



LETTER

Anchoring Vignettes as a Diagnostic Tool for Cross-National (in)Comparability of Survey Measures: The Case of Voters' Left-Right Self-Placement

Nick Lin¹  and Seonghui Lee² 

¹Institute of Political Science, Academia Sinica, Taipei, Taiwan and ²Department of Government, University of Essex, Colchester, UK

Corresponding author: Nick Lin; Email: nlin@gate.sinica.edu.tw

(Received 19 April 2022; revised 17 January 2023; accepted 26 April 2023; first published online 9 August 2023)

Abstract

There are potentially multiple sources that make it difficult to compare the typical survey measure of the left-right self-placement cross-nationally. We focus on differential item functioning (DIF) due to the different use of response scales when the left-right is framed as an aggregate dimension of policies. We also examine whether and to what extent ordinary citizens' use of the scale is cross-nationally comparable. Our goal is twofold. First, we assess the cross-national comparability of the left-right self-placement scale using the anchoring vignette method used in nine European countries. Second, we propose a measure that quantifies the extent of DIF at the country level. Our original survey and other benchmark studies suggest that the size of cross-national DIF (CN-DIF) in citizens' use of a left-right scale is relatively small when the left-right concept is considered in policy terms and when a comparison is made between Western European countries.

Keywords: anchoring vignettes; differential item functioning; cross-national comparability; left-right self-placement

Political scientists often rely on individual responses to survey questions to empirically capture important theoretical concepts and compare individuals from different groups. The ideological left-right (LR) is one such important concept. This broad, shared summary of complex, political reality emerges, as Benoit and Laver remind us, 'because people over the years have found them simple and effective ways to communicate their perceptions of [the] similarity and difference[s]' between political parties, politicians, and voters (2012, 198). Given this, it is natural that researchers frequently use it to develop and test theories of mass political behaviour. Moreover, the LR metaphor in politics is ubiquitous, from the daily conversation of citizens to the debates among political elites across boundaries and around the globe. Hence, the question about parties, politicians, and voters' positions on the LR scale has been regularly employed in many cross-national surveys (for example: CSES, the European Social Survey, the ISSP, Latinobarometro, the World Values Survey, and other national election surveys). As a result, it is not uncommon to find that the LR self-placement scale is used for making cross-national comparisons.

Indeed, some scholars have directly compared voters' LR self-placement cross-nationally, with an assumption that LR scales are generally an appropriate instrument for cross-national tests (Dassonneville 2021; Freire 2008; Knutsen 1998; Medina 2015; Meyer 2017; Noelle-Neumann 1998). For example, in his study on the stability of ideological orientations of electorates in West European countries, Knutsen (1998, 297) used the population mean of LR self-placements and described 'the mass public in Ireland was the most rightist. [...] followed by Germany,

Belgium [...]. Similarly, Medina (2015) relied on the country mean of LR self-placement scores to describe which electorate was further to the right or left among European countries.

Nevertheless, other scholars have addressed concerns about such cross-national comparisons because of the general potential of Differential Item Functioning (DIF) to the LR concept. For instance, some investigated interpersonal differences in how individuals interpret the left and right metaphor more generally (for example, Bauer et al. 2017; Thorisdottir et al. 2007; Zuell and Scholz 2019); others have focused on developing scaling techniques to make placements of respondents and/or political actors (for example, political parties) more comparable (for example, Lo, Proksch, and Gschwend 2014; Weber 2011). However, despite these efforts, the findings are rather inconclusive with regard to the comparability of the LR self-placement, depending on the sample and the methodological approaches. Moreover, even if we could tell whether the LR self-placement is cross-nationally (in)comparable, little is known about the extent of such cross-national (in)comparability and how to sample a set of countries that are more comparable.

In this note, we join the above literature by focusing on a specific kind of DIF that may make the LR self-placement cross-nationally incomparable – the DIF that occurs when respondents in different countries systematically differ in how they map the underlying continuous scale of an attitudinal variable to be self-rated to its ordinal answer categories. Since we are interested in cross-national (in)comparability (rather than an interpersonal one), we call this problem ‘cross-national DIF’ (CN-DIF) in this paper. In our definition, this type of response category CN-DIF occurs when, for example, citizens in Spain assess a health condition of a woman in her fifties ‘feeling chest pain and getting breathless after walking 200 meters’ as six on an eleven-point healthiness scale, while citizens in France assess the same health condition as three on the same scale. Of course, there could be differences in the use of the scale among individuals; however, our interest is the incomparability problem when the response category is interpreted systematically differently across populations in different countries.

Using anchoring vignettes as a diagnostic tool, we (1) quantify the degree of CN-DIF of a given concept (and the scale that measures it) and (2) identify problematic cases (that is, countries that are relatively incomparable to others) in which respondents use the scale differently from respondents in other countries. We then apply our measure of CN-DIF – $R_{\text{CN-DIF}}$ – to our original surveys in nine European countries as well as to several benchmark studies in political science that utilized anchoring vignettes to assess the cross-national comparability of other important concepts – namely, democracy, political interest, political efficacy, and experts’ assessment of the LR positions of parties.

With our proposed measure and the original survey in which we ask respondents to place several hypothetical parties on the traditional eleven-point LR scale, we find that the LR scale suffers relatively little from the kind of CN-DIF we investigate here (that is, the cross-national difference in the use of the response scale) in so far as the concept is considered in policy terms and the comparison is made between Western European countries. Moreover, our results are in line with previous findings in identifying heterogeneous entities that scholars should be wary of when making comparisons across groups or when determining their grouping strategies. Overall, our work makes a methodological contribution to the broad literature on comparative political behaviour by offering a useful diagnostic tool for survey practitioners, particularly those who are interested in making cross-national comparisons to empirically capture the extent to which a given theoretical concept suffers from CN-DIF, and to identify problematic cases causing greater incomparability within a given sample of countries.

Assessing Cross-National DIF Using Anchoring Vignettes

DIF refers to the problem in which individuals in different groups (in our case, country) provide systematically different answers to survey questions because of artefactual elements in the measurement process. For example, despite the common wording used for LR self-placements

in many surveys, respondents may interpret the question and scale in ways that undermine the cross-national comparability of the resulting measures. In general, there are three major sources of such cross-national incomparability in the use of the LR scale.

First, while many people may think of the LR as an aggregate dimension of various policy domains, other factors such as individual partisanship, long-term values, and social position also play a prominent role in predicting LR self-placements (for example, Dassonneville 2021; Freire 2006; Inglehart and Klingemann 1976; Medina 2015). If the importance of these sources differs across countries, cross-national comparability will suffer. Second, even when they think of the LR in policy terms, respondents in different countries aggregate different kinds of policies into their LR position.¹ Third, even if respondents think of the concept of LR in a similar fashion (for example, the same set of policy areas), respondents in different countries may systematically differ in how they map the underlying scale of the concept to be self-rated to its ordinal answer categories.

We examine this last potential source of cross-national incomparability – the CN-DIF that results from respondents in some countries systematically using the typical response scales differently from respondents in other countries. For instance, when the underlying extent of ‘left-ness’ that causes someone in one country to label herself a ‘4’, this may be the same level that causes a respondent in a different country to label herself a ‘2’. At the same time, we focus the content of the vignettes on the policy contents. Specifically, we use vignettes to explore the extent of CN-DIF resulting from the differential interpretation of answer categories by priming respondents to think of the LR question explicitly in policy terms with respect to a specific set of policies. For example, suppose voters across European countries define the concept of LR using a similar set of policies. In that case, the remaining source of cross-country incomparability of the LR concept will primarily come from the differential interpretation of the answer categories.

Anchoring vignettes is a useful tool to identify and ameliorate DIF caused by differing interpretations of the ‘cut-points’ defining answer categories (King et al. 2004). It does so by utilizing respondents’ assessments of one or more vignettes, which are then used to assess the extent to which (groups of) individuals use the scale differently (identification) and to re-scale their self-assessments relative to where they place the corresponding vignettes (correction).² In this note, we use the anchoring vignette technique primarily for diagnostic purposes rather than corrective to assess the extent to which these data can be used to make reliable comparisons of LR self-placements across countries. Such diagnostic effort is very much in keeping with the scholarly agenda articulated in King et al. (2004), which clearly anticipated (and implicitly encouraged) this kind of diagnostic use by stating that ‘[...] researchers who are confident that their survey questions are already clearly conceptualized, are well measured, and have no DIF now have the first real opportunity to verify empirically these normally implicit but highly consequential assumptions’ (p.205). In our diagnostic use of anchoring vignettes, our primary concern is the DIF that is systematically related to the respondent’s nationality rather than more general forms of DIF among individuals.

Constructing Vignettes for the Placement of Parties on the Left-Right Dimension

Our first task is to construct a set of vignettes that describe different LR ideological positions. The vignettes must be designed in a way that promotes respondents to perceive the scale as

¹Indeed, several studies have asked if voters in different countries use the LR as a summary of the same underlying policies (e.g., de Vries, Hakhverdian, and Lancee 2013; Huber 1989; Knutsen 1997; Thorisdottir et al. 2007). On balance, these studies suggest that, at least in Western democracies, the content of the LR is quite similar.

²Many researchers, such as in studies of subjective health have employed anchoring vignettes to improve the comparability of self-assessed responses across individuals or countries (e.g., Grol-Prokopczyk, Freese, and Hauser 2011; Rice, Robone, and Smith 2012; Salomon, Tandon, and Murray 2004). Likewise, political scientists have offered cross-nationally comparable estimates of party positions by combining anchoring vignettes and scaling techniques (e.g., Bakker, Jolly, and Polk 2022; Struthers, Hare, and Bakker 2020).

unidimensional – in our case, a unidimensional aggregate of a specific set of policy dimensions.³ To achieve this, we focus on four specific sub-dimensions that largely constitute the LR concept, particularly in Western democracies: *regulation of the economy, support for redistribution, the size and scope of government, and attitudes toward cultural diversity*. These policy areas are known to be linked with the LR dimension at both the elite and individual levels across Western European countries (for example, Benoit and Laver 2012; Van der Brug and Van Spanje 2009; Wojcik, Cislak, and Schmidt 2021).⁴ We then vary the level of each sub-dimension monotonically such that the levels of each of these dimensions move in the same direction from the leftist vignette to the rightist vignette. In other words, the leftist vignette describes the most lefty positions on all four policy issues; the rightist vignette describes the rightist positions on all policy areas; and the centrist vignettes describe positions in-between these. By explicitly providing these vignettes *before* asking respondents to place themselves on the same scale, we prime the respondent to think about the LR concept as an aggregator of these dimensions (Hopkins and King 2010). Doing so allows us to concentrate our analysis on the kind of DIF that results from the differential interpretation of ordinal answer categories rather than alternative interpretations of the whole concept.

For several reasons, we construct the vignettes to describe hypothetical parties (rather than hypothetical individuals). As Bauer et al. (2017) empirically demonstrate, citizens often associate political parties with the concepts of left and right. Indeed, political parties play a critical role in making up citizens' LR orientation (Inglehart and Klingemann 1976). Doing so also mimics how typical surveys are designed when asking about respondents' self-placement on the LR scale: many cross-national and national election surveys ask about the respondents' self-placement along with the questions about how they perceive the positions of other political actors, such as political parties, governments, and prominent political figures. Therefore, similar to the vignettes used in Bakker et al. (2014), we create several hypothetical parties as our vignettes to gauge the way citizens use the LR scale. Implicit in this process is to have the respondents make comparisons between the parties described in the vignettes and the respondents themselves, which is exactly the key process expected in the 'anchoring vignettes' approach.

Figure 1 presents the instruction and vignettes used in our survey. The vignettes can be arranged in an ordered, unidimensional scale, ranging from the left-wing party (Party A) to the centre (or centre-right) party (Party B) and the right-wing party (Party C). We randomized the order of the vignettes and asked the respondents to evaluate the LR position of the three hypothetical parties, followed by the respondents rating themselves using an eleven-point scale. Our original cross-national survey was fielded in early 2020 in nine European countries, using internet panels of respondi AG (roughly 2,000 responses per country): France, Germany, Hungary, Italy, the Netherlands, Poland, Spain, Sweden, and the UK.⁵ Given that the broader literature suggests the discrepancies in the meaning of the LR between Western and Eastern European countries (for example, Tavits and Letki 2009; Wojcik, Cislak, and Schmidt 2021), the two Eastern European countries – Poland and Hungary – were included to litmus test the

³The vignettes should be constructed to describe the 'unidimensional' space of LR ideology simply because the target measure, which we are interested in regarding the cross-national comparability, is the unidimensional LR scale ranging from left to right. One may relate this unidimensionality to the discussions on the (issue) dimensionality in political competition in European countries; however, we want to highlight that such discussions are only remotely relevant to this part of the vignette's design. Technically speaking, in the anchoring vignettes approach, even for inherently multi-dimensional concepts, the unidimensionality assumption can still be satisfied by writing the vignettes appropriately: 'For an inherently multi-dimensional concept [...] we need to invoke the same, important sub-dimensions of interest in each vignette and make sure the levels of each of these aspects move monotonically and in the same direction from the low-interest vignette to the high-interest vignette' (Lee, Lin and Stevenson 2015, 211).

⁴Admittedly, the saliency of these issues (or sub-dimension) may differ across countries. However, we have no reason to believe that issue saliency changes how respondents use the LR scale (i.e., the specific type of DIF we examine here). In other words, whether issue saliency contributes to CN-DIF is beyond the scope of our work, and we leave it to future research.

⁵See Online Appendix A for more details about the survey.

Below we present three hypothetical political parties with different policy goals on social-economic issues. Using a scale from 1 to 11, where 1 means left and 11 right, please tell us where would you place the three political parties?

Party A believes in active government regulation of the economy to ensure fairness, health and safety, and to prevent market failures. It always supports redistributive social policies that tax higher income people to fund social programs. It also believes that government should play a key role in addressing most social and economic problems, and that greater cultural diversity is good for the country.



Party B believes in government regulation of the economy to ensure health and safety. It sometimes supports redistributive social policies that tax higher income people to fund social programs. It also believes that government should play some role in addressing some social and economic problems but that others should be left to markets and civil society, and it believes that greater cultural diversity has both good and bad impacts on the country.



Party C believes in minimal government regulation of the economy, only to ensure health and safety. It opposes almost all redistributive social policies that tax higher income people to fund social programs. It also believes that government should play a minimum role in addressing social and economic problems, which should be left to markets and civil society, and it believes that greater cultural diversity is bad for the country.



Figure 1. The vignette assessment question.

performance of our CN-DIF measure, in the expectation that these countries were relatively incomparable to other Western (and Southern) European countries in our sample.

We performed a series of tests that had been carried out in prior research using anchoring vignettes (for example, Bratton 2010; King et al. 2004; Lee, Lin, and Stevenson 2015). We provide the results and relevant discussions in Online Appendix B while summarizing them as follows. First, we evaluated the extent to which respondents perceive the scale of interest as unidimensional – that is, the vignette equivalence test by looking at how many respondents place the vignettes on the same scale in the order we expected. Second, we investigated whether there are systematic differences in vignette placements because even when respondents place vignettes in the same order, people in some countries may systematically shift all vignettes to the left or right of the scale or use only part of the scale. Finally, we compared respondents’ self-placements before and after any DIF was corrected (as described by King et al. 2004). This last test includes a parametric and non-parametric approach to correcting DIF and comparing country rankings before and after correction. The presence of severe CN-DIF would lead to dramatic differences between the raw self-placements and the corrected measures. While these tests help to understand the extent of CN-DIF in the standard measure of the LR self-placement, the answer is not completely straightforward.⁶

⁶For example, to what extent do we accept the ranking changes? Should we only be assured when the rankings remain the same, or do we accept small changes in rankings? There has been no clear standard for such assessments, if not arbitrary.

We thus reanalysed several benchmark studies to compare the results. We chose the benchmark studies because they examined DIF in survey measures of other important concepts in political science – namely, King et al.’s (2004) study on political efficacy; Bratton’s (2010) research on the assessment of democracy; Bakker et al.’s (2014) work on experts’ LR placement of political parties; and Lee, Lin, and Stevenson’s (2015, 2016) studies on political interest. All results from the above-mentioned tests, along with re-analyses of benchmark studies, are reported in Online Appendix B. To summarize, our diagnosis tests suggest that, in general, the LR scale suffers relatively little from CN-DIF and is clearly less problematic than the case of political efficacy in King et al. (2004).

We understand that these diagnoses may not be sufficient to answer whether the extents of CN-DIF we reveal in these analyses are low enough to assure researchers that the LR self-placement is truly cross-nationally comparable. The answers to this question might only be suggestive and relative rather than definite. Nevertheless, there could be ways through which we could at least obtain better insights and have better standards for such an evaluation. In the following section, we propose a measure indicating the extent of CN-DIF and compare our results with other benchmark studies beyond an obvious high CN-DIF case of political efficacy.

Assessing the Extent of Systematic Cross-National Variation in Vignette Placement

Measuring CN-DIF

To approximate CN-DIF, we take a ‘parsing variance’ approach. A typical example of this approach is the estimation of the intraclass correlation coefficient (ICC), which measures the ratio of between-group variance to the total variance. Greater values of ICC indicate that a large portion of the total variance is attributable to between-group differences. In our case, theoretically, the total variance of vignette placements for a sample of individuals grouped by country and vignette can be decomposed into variance between vignettes (on average), the average variance between countries (within vignettes), and the remaining variance across individuals within countries and vignettes. When CN-DIF exists, a large portion of the total variance is expected to come from the variation across countries. We can formally compute these quantities to get a sense of the levels of CN-DIF. A simple way to compute these quantities is to estimate a multi-level model for vignette placements in which

$$y_{kji} = \alpha + u_k + u_j + u_{jk} + e_{jki}, \quad (1)$$

where y_{kji} denotes the placement of vignette j in country k by respondent i . Here, α represents the constant, u_k represents the random intercepts for $k = 1 \dots K$ countries, u_j represents random intercepts for each of $j = 1 \dots J$ vignettes, u_{jk} represents the random intercepts for $J \times K$ vignettes, and e_{jki} represents the residual that captures the random effect on placements attributable to unmeasured factors idiosyncratic to the individual-country-vignette.

Assuming that each of the random effect terms is distributed independently normal with zero means, each will contribute a variance term to the likelihood function and estimates of these variance terms can be used to produce direct measures of the proportion of variance attributable to vignettes vs. countries (as well as estimates in the uncertainty around this proportion). We take the proportion of the variance attributable to the country (as against the total defined by variances attributable to country, vignettes, and country vignettes) as a measure of CN-DIF. The ratio is calculated with:

$$R_{\text{CN-DIF}} = \frac{(\sigma_k^2 + \sigma_{jk}^2)}{(\sigma_j^2 + \sigma_k^2 + \sigma_{jk}^2)} \quad (2)^7$$

⁷This is a form of the intraclass correlation coefficient (Fisher 1954; Koch 1982).

As it is a measure of proportion, the $R_{\text{CN-DIF}}$ ranges from 0 to 1. Suppose the proportion of variance attributable to country is near zero. In that case, this indicates that little of the variation in the vignette placements can be attributed to unmeasured factors relevant to the respondents' nationalities. In contrast, if the proportion is near 1, it suggests that the variation in vignette placements is much more closely related to respondents' nationalities than differences in vignettes. Therefore, we estimate the necessary quantities by using the individual-level data.⁸ Table 1 presents the estimates for $R_{\text{CN-DIF}}$, the proportion of the variance attributable to between-country differences for each of the benchmark studies, and our own data.

Our results from the benchmark studies are largely consistent with what the authors concluded in their works (in the third column). For instance, the proportion of variance attributable to country relative to vignette is around 0.9 in King et al.'s study of political efficacy, which is much greater than what we observe in other studies, and this is in line with King et al.'s conclusion that political efficacy is not comparable across countries, at least between China and Mexico. Meanwhile, the lower levels of variance associated with the country for the works on democracy, experts' LR placement of parties, and political interest indicate that the concepts studied in these works do not suffer serious CN-DIF. Again, this is consistent with what the authors of these studies initially concluded. For instance, Bratton (2010) claims that '[the study's] result provides a preliminary rebuttal against the cynical claim that the original "D-word" formulation is completely incomparable' (p.112).

When it comes to our own study, $R_{\text{CN-DIF}}$ is 0.112. This result on our LR policy measure clearly looks much more like the concepts without significant CN-DIF (for example, political interests and democracy) than the opposite (for example, political efficacy). Moreover, while we have a similar study examining whether LR placements among experts are comparable across countries (Bakker et al. 2014), the results we present here suggest that average voters perform only slightly worse than experts. We take the results as evidence against the idea that CN-DIF might be a serious problem for comparing the LR self-placements across countries. Nevertheless, as an aggregate measure (that is, at the level of a concept), $R_{\text{CN-DIF}}$ is largely determined by what countries are included in a study. In the next section, we measure the country-specific $R_{\text{CN-DIF}}$ to help researchers identify the countries that might be less comparable to others.

Identifying Problematic Cases That Behave Differently From Others

The specific issue of CN-DIF addressed in this study arises from the possibility that respondents from different countries systematically use a response scale differently. If this is the case, we expect to see a greater $R_{\text{CN-DIF}}$ score when grouping countries that are very different from each other to examine a particular concept, compared to the score obtained when studying the same concept based on a group of homogeneous countries. In essence, by comparing $R_{\text{CN-DIF}}$ scores from different groupings of countries, it further enables us to compute a country-specific score that indicates whether a country is suitable for inclusion when comparing a concept of interest across countries. This score could be useful guidance to researchers about the dangers of making comparisons with specific countries.

To calculate a country-specific measure of CN-DIF for each country in each study, we estimate the same random-effect model (as shown in Table 1) for each *pair* of countries included in the dataset. We then average the scores from all country-pairs that include a given country. The resulting score for a specific country is the average variance in vignette placements explained by country (than by vignettes) from all possible pairs of countries that contain a given country. When a country deviates substantially from others in terms of how respondents place vignettes,

⁸Online Appendix C details how we compute these quantities and discusses using individual-level vs. country-level data to estimate CN-DIF. Using country-level data (obtained by collapsing individual-level data) essentially yields the same substantive conclusion. See Figure C1 as a robustness check of what we present in Figure 2.

Table 1. CN-DIF: the proportion of the variance attributable to country

	R_{CN-DIF}	Levels of DIF original study concluded
Political efficacy (King et al. 2004)	0.903	High
Democracy (Bratton 2010)	0.054	Low to medium
Experts' LR Placement (Bakker et al. 2014)	0.011	Low
Political interest (Lee, Lin, and Stevenson 2015)	0.069	Low
Political interest (Lee, Lin, and Stevenson 2016)	0.033	Low
LR Placement	0.112	-

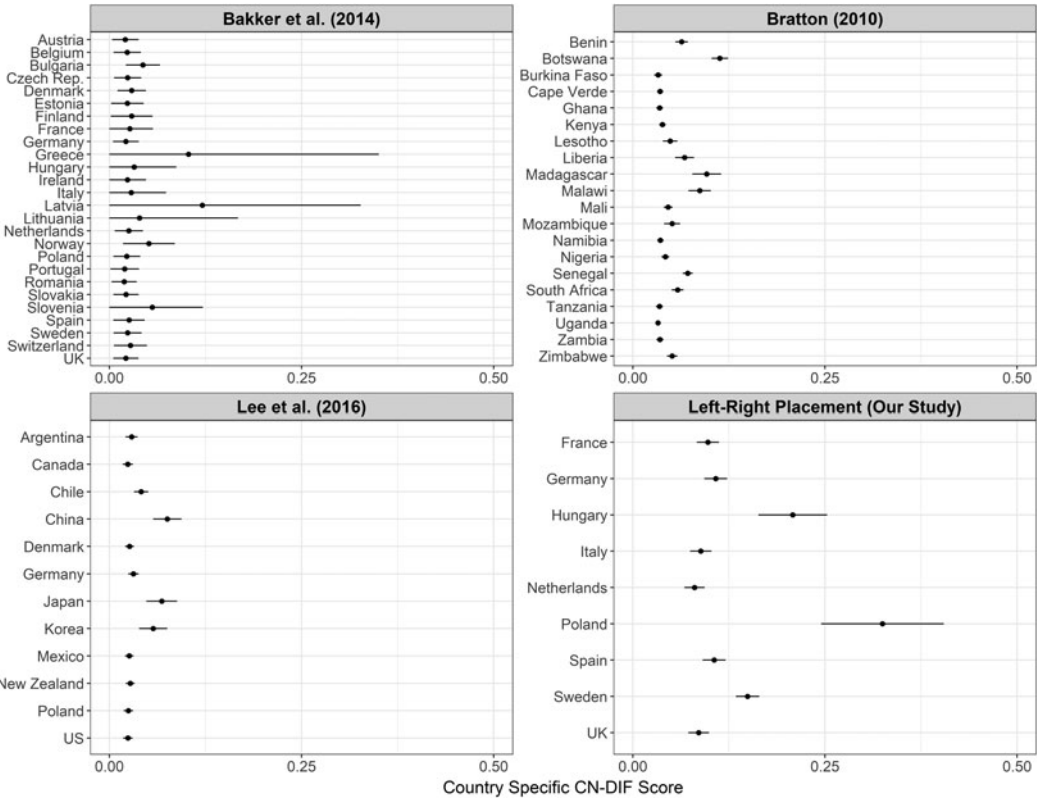


Figure 2. Country-specific CN-DIF scores in four studies.

we expect to observe a greater score. We compute the country-specific CN-DIF scores and 95 per cent confidence intervals by bootstrapping the individual-level data and repeating the estimation process 1,000 times. The results for our study of LR placement, as well as those for Bratton (2010); Bakker et al. (2014); and Lee, Lin, and Stevenson (2016) are illustrated in Fig. 2.⁹

Figure 2 provides some guidance regarding countries that should be included and those that may be better excluded. For instance, the results from Bratton’s data (upper-right panel) suggest that when studying the perception of democracy or similar concepts cross-nationally in Africa, researchers may want to exclude countries like Botswana, Madagascar, and Malawi to ensure that inference or conclusion does not suffer from biases due to potential CN-DIF. Likewise, when comparing political interests cross-nationally (Lee, Lin, and Stevenson 2016, bottom-left

⁹We exclude King et al. (2004) and Lee, Lin, and Stevenson (2015) as they include only a few countries.

panel), researchers may opt against directly comparing the raw reported levels of political interest from Asian countries such as China, Japan, and Korea with those from North American and European countries. The problematic cases identified here largely overlap the ones the benchmark studies pointed out as countries that needed special attention. For instance, Bratton (2010) mentions Botswana and Malawi as countries ‘whose [corrected] ranks change radically’ (p.112). Lee, Lin, and Stevenson (2016) also specify that China, Korea, and Japan are countries with ‘relatively high scores for the low interest case,’ although they contend that there is not enough CN-DIF to substantially undermine their general conclusions.

When it comes to LR party placements made by political experts, our results indicate that there is little CN-DIF (in the sense that there is no country whose score is significantly different from other countries). At the same time, the extremely wide confidence intervals for Greece, Latvia, Lithuania, and Slovenia warrant further attention.¹⁰ Two of these countries – Greece and Latvia – are described as being less comparable in Bakker et al.’s original work (p.5) based on their pair-wise comparison of the mean placements (of vignettes).

Finally, the results regarding citizens’ LR self-placements suggest that we should be wary of comparing Eastern and Western European countries on LR policy ideology. In our data, respondents in Hungary and Poland seem to use the scale of the LR position very differently from their Western European counterparts. These findings could be explained by several empirical features, such as the tendency for leftist parties in post-communist countries to lean economically conservative while rightist parties tend to be more socially liberal (for example, Kitschelt 1992; Tavits and Letki 2009; Vachudova 2008), and the rise of right-wing populist parties in recent years, which has taken over the traditional voting base of leftist parties in Poland and Hungary (Berman and Snegovaya 2019).¹¹ Although it is not a new insight that there are political and cultural differences between Western and Eastern European countries, our CN-DIF measure effectively highlights these differences by providing a quantified description of the extent to which each of the countries in the sample is comparable to others.

Conclusion and Discussion

A general message from this study is that when voters are primed to consider LR in terms of the usual policy debates prevalent in Western societies, they attribute a similar meaning to the ordinal categories on which they are asked to record their responses. That is, a person in Spain seems to think of the meaning of a ‘2’ on this LR policy scale similarly to a person in Germany. While this conclusion may seem narrow in scope, it is directly useful for researchers who are specifically interested in measuring the policy content rather than the general notion of LR. Our study suggests that priming respondents to interpret the LR question in relation to specific policy content can be an effective strategy for creating cross-national comparability of the resulting measures, at least within Western democracies. At the same time, this may provide empirical grounds for previous studies that used a direct comparison of voters’ LR self-placements cross-nationally (for example, Dassonneville 2021; Knutsen 1998; Medina 2015). Beyond this narrow conclusion, this study should also count against blanket critiques of LR self-placement scales that assume voters must use these scales differently across countries.

In addition to our general finding, this work contributes to the field of comparative political behaviour by providing quantitative tools to assess the extent of CN-DIF. Specifically, the measure of CN-DIF we propose, R_{CN-DIF} , can assist future research in identifying the extent of cross-national comparability for a particular concept within a sample of countries. Our country-specific metric further indicates the extent to which a specific country ascribes to the incomparability of the concept

¹⁰See more discussions about this result in Online Appendix D.

¹¹An alternative interpretation for the distinctiveness of Poland and Hungary may be that the ideological range of possible party positions is perceived to be greater in these countries than in others. However, our investigation of the ideological range of respondents’ placement of actual parties suggests this is not the case.

within the sample of countries. As demonstrated, the cross-national comparability of citizens' LR self-placement may suffer when attempting to compare Western European countries to Eastern European ones. While the primary goal of previous works was to reveal the presence of cross-national incomparability in LR placements (for example, Bauer et al. 2017; Lo, Proksch, and Gschwend 2014; Zuell and Scholz 2019), our goal was to quantify the degree of cross-national incomparability and to identify potentially problematic cases using our CN-DIF measure. Finally, our application of $R_{\text{CN-DIF}}$ to benchmark studies found that our quantification of CN-DIF yields similar conclusions to those of the original articles, which employed different analytic approaches. We believe this validates the effectiveness of our proposed CN-DIF measure.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0007123423000236>.

Data availability statement. Replication Data for this article can be found in Harvard Dataverse at: <https://doi.org/10.7910/DVN/LC4QVF>.

Acknowledgements. Both authors contributed equally to this work, and the author's ordering reflects the principle of rotation. Finally, the authors thank Randy Stevenson, Ryan Bakker, and the anonymous reviewers for their thoughtful feedback and suggestions.

Financial support. The authors acknowledge financial support from the German Research Foundation (DFG) via Collaborative Research Center (SFB) 884 'The Political Economy of Reforms' (projects C1) at the University of Mannheim.

Competing interests. The authors declare no conflicts of interest in this research.

References

- Bakker R, Jolly S, and Polk J** (2022) Analyzing the cross-national comparability of party positions on Europe's socio-cultural and EU dimensions. *Political Science Research and Methods* 10(2), 408–18.
- Bakker R et al.** (2014) Anchoring the experts: Using vignettes to compare party ideology across countries. *Research and Politics* 3(1), 1–8.
- Bauer PC et al.** (2017) Is the left-right scale a valid measure of ideology? Individual-level variation in associations with "left" and "right" and left-right self-placement. *Political Behavior* 39(3), 553–83.
- Benoit K and Laver M** (2012) The dimensionality of political space: Epistemological and methodological considerations. *European Union Politics* 13(2), 194–218.
- Berman S and Snegovaya M** (2019) Populism and the decline of social democracy. *Journal of Democracy* 30(3), 5–19.
- Bratton M** (2010) The meanings of democracy: Anchoring the 'D-Word' in Africa. *Journal of Democracy* 21(4), 106–13.
- Dassonneville R** (2021) Change and continuity in the ideological gender gap a longitudinal analysis of left-right self-placement in OECD countries. *European Journal of Political Research* 60(1), 225–38.
- de Vries CE, Hakhverdian A, and Lancee B** (2013) The dynamics of voters' left/right identification: the role of economic and cultural attitudes. *Political Science Research and Methods* 1(2), 223–38.
- Fisher RA** (1954) *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Freire A** (2006) Bringing social identities back in: The social anchors of left-right orientation in Western Europe. *International Political Science Review* 27(4), 359–78.
- Freire A** (2008) Party polarization and citizens' left – right orientations. *Party Politics* 14(2), 189–209.
- Grol-Prokopczyk H, Freese J, and Hauser RM** (2011) Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior* 52(2), 246–61.
- Hopkins DJ and King G** (2010) Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly* 74(2), 201–22.
- Huber JD** (1989) Values and partisanship in left-right orientations: Measuring ideology. *European Journal of Political Research* 17, 599–621.
- Inglehart R and Klingemann H-D** (1976) Party identification, ideological preference and the left-right dimension among Western mass publics. In Budge I, Crewe I and Farlie D (eds), *Party Identification and Beyond Representations of Voting and Party Competition*. Chapter 13. London: Wiley, pp. 243–273.
- King G et al.** (2004) Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98(1), 191–207.
- Kitschelt H** (1992) The formation of party systems in East Central Europe. *Politics and Society* 20(1), 7–50.
- Knutsen O** (1997) The partisan and the value-based component of left-right self-placemen: A comparative study. *International Political Science Review* 18(2), 191–225.

- Knutsen O** (1998) Europeans move towards the center: A comparative longitudinal study of left-right self-placement in Western Europe. *International Journal of Public Opinion Research* **10**(4), 292–316.
- Koch GG** (1982) Intraclass correlation coefficient. In Kotz S and Johnson NL (eds), *Encyclopedia of Statistical Sciences*, Vol. 4. New York: John Wiley & Sons, 213–17.
- Lee S, Lin N, and Stevenson RT** (2015) Evaluating the cross-national comparability of survey measures of political interest using anchoring vignettes. *Electoral Studies* **39**, 205–18.
- Lee S, Lin N, and Stevenson RT** (2016) An expanded empirical evaluation of the cross-national comparability of survey measures of political interest using anchoring vignettes: A research note. *Electoral Studies* **44**, 423–8.
- Lin N and Lee S** (2023) “Replication Data for “Anchoring Vignettes as a Diagnostic Tool for Cross-national (In)comparability of Survey Measures: The Case of Voters’ Left-Right Self-placement””, <https://doi.org/10.7910/DVN/LC4QVF>, Harvard Dataverse, V1.
- Lo J, Proksch S-O, and Gschwend T** (2014) A common left-right scale for voters and parties in Europe. *Political Analysis* **22**(2), 205–23.
- Medina L** (2015) Partisan supply and voters’ positioning on the left–right scale in Europe. *Party Politics* **21**(5), 775–90.
- Meyer AG** (2017) The impact of education on political ideology: Evidence from European compulsory education reforms. *Economics of Education Review* **56**, 9–23.
- Noelle-Neumann E** (1998) A shift from the right to the left as an indicator of value change: A battle for the climate of opinion. *International Journal of Public Opinion Research* **10**(4), 317–34.
- Rice N, Robone S, and Smith PC** (2012) Vignettes and health systems responsiveness in cross-country comparative analyses. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **175**(2), 337–69.
- Salomon JA, Tandon A, and Murray CJL** (2004) Comparability of self-rated health: Cross sectional multi-country survey using anchoring vignettes. *BMJ* **328.7434**, 258.
- Struthers CL, Hare C, and Bakker R** (2020) Bridging the pond: Measuring policy positions in the United States and Europe. *Political Science Research and Methods* **8**(4), 677–91.
- Tavits M and Letki M** (2009) When left is right: Party ideology and policy in post-communist Europe. *American Political Science Review* **103**(4), 555–69.
- Thorisdottir H et al.** (2007) Psychological needs and values underlying left–right political orientation: Cross-national evidence from Eastern and Western Europe. *Public Opinion Quarterly* **71**(2), 175–203.
- Vachudova MA** (2008) Center-right parties and political outcomes in East Central Europe. *Party Politics* **14**(4), 387–405.
- Van Der Brug W and Van Spanje J** (2009) Immigration, Europe and the ‘new’ cultural dimension. *European Journal of Political Research* **48**(3), 309–34.
- Weber W** (2011) Testing for measurement equivalence of individuals’ left-right orientation. *Survey Research Methods* **5**(1), 1–10.
- Wojcik AD, Cislak A, and Schmidt P** (2021) The left is right: Left and right political orientation across Eastern and Western Europe. *The Social Science Journal*, 1–17. doi: 10.1080/03623319.2021.1986320
- Zuelli C and Scholz E** (2019) Construct equivalence of left-right scale placement in a cross-national perspective. *International Journal of Sociology* **49**, 77–95.