04$  ( $p < 0.05 = 0.98$ ), and the step can be substantively justified: In the United Kingdom, the association between high levels of external efficacy and the disagreement that one could do as good of a job in public office as others can be interpreted to mean that individuals have a lingering tendency to understand that the “experts” may have a better grasp than them when it comes to representing the public. In the United States, it is the NOTSURE indicator that has a secondary negative loading on the external efficacy dimension; similar to the situation in the United Kingdom, this result suggests that those high in levels of external efficacy might not be those who want to be looked at as experts, even in casual conversation. Adding this single negative loading yields a Confirmatory Factor model for the United States samples that produces adequate approximate fit to the data (US:  $\chi^2_{WLSMV} = 220.9$  ( $df = 29$ );  $CFI = 0.99$ ;  $RMSEA = 0.05$  ( $p < 0.05 = 0.17$ )).

Further improvements in fit are possible for both the United States and United Kingdom samples by allowing the errors of the indicators to correlate. We do not take this further step because doing so because would not only be atheoretical, but could potentially change the meanings of the latent variables, complicating comparisons across time and space. We proceed with a multi-group analysis of a model of political efficacy where

<sup>8</sup>The Comparative Fit Index (CFI) compares the estimated model to one where all indicators are uncorrelated via the  $\chi^2_{WLSMV}$  statistics obtained for each, weighted for the degrees of freedom. As a ratio index, values closer to 1.0 indicate superior models, and Byrne (1994) recommends that approximately fitting models should have values above 0.93. The RMSEA is the average discrepancy between the observed and model implied (asymptotic) covariances weighted by the degrees of freedom in the model (Kline 2005:137–140; Byrne (2012)). RMSEA values greater than 0.10 indicate that a hypothesized SEM has a poor fit to the observed data, and values in the 0.08–0.10 range indicate a “mediocre” fit. Values less than 0.08 indicate “reasonable errors of approximation in the population,” and values lower than 0.05 indicate a close fit.

<sup>7</sup>As a reminder, the variances on the latent variables and the covariance between the internal and external efficacy dimensions are free across groups, and we fix initially the error variances on all of the ten indicators to 1.0 in both groups.

the factor structure of the model is equivalent, save the single exception that the PUBOFF indicator loads on both dimensions in the United Kingdom and the NOTSURE indicator loads on both dimensions in the United States. (Standardised loadings and thresholds may be found in the Online Appendix, along with schematic representations of the hypothesized and final measurement models).

### 5.2 Configural Invariance

Step 1 in multi-group analysis reaffirms that model fit and structure are acceptable for both nations when CFAs for both groups are estimated simultaneously. Additionally, this estimation establishes an unrestricted model to serve as a baseline model in the multigroup setting. As discussed above, testing for simple “configural” invariance involves obtaining parameters from a model where the structure is the same but both factor loadings and thresholds obtained via ordered probit are allowed to vary across groups.<sup>9</sup> Applying this step in our case produces a model with acceptable approximate fit to the data. The combined  $\chi^2_{WLSMV}$  statistic is 533.8 ( $df = 70$ ) and the contribution from each group balances reasonably (United States  $\chi^2_{WLSMV} = 293.5$  and United Kingdom  $\chi^2_{WLSMV} = 240.2$ ), the RMSEA is 0.05 ( $p < 0.05 = 0.08$ ), and the CFI is 0.98. In short, we judge that, with the small caveats involving the one dual loading in each group and the need to include a Methods Factor due to the mode of survey delivery, the latent variable model of internal and external efficacy proposed by Craig et al. (1990) using alternative indicators “travels well” to the United Kingdom.

**Table 3** presents the standardized loadings of each of the indicators on the two dimensions, by country.<sup>10</sup> In both the United States and the United Kingdom, the six indicators hypothesized to load on the internal efficacy dimension do so and we judge them to be significant manifestations of the latent dimensions. Although things can change when equality constraints are placed on the loadings and thresholds, a first glance at the table shows the primary loadings to be similar across the two nations. In a departure from the high correlations we see when employing the traditional indicators of internal and external efficacy (e.g., Acock and Clarke 1990), the correlations between the dimensions are modest in the case of the United States and the two dimensions are nearly orthogonal to one another in the United Kingdom group. The unstandardised variances reported in the bottom half of **Table 3** suggest that respondents in both groups are more dispersed on the internal efficacy dimension and that the latent levels of both types of efficacy for Americans is more dispersed, possibly a function of the

greater number of group respondents who provide “strongly agree” or “strongly disagree” responses to the ten statements.

Loadings of the four indicators hypothesized to load on the external efficacy dimension differ slightly more, and while MAKELSTN is a significant reflective indicator of external efficacy in the United States, the magnitude of the loading is quite modest. The methods factor absorbs measurement error that comes about as a result of the reverse direction of these statements (or other common extraneous covariance that may exist among these indicators). Even after controlling for this form of error, the loadings of these statements on the substantive latent variables are as predicted. Nonetheless, the magnitude of the loadings of these statements on the methods factor is large and, in the case of the NOSAY and MAKELSTN indicators in the United States group greater than the size of the loadings on the predicted external efficacy dimension. We return to this particular finding in the discussion.

**Table 4** lists the standardized thresholds for each indicator. Interpretation is a bit challenging—they are the z-score or standard scores obtained when the factor scores are zero. Not surprisingly, a conversion of the z-scores into probabilities reveals very low probabilities of obtaining points at the extreme (strongly agree, strongly disagree) ends of the scale. Take the “INFORM” indicator that has an estimated threshold splitting category 1 (strongly disagree to disagree) of  $-1.633$ . This signifies that approximately 5.1% of Americans are predicted to be in the strongly disagree category when they have a factor score of 0.0 on the internal efficacy dimension. For the British group, the z-score is estimated to be  $-1.731$  giving a United Kingdom respondent at this location on the latent Internal dimension a 4.2% probability of being in the lowest category. In establishing this baseline model, we hold the factor means across groups to 0.0—when we allow factor means to vary as we will below, it is important to remember that scores of 0.0 on the latent variables for non-baseline groups may represent respondents at very different points of the latent variable distribution.

### 5.3 Models With Equality Constraints

When examining models with equality constraints, we first compare a very restrictive model to the configural baseline estimated above. Model A restricts all thresholds, factor loadings, and means to be equal across the two groups. As shown in the Appendix, the magnitudes of the parameters do not change dramatically, but the fit of the model is poor ( $\chi^2_{WLSMV} = 1952.04(109df)$  with group contributions United States = 853.27 and UK = 1,098.76). An adjusted 39 degrees of freedom come from the imposition of equality constraints. However, a  $\chi^2$  difference test modified for the WLSMV estimator (Asparouhov, Muthén and Muthén 2006) indicates that the constraints imposed lead to a model fit that is significantly worse than the baseline ( $\chi^2_{WLSMV} Diff = 1,309.54$ ). Close fit statistics for Model A also are poor (RMSEA = 0.09 ( $p < 0.05 = 0.00$ ); CFI = 0.92).

An inspection of the Mplus output Modification Indices (MIs), which constitute empirical suggestions as to the contribution of each model restriction to the overall  $\chi^2_{WLSMV}$  statistic (Byrne 2012), suggests that the equality of group mean restrictions placed on the three latent variables are not tenable. The MIs for the United Kingdom group are as follows: internal

<sup>9</sup>Technically, we only test for partial invariance because of the dual loadings of the PUBOFF indicator in the United Kingdom and NOTSURE indicator in the US. To avoid wordiness, we do not always reference this below. For the testing of latent mean differences and differences in the magnitude of the errors, minor violations of strict invariance usually are judged as permissible (cf. Steenkamp et al. (2010)).

<sup>10</sup>Although the *unstandardized* factor loadings for the UNDERSTAND and NOSAY indicators are fixed to 1.0 and the bottom two thresholds for these indicators are set to equality across groups (as are the lowest thresholds for *all* indicators), the *standardized* loadings of fixed parameters are not equivalent across the groups. This is because standardization takes into account parameters that are not fixed to equality (e.g. the variances of the latent variables).

**TABLE 3 |** Configural model: Factor loadings.

Indicator	United States			United Kingdom		
	Internal	External	Methods	Internal	External	Methods
PUBOFF	0.66 (0.02)	-	-	0.54 (0.02)	-0.21 (0.03)	-
INFORM	0.85 (0.01)	-	-	0.83 (0.01)	-	-
NOTSURE	0.69 (0.02)	-0.22 (0.03)	0.51 (0.02)	0.68 (0.02)	-	0.38 (0.03)
UNDERSTAND	0.86 (0.01)	-	-	0.86 (0.04)	-	-
SELFQUAL	0.82 (0.01)	-	-	0.81 (0.01)	-	-
COMPLEX	0.54 (0.02)	-	0.53 (0.02)	0.58 (0.02)	-	0.51 (0.03)
LEGAL	-	0.76 (0.02)	-	-	0.80 (0.02)	-
FINALSAY	-	0.54 (0.02)	-	-	0.61 (0.02)	-
MAKELSTN	-	0.38 (0.03)	0.56 (0.03)	-	0.47 (0.02)	0.45 (0.03)
NOSAY	-	0.52 (0.03)	0.54 (0.03)	-	0.64 (0.02)	0.51 (0.03)
Correlation						
Internal	1.00 (2.82)	-	-	1.00 (2.86)	-	-
External	0.28 (0.03)	1.00 (0.62)	-	0.08 (0.03)	1.00 (1.26)	-
Methods	0.00 <sup>a</sup>	0.00 <sup>a</sup>	1.00 (0.68)	0.00 <sup>a</sup>	0.00 <sup>a</sup>	1.00 (0.80)

Notes: Standardized factor loadings shown with standard errors in parentheses. All loadings and estimated correlations significant at  $p < 0.01$ .

<sup>a</sup>Correlations between Methods and substantive factors fixed to zero. Unstandardized variances for the factors shown in parentheses in correlation matrix.

efficacy: 510.27; external efficacy: 452.29; and the Methods Factor: 480.24.

Model B retains equality constraints on all factor loadings and thresholds but sacrifices an adjusted four degrees of freedom by allowing the latent means to vary across the two countries. The result is that fit improves dramatically:  $\chi^2_{WLSMV} = 799.73$ ;  $RMSEA = 0.05$  ( $p < 0.05 = 0.06$ );  $CFI = 0.97$ .<sup>11</sup> Following a pattern that persists over various modifications of the model, the standardized factor loadings and thresholds, as reported in the Online Appendix, change little in comparison to the baseline model, but the standardised differences in latent means are large—for internal efficacy, Britons score nearly half a standardized latent variable unit below their American counterparts (-0.49) and the difference across groups for external efficacy is even greater with the United Kingdom averaging 0.58 of a standardized latent unit less.<sup>12</sup>

Parameter estimates provided by Model B with the large number of equality constraints generates a model with approximate fit to the data. However, compared to the modified configural model where latent means are freed, the fit is still worse  $\chi^2_{WLSMV}DIFF = 369.03$  ( $df = 39$  ( $p < 0.00$ )). Inspection of the MIs suggests that the variances of indicators vary, often quite substantially, across groups. As factor loadings and thresholds all are constrained to equality, there are now enough restrictions on the model to allow the ten equality

restrictions on the error variances to be free without causing model under-identification.

Model C frees the ten error variances in the United Kingdom group, and the exact and approximate fit statistics generated by this model are promising:  $\chi^2_{WLSMV} = 520.20$  ( $96df$   $p < 0.00$ );  $RMSEA = 0.04$  ( $p < 0.05 = 0.997$ );  $CFI = 0.98$ .<sup>13</sup> The removal of equality constraints on the error variances of the indicators poses a challenge for difference testing of Model C against configural and modified configural models as they are initially necessary to identify the model. Although Model C is not nested in either of the configural models and therefore not directly comparable, its  $\chi^2_{WLSMV}$  statistic is in the range of the two configural models, and, importantly, it contains 26 and 29 more adjusted degrees of freedom than the configural and modified configural models, respectively. A difference test between the less restrictive Model C and its more restrictive baseline, Model B, demonstrates that the empirical decision to free the error variances of the indicators is valid—that is Model B with the constraints fits the data worse ( $\chi^2_{WLSMV}DIFF = 373.66$  ( $df = 10$  ( $p < 0.00$ ))).

An examination of the remaining MIs from the output of Model C suggests that the fixed paths contributing to large spikes in the  $\chi^2_{WLSMV}$  statistic are those we decided to constrain initially or are fixed for model identification purposes (e.g. correlated error covariances for the indicators, the means of the indicators ( $\alpha_{ig}$ )). The approximate fit statistics are quite good for this model and it is one that constrains all thresholds and, with the exception of the secondary loading of PUBOFF on external efficacy in the United Kingdom and NOTSURE on this dimension in the United States, all of the indicators to be equal in structure and magnitude across groups. Such a large number of constraints on the substantive parameters connecting the factors to indicators allow us to state that latent variable models of internal and external efficacy with the revised indicators are nearly the same in the United States and the United Kingdom. More

<sup>11</sup>The  $\chi^2_{WLSMV}$  contribution from the US group = 421.52 and United Kingdom group = 378.21. A comparison of a modified configural model where all the factor loadings, thresholds, and means vary across groups to a more restrictive one that fixes just the factor means empirically demonstrates equality constraints on the latent variable means are not justified ( $\chi^2_{WLSMV}DIFF = 351.75$  ( $df = 3$ ;  $p < 0.00$ )). The fit statistics for the modified Configural model are:  $\chi^2_{WLSMV} = 480.02$  ( $df = 67$ );  $RMSEA = 0.05$  ( $p < 0.05 = 0.25$ );  $CFI = 0.98$ . The  $\chi^2_{WLSMV}$  contribution from the US group = 279.62 and United Kingdom group = 200.41.

<sup>12</sup>The standardized difference in latent variable means for the Methods Factor is -0.19 ( $p < 0.05$ ).

<sup>13</sup>Contribution to the  $\chi^2_{WLSMV}$ : US = 302.36; UK = 217.84.

**TABLE 4 |** Configural model: Unites States and United Kingdom thresholds.

Indicator and threshold number	Unites States value	Unites States std err	United Kingdom value	United Kingdom std err
PUBOFF-1	-1.524	0.042	-1.673	0.045
PUBOFF-2	-0.927	0.036	-0.777	0.038
PUBOFF-3	-0.178	0.031	-0.018	0.036
PUBOFF-4	0.703	0.034	1.075	0.048
INFORM-1	-1.633	0.051	-1.731	0.058
INFORM-2	-1.094	0.039	-1.109	0.043
INFORM-3	-0.410	0.032	-0.253	0.037
INFORM-4	0.508	0.032	1.167	0.048
NOTSURE-1	-1.559	0.051	-1.774	0.056
NOTSURE-2	-0.885	0.036	-0.579	0.038
NOTSURE-3	-0.203	0.031	0.143	0.036
NOTSURE-4	0.446	0.031	1.124	0.044
UNDERSTAND-1	-1.976	0.064	-1.966	0.060
UNDERSTAND-2	-1.311	0.043	-1.305	0.039
UNDERSTAND-3	-0.604	0.034	-0.306	0.037
UNDERSTAND-4	0.457	0.031	1.139	0.047
SELFQUAL-1	-1.339	0.039	-1.378	0.041
SELFQUAL-2	-0.848	0.035	-0.458	0.037
SELFQUAL-3	-0.059	0.031	0.417	0.037
SELFQUAL-4	0.756	0.033	1.425	0.057
COMPLEX-1	-1.404	0.042	-1.372	0.043
COMPLEX-2	-0.541	0.032	-0.306	0.037
COMPLEX-3	0.050	0.031	0.344	0.036
COMPLEX-4	0.715	0.033	1.222	0.049
LEGAL-1	-1.535	0.044	-1.419	0.048
LEGAL-2	-0.903	0.035	-0.382	0.038
LEGAL-3	-0.145	0.031	0.386	0.038
LEGAL-4	0.994	0.037	1.752	0.064
FINALSAY-1	-0.643	0.026	-0.609	0.026
FINALSAY-2	0.091	0.031	0.304	0.038
FINALSAY-3	0.643	0.033	0.974	0.042
FINALSAY-4	1.269	0.042	1.991	0.070
MAKELSTN-1	-0.922	0.030	-0.950	0.032
MAKELSTN-2	-0.232	0.030	0.192	0.037
MAKELSTN-3	0.221	0.031	0.640	0.037
MAKELSTN-4	0.978	0.037	1.662	0.061
NOSAY-1	-1.070	0.036	-0.927	0.037
NOSAY-2	-0.174	0.025	-0.151	0.022
NOSAY-3	0.275	0.031	0.706	0.036
NOSAY-4	1.066	0.037	1.740	0.055

importantly, these equality constraints, taken together, go well beyond the minimal constraints generally deemed necessary for difference of means testing.

An inspection of the factor loadings from Model C in **Table 5** demonstrates a pattern similar to the initial configural model presented in **Table 3**. Comparing these tables, loadings are generally within 0.02 of a standardized unit from one another. It is important to emphasize that the unstandardized loadings are equivalent across groups, but because we allow other elements of the model to vary freely across groups, the standardized loadings vary.

The bottom half of **Table 5**, however, reveals important aspects to distinguish the groups. The unstandardised variance of the internal efficacy factor (3.41) in the Unites States group is more than 175% greater than what we see in for the United Kingdom group (1.91). On the external efficacy dimension, dispersion in the Unites States (0.72) also is slightly greater than that obtained in the United Kingdom (0.63). As is the case in earlier model iterations, the correlations between the internal and external latent variables

in Model C is much lower than what we see when the traditional efficacy indicators are employed. Similar to the initial models, there is a modest positive correlation between the two substantive dimensions in the US and only a slight association between the two in the United Kingdom. Finally, the standardized mean differences on the two substantive dimensions is large and statistically significant—in a measurement model that is nearly equivalent across groups, Britons come out with levels of internal and external efficacy that are, on average, more than half a standardised unit below their American counterparts.<sup>14</sup>

<sup>14</sup>The mean difference on the Methods Factor just crosses into the realm of statistical significance—We interpret the difference as meaning Britons are slightly less prone than their American counterparts to fall into the “trap” of acquiescence bias. However, the robust and equivalent loadings of these negatively phrased questions on the Methods factor across the two groups leaves little doubt of its presence on representative samples of internet respondents in both nations.

**TABLE 5 |** Final model (C): Factor loadings and variance matrix.

Indicator	United States			United Kingdom		
	Internal	External	Methods	Internal	External	Methods
PUBOFF	0.65 (0.02)	-	-	0.56 (0.02)	-0.21 (0.03)	-
INFORM	0.83 (0.01)	-	-	0.85 (0.01)	-	-
NOTSURE	0.69 (0.02)	-0.21 (0.03)	0.49 (0.02)	0.68 (0.01)	-	0.42 (0.02)
UNDERSTAND	0.88 (0.01)	-	-	0.84 (0.01)	-	-
SELFQUAL	0.83 (0.01)	-	-	0.81 (0.01)	-	-
COMPLEX	0.54 (0.02)	-	0.53 (0.02)	0.57 (0.02)	-	0.49 (0.02)
LEGAL	-	0.79 (0.02)	-	-	0.81 (0.02)	-
FINALSAY	-	0.51 (0.02)	-	-	0.60 (0.02)	-
MAKELSTN	-	0.38 (0.03)	0.55 (0.02)	-	0.48 (0.02)	0.48 (0.02)
NOSAY	-	0.52 (0.03)	0.56 (0.02)	-	0.62 (0.02)	0.45 (0.02)
Correlation						
Internal	1.00 (3.41)	-	-	1.00 (1.91)	-	-
External	0.28 (0.03)	1.00 (0.72)	-	0.08 (0.03)	1.00 (0.63)	-
Methods	0.00 <sup>a</sup>	0.00 <sup>a</sup>	1.00 (0.78)	0.00 <sup>a</sup>	0.00 <sup>a</sup>	1.00 (0.34)

Notes: Standardized factor loadings shown with standard errors in parentheses. All loadings and estimated correlations significant at  $p < 0.01$ .  
<sup>a</sup>Correlations between Methods and substantive factors fixed to zero. Unstandardized variances for the factors shown in parentheses in correlation matrix. Standardized Differences in Means on Latent Variables between United States and United Kingdom: Internal: -0.57 (0.05); External: -0.59 (0.05); Methods: -0.17 (0.07).

The standardized thresholds for Model C, as reported in Table 6 do differ across groups, sometimes to a greater degree than we observe in Table 4. Again, this can be explained by the fact that there are substantial differences in the factor means, variances, and covariances in the United States and United Kingdom cases. In the United Kingdom, the standardized thresholds still allow us to predict the breakdown of responses to the statements we are likely to observe when an individual has a score of 0.0 on the latent Internal and/or External (and/or Methods) dimension(s). However, recall that Model C allows means to vary, and the British, on average, score significantly lower on the substantive latent dimensions. Hence a British respondent with a score(s) of 0.0 on the latent dimension(s) is one that has higher levels of political efficacy than the average United Kingdom respondent.

Table 7 shows the standardized error variances for each of the indicators that this model allows to vary across groups. As is common across the social sciences, these estimates are all statistically significant which signifies the factors do not explain all the variation in responses to the ten indicators. The differences in the magnitudes of the standardized error variances across groups also varies.

Further iterations of the model explore the viability of placing constraints on the covariance between the internal and external efficacy indicators, the variances of the three latent variables and re-establishing the equivalence constraint on the latent variable means for the Methods factor. All restrictions on the variance/covariance matrix are untenable. The equivalence restriction on the average score of the Methods factor produced a difference test that just creeps into the realm making the fit of the model significantly worse at  $p < 0.05$  (but not  $p < 0.01$ ). In multivariate structural models where the efficacy dimensions are employed as predictors or outcome variables, it likely would cause little harm to fix the group means to zero of what is effectively a latent variable that sweeps up the measurement error resulting from placing reverse ordered agree/disagree questions on an internet survey.

## 6 SUBSTANTIVE DISCUSSION

Political efficacy receives pride of place in many models of political participation and, in the case of cross-cultural group or country comparisons, some suggest that efficacy is an important signal of the health of representative democracies or anomie among marginalised groups. Large cross-cultural studies of political and social behaviour such as the Comparative Study of Electoral Systems (CSES) and European Social Survey (ESS) still employ the traditional indicators of political efficacy. In empirical analyses utilizing the techniques described above, cross-national comparisons by Xena (2015) finds such indicators do not perform well.

Work by Richard Niemi, Stephen Craig, and their associates (1990, 1991) argue for a set of indicators they believe to clearly reflect the latent concepts of internal and external efficacy and better distinguish between the two. Subsequent work by Morrell (2003) reviews use of their revised indicators in the literature and conducts his own analyses to attest to their internal and external validity, particularly insofar as internal efficacy is concerned. However there is still much work to be done in the area of cross-cultural validation. A failure to do so leaves open the possibility that we are comparing apples to oranges when discussing how efficacy affects politics and political systems.

The above analyses scratch the surface of cross-nationally validating the indicators in two English speaking longstanding democracies, using multi-group confirmatory factor analysis (MGCFA). Although MGCFA dates back to the work of Jöreskog (1971), the ordinal nature of the response choices for the revised indicators necessitate techniques appropriate for the measurement level of the variables. Building on the work of Millsap and Yun-Tein, (2004), Temme (2006), and Davidov et al. (2011, 2018), Section 4 details the ordinal probit model the widely used software package, Mplus, utilizes to obtain parameters for the MGCFA. Attention is given to the constraints necessary to identify the model, and the importance for considering both the equivalence of factor

**TABLE 6 |** Standardized thresholds to accompany **Table 5**.

Indicator and threshold number	Unites States value	Unites States std err	United Kingdom value	United Kingdom std err
PUBOFF-1	-1.583	0.046	-1.803	0.052
PUBOFF-2	-0.887	0.031	-1.011	0.038
PUBOFF-3	-0.182	0.026	-0.207	0.031
PUBOFF-4	0.731	0.032	0.832	0.035
INFORM-1	-1.633	0.048	-2.214	0.067
INFORM-2	-1.142	0.033	-1.549	0.051
INFORM-3	-0.487	0.027	-0.661	0.041
INFORM-4	0.507	0.03	0.688	0.039
NOTSURE-1	-1.639	0.044	-2.147	0.062
NOTSURE-2	-0.829	0.03	-1.085	0.042
NOTSURE-3	-0.224	0.027	-0.294	0.037
NOTSURE-4	0.475	0.029	0.622	0.037
UNDERSTAND-1	-1.947	0.058	-2.482	0.072
UNDERSTAND-2	-1.355	0.041	-1.727	0.052
UNDERSTAND-3	-0.61	0.03	-0.778	0.041
UNDERSTAND-4	0.483	0.03	0.616	0.037
SELFQUAL-1	-1.381	0.039	-1.795	0.052
SELFQUAL-2	-0.763	0.029	-0.992	0.041
SELFQUAL-3	-0.045	0.027	-0.059	0.036
SELFQUAL-4	0.751	0.032	0.976	0.041
COMPLEX-1	-1.329	0.04	-1.872	0.054
COMPLEX-2	-0.522	0.028	-0.736	0.04
COMPLEX-3	-0.008	0.026	-0.012	0.037
COMPLEX-4	0.648	0.03	0.913	0.041
LEGAL-1	-1.63	0.043	-1.774	0.051
LEGAL-2	-0.844	0.031	-0.918	0.036
LEGAL-3	-0.116	0.028	-0.126	0.031
LEGAL-4	1.047	0.036	1.139	0.044
FINALSAY-1	-0.709	0.027	-0.897	0.035
FINALSAY-2	0.026	0.025	0.033	0.031
FINALSAY-3	0.569	0.029	0.719	0.035
FINALSAY-4	1.28	0.04	1.618	0.055
MAKELSTN-1	-0.962	0.032	-1.261	0.04
MAKELSTN-2	-0.177	0.025	-0.231	0.033
MAKELSTN-3	0.216	0.026	0.283	0.034
MAKELSTN-4	0.983	0.035	1.289	0.047
NOSAY-1	-1.092	0.035	-1.34	0.041
NOSAY-2	-0.302	0.027	-0.371	0.033
NOSAY-3	0.245	0.027	0.301	0.033
NOSAY-4	1.063	0.034	1.306	0.045

loadings and thresholds in a simultaneous fashion. This is an important departure from the separate steps of metric and scalar invariance those conducting MGCFA with continuous indicators employ.

When we conduct analyses with the appropriate techniques for analyzing latent variable models with ordinal indicators, we demonstrate that the alternative indicators perform extremely well, with separate confirmatory factor models for Unites States and United Kingdom data fitting the hypothesized model with the exception of a single separate substantive dual loading in each group (and the need to designate a “methods” factor for reverse coded indicators). We use MGCFA to demonstrate that, save the exception of the dual loading of NOTSURE on external efficacy in the US and PUBOFF on external efficacy in the United Kingdom, the models have statistically equivalent loadings and thresholds on all three latent variables.

The equivalence of the loadings and thresholds allow us to free the latent variable means, and it is clear that Britons have much

lower levels of both internal and external efficacy than do their American counterparts. Americans vary more in their latent levels of political efficacy than do Britons. Finally, in what we believe is a first in the literature, we leverage the equality constraints on the loadings and thresholds to free the error variances on the indicators. Although the indicator variances significantly vary across groups, the standardized estimates presented in **Table 7** suggest that the magnitude of these differences is relatively small. This leads to an important substantive point: Recall from **Tables 1, 2** that the observed distributions of the American responses to the ten indicators vary more than those of the Britons. The equality of loadings and thresholds coupled with the large differences in the latent variances on both substantive and the “Methods” dimensions suggest that the variation we observe is a function of “true” latent variation on the dimensions.

The analyses presented in this paper do not free us from some remaining substantive questions concerning the measurement of

**TABLE 7 |** Standardized indicator error variances to accompany **Table 5**.

Indicator	Unites States value	Unites States standard error	United Kingdom value	United Kingdom standard error
PUBOFF	0.576	0.018	0.668	0.02
INFORM	0.307	0.016	0.286	0.017
NOTSURE	0.321	0.028	0.364	0.022
UNDERSTAND	0.227	0.015	0.297	0.016
SELFQUAL	0.311	0.015	0.349	0.016
COMPLEX	0.422	0.023	0.428	0.023
LEGAL	0.372	0.033	0.35	0.032
FINALSAY	0.745	0.018	0.644	0.021
MAKELSTN	0.544	0.026	0.544	0.022
NOSAY	0.399	0.026	0.415	0.023

political efficacy using the revised indicators. The final model presented in **Table 5** produces standardised estimates for MAKELSTN that are below 0.5 in both groups and the loadings for most of the indicators on the external efficacy dimension are lower than those designated as reflective of the internal efficacy dimension. Previous work gives considerably more attention to the robustness of the latter (see Morrell 2003), and future work may wish to ask whether it is time to pay closer attention to the former. Craig et al. (1990) contend that the four indicators on this dimension appropriately combine citizen beliefs concerning “Incumbent” and “Regime” based efficacy. The modest loadings we observe may end up being a function of trying to combine two concepts that, especially in recent times, are distinct. In both the United Kingdom and United States, support for politicians are at all time lows but support for institutions of government remains comparatively stronger.

Models of internal and external efficacy utilizing the traditional indicators yield correlations between the two dimensions that are sometimes greater than 0.90, and this calls into question the ability of the indicators to distinguish between the two concepts. In sharp contrast, the correlations we report above are much more modest. In the United States, the correlation of 0.28, exactly matches the coefficient obtained by Craig et al. (1990) a quarter century ago. In the United Kingdom, the dimensions are nearly orthogonal. There are competing theoretical arguments as to whether the dimensions should be distinct or interrelated. Coleman and Davis, (1976: 191–193) believe in the close association, noting “[individuals] who believe the system is responsive to people like themselves will be more likely to believe that they personally have the skills to induce government officials to act.” In contrast, Craig and Maggiotto (1982) contend that there is no reason that beliefs about internal “political effectiveness” should be related to attitudes concerning “system responsiveness”. Further theoretical and empirical work is necessary to adjudicate between these rival viewpoints.

We would be remiss if we did not remark on the magnitude of loadings the negatively worded indicators on the Methods Factor, some of which are higher than the substantive loadings of the indicators in question. The agree-disagree statements were put to respondents in a grid based format delivered via an internet survey. Consistent with Kaminska et al. (2010), this finding suggests that a not insignificant number of respondents likely engaged in

“satisficing” or quickly filling out a pattern of answers regardless of question content to move to the next screen. The inclusion of the Methods Factor allows us to “purge” the substantive factors of measurement error likely related to satisficing. However, the need for a “Methods Factor” reinforces the argument that the use of grids in survey questionnaires have tradeoffs—they allow respondents to move through a survey more quickly and this allows more questions to be placed on the survey, but this decision comes at the cost of increased measurement error related to question ordering. On a more positive note, the similarity in the structure of this “nuisance” factor across the two groups suggests that representative samples of respondents in the United States and United Kingdom in approach grid based questionnaires in a similar manner.

These potential problems notwithstanding, Xena’s (2015) finding that the traditional indicators found on cross-national surveys such as the ESS are completely lacking in cross-cultural validity and our analysis that the revised indicators for internal and external efficacy are equivalent across two major English speaking democracies suggest that the modified indicators are a better jumping off points for future (minor) revisions to the battery. Future work should also extend MGCFA to include non-English speaking countries. If, with the addition of additional nations, equivalence holds, the latent dimensions with the revised indicators can be put to use in testing aggregate or multilevel theories concerning the role of nationwide and individual efficacy levels play in a variety of contexts.

## 7 METHODOLOGICAL CONCLUSION: WAS IT WORTH IT?

As this manuscript demonstrates, multi-group analysis where indicator continuity cannot be assumed is a complex process. We hope the above provides a useful guide to the mathematical basis and procedure for multi-group invariance testing with ordinal data. Whether to treat data as ordinal or continuous remains a contentious debate and disciplinary practices vary. Given differences, the results of a single, two-group comparison cannot form the basis for a definitive answer. The procedure chosen may lie with the question the researcher is attempting to answer. Recall that **Table 1** indicates cross-group (UK-US) differences in the intensity of ordinal responses to the

indicators, with United States respondents more likely to indicate that they Agree or Disagree “strongly” to the items. If researcher interest concerns substantive reasons behind the difference in observed preferences, then the only way of attacking the problem would be to treat the indicators as ordinal.

Countering this purposive reason to treat our data as ordinal, Robitzsch (2020) notes that the assumptions underlying the construction of factor scores when the indicators are treated as continuous or ordinal are quite different, and only through examining the covariance of the efficacy factors with known correlates can the researcher evaluate whether treating the indicators as ordinal are superior to more simplistic assumptions of continuity. External validation of the factors is beyond the scope of this article, but we ran supplementary estimations of Models A-C that employ Robust Maximum Likelihood estimators and treat the indicators as continuous. Results suggest that if we put aside the complexities of factor scoring, full structural models employing the efficacy measures as predictors should behave quite similarly. Differences in the latent variable means and variances are minor, and the fit of the models employing MLR estimator are similar. In short, the decision on how to treat the measurement of indicators in multigroup analysis is one driven both by substantive and statistical assumptions and questions that often lie with the researcher and research question at hand.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found below: <https://reshare.ukdataservice.ac.uk/851142/>.

## REFERENCES

- Acock, A. C., and Clarke, H. D. (1990). Alternative Measures of Political Efficacy: Models and Means. *Qual. Quantity* 24 (1), 87–105. doi:10.1007/bf00221386
- Acock, A., Clarke, H. D., and Stewart, M. C. (1985). A New Model for Old Measures: A Covariance Structure Analysis of Political Efficacy. *J. Polit.* 47 (4), 1062–1084. doi:10.2307/2130807
- Asparouhov, T., Muthén, B., and Muthén, B. O. (2006). Robust Chi Square Difference Testing with Mean and Variance Adjusted Test Statistics. *matrix* 1 (5), 1–6.
- Balch, G. I. (1974). Multiple Indicators in Survey Research: The Concept “Sense of Political Efficacy”. *Polit. Methodol.* 1, 1–43.
- Bollen, K. A., and Jackman, R. W. (1989). Democracy, Stability, and Dichotomies. *Am. Sociological Rev.* 54, 612–621. doi:10.2307/2095882
- Brehm, J., and Rahn, W. (1997). Individual-level Evidence for the Causes and Consequences of Social Capital. *Am. J. Polit. Sci.* 41, 999–1023. doi:10.2307/2111684
- Byrne, B. M. (1994). *Structural Equation Modeling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*. Thousand Oaks: Sage.
- Byrne, B. M. (2012). *Structural Equation Modeling with MPlus*. Oxfordshire, England: Routledge.
- Campbell, A., Gurin, G., and Miller, W. E. (1954). *The Voter Decides*. Row, Peterson, and Co.
- Clarke, H. D., and Acock, A. C. (1989). National Elections and Political Attitudes: The Case of Political Efficacy. *Br. J. Polit. Sci.* 19 (4), 551–562. doi:10.1017/s0007123400005639
- Coleman, K. M., and Davis, C. L. (1976). The Structural Context of Politics and Dimensions of Regime Performance. *Comp. Polit. Stud.* 9 (2), 189–206. doi:10.1177/001041407600900203
- Coromina, L., and Bartolomé Peral, E. (2020). Comparing Alignment and Multiple Group CFA for Analysing Political Trust in Europe during the Crisis. *Methodology* 16 (1), 21–40. doi:10.5964/meth.2791
- Craig, S. C., and Maggiotto, M. A. (1982). Measuring Political Efficacy. *Polit. Methodol.* 85–109.
- Craig, S. C., Niemi, R. G., and Silver, G. E. (1990). Political Efficacy and Trust: A Report on the NES Pilot Study Items. *Polit. Behav.* 12 (3), 289–314. doi:10.1007/bf00992337
- Davidov, E., Datler, G., Schmidt, P., and Schwartz, S. H. (2011). Testing the Invariance of Values in the Benelux Countries with the European Social Survey: Accounting for Ordinality. *Cross-Cultural Anal. Methods Appl.*, 149–168.
- Davidov, E., Schmidt, P., Billiet, J., and Meuleman, B. (2018). *Cross-cultural Analysis: Methods and Applications*. Oxfordshire, England: Routledge.
- Davidov, E., Schmidt, P., and Schwartz, S. H. (2008). Bringing Values Back in: The Adequacy of the European Social Survey to Measure Values in 20 Countries. *Public Opin. Q.* 72 (3), 420–445. doi:10.1093/poq/nfn035
- Easton, D., and Dennis, J. (1967). The Child’s Acquisition of Regime Norms: Political Efficacy. *Am. Polit. Sci. Rev.* 61 (1), 25–38. doi:10.2307/1953873
- Finifter, A. W. (1970). Dimensions of Political Alienation. *Am. Polit. Sci. Rev.* 64 (2), 389–410. doi:10.2307/1953840
- Gomez, B. T., and Wilson, J. M. (2006). Rethinking Symbolic Racism: Evidence of Attribution Bias. *J. Polit.* 68 (3), 611–625. doi:10.1111/j.1468-2508.2006.00449.x

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Essex, ethics committee. Survey respondents consisted of those who voluntarily opted into the YouGov panel, and the neither identity of survey respondents nor identifying information was made available to the authors.

## AUTHOR CONTRIBUTIONS

TS wrote the initial version of the article, read up on the technical aspects of multi-group analysis with ordinal variables and ran the analyses in Mplus. TS was a principal investigator on the survey that collected the data used in the analyses presented in this paper. CX conducted extensive research that identified the problem of considerable cross-cultural variance in the political efficacy indicators on the European Social Survey, which motivated the paper and assisted in the writing and editing of the first draft. JR was a co-designer of the survey that collected the data and extensively edited a revised version of the article that constitutes our submission.

## FUNDING

The ESRC funded the surveys that generated the data used in the paper and TS’s time on the project. Funding for this project was provided by ESRC Grant #RES-061-25-0405.

- Hong, S., Malik, M. L., and Lee, M.-K. (2003). Testing Configural, Metric, Scalar, and Latent Mean Invariance across Genders in Sociotropy and Autonomy Using a Non-western Sample. *Educ. Psychol. Meas.* 63 (4), 636–654. doi:10.1177/0013164403251332
- Jöreskog, K. G. (1971). Simultaneous Factor Analysis in Several Populations. *Psychometrika* 36 (4), 409–426. doi:10.1007/bf02291366
- Kaminska, O., McCutcheon, A. L., and Billiet, J. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opin. Q.* 74 (5), 956–984. doi:10.1093/poq/nfq062
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. Vol. 366, New York, NY: Gilford Publications.
- Lane, R. E. (1959). *Political Life: Why People Get Involved in Politics*. Mumbai: Free Press.
- Little, T. D. (1997). Mean and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues. *Multivariate Behav. Res.* 32 (1), 53–76. doi:10.1207/s15327906mbr3201\_3
- Lubke, G. H., and Muthén, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Struct. Equation Model. A Multidisciplinary J.* 11 (4), 514–534. doi:10.1207/s15328007sem1104\_2
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika* 58 (4), 525–543. doi:10.1007/bf02294825
- Meuleman, Bart., and Billiet, Jaak. (2012). Measuring Attitudes toward Immigration in Europe: The Cross-Cultural Validity of the ESS Immigration Scales. *Ask. Res. Methods* 21 (1), 5–29.
- Miller, W. E., and Traugott, S. (1989). *American National Election Studies Data Sourcebook, 1952-1986*. Cambridge: Harvard University Press.
- Millsap, R. E., and Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behav. Res.* 39 (3), 479–515. doi:10.1207/s15327906mbr3903\_4
- Morrell, M. E. (2003). Survey and Experimental Evidence for a Reliable and Valid Measure of Internal Political Efficacy. *Public Opin. Q.* 67 (4), 589–602. doi:10.1086/378965
- Muller, E. N. (1970). Cross-national Dimensions of Political Competence. *Am. Polit. Sci. Rev.* 64 (3), 792–809. doi:10.2307/1953463
- Muthén, L. K., and Muthén, B. (2017). *Mplus User's Guide: Statistical Analysis with Latent Variables, User's Guide*.
- Niemi, R. G., Craig, S. C., and Mattei, F. (1991). Measuring Internal Political Efficacy in the 1988 National Election Study. *Am. Polit. Sci. Rev.* 85, 1407–1413. doi:10.2307/1963953
- Pateman, C. (1970). *Participation and Democratic Theory*. Cambridge: Cambridge University Press.
- Reeskens, T., and Hooghe, M. (2008). Cross-cultural Measurement Equivalence of Generalized Trust. Evidence from the European Social Survey (2002 and 2004). *Soc. Indicators Res.* 85 (3), 515–532. doi:10.1007/s11205-007-9100-z
- Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Front. Educ.* 5, 177. doi:10.3389/educ.2020.589965
- Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *J. Soc. Issues* 50 (4), 19–45. doi:10.1111/j.1540-4560.1994.tb01196.x
- Seddig, D., Maskileyson, D., and Davidov, E. (2020). The Comparability of Measures in the Ageism Module of the Fourth Round of the European Social Survey, 2008-2009. *Surv. Res. Methods* 14, 351–364. doi:10.18148/srm/2020.v14i4.7369
- Steenkamp, J. B. E., De Jong, M. G., and Baumgartner, H. (2010). Socially Desirable Response Tendencies in Survey Research. *J. Mark. Res.* 47 (2), 199–214.
- Stewart, M. C., Kornberg, A., Clarke, H. D., and Acock, A. (1992). Arenas and Attitudes: A Note on Political Efficacy in a Federal System. *J. Polit.* 54 (1), 179–196. doi:10.2307/2131649
- Tarman, C., and Sears, D. O. (2005). The Conceptualization and Measurement of Symbolic Racism. *J. Polit.* 67 (3), 731–761. doi:10.1111/j.1468-2508.2005.00337.x
- Temme, D. (2006). “Assessing Measurement Invariance of Ordinal Indicators in Cross-National Research,” in *International advertising and communication* (Berlin: Springer), 455–472.
- Treier, S., and Jackman, S. (2008). Democracy as a Latent Variable. *Am. J. Polit. Sci.* 52 (1), 201–217. doi:10.1111/j.1540-5907.2007.00308.x
- Xena, C. (2015). “Old Measures, New Implications: The Meaning of Political Efficacy across Political Contexts,” (Colchester: University of Essex, 2015). PhD thesis.

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

Copyright © 2021 Scotto, Xena and Reifler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.