

# Public opinion surveys: Evaluating existing measures

*Steven M. Van Hauwaert, Christian H. Schimpf, and Flavio Azevedo*

## Reference

Van Hauwaert, S. M., Schimpf, C. H., & Azevedo, F. (2018). Public opinion surveys: Evaluating existing measures 1. In *The Ideational Approach to Populism* (pp. 128-149). Routledge.

Over the last three decades, populist actors have established themselves as influential political forces around the world (de la Torre 2015). To grasp the phenomenon, the discipline puts a particular focus on populism's conceptualisation (e.g., Albertazzi and McDonnell 2008; Canovan 2005; Mudde 2004; Weyland 2001), as well as the identification of its underlying causes (e.g. Kriesi and Pappas 2015; Mudde 2007) and the evaluation of its consequences (e.g. Mudde and Rovira Kaltwasser 2012; Rooduijn, de Lange, and van der Brug 2014). Until recently, however, most of these studies almost exclusively focused on populism as a supply-side phenomenon.

In contrast, the study of populism as a demand-side phenomenon, particularly in the form of *populist potential amongst individuals*, has received much less attention. Yet, it represents an important component in understanding the rise and fall of populist forces (cf. Hawkins and Rovira Kaltwasser in the Introductory chapter of this book). At this time, we do not (yet) have a comprehensive understanding of its measurement, its empirical evaluation and its implications. Yet, a series of recent studies is making gradual intakes into this field. Most commonly, populism is theorised as an ideational construct, which – as a set of ideas – can manifest itself at the individual level and, therefore, can be measured (Akkerman, Mudde, and Zaslove 2014; Hawkins and Riding 2010; Hawkins, Riding, and Mudde 2012; Stanley 2011).<sup>2</sup> Either implicitly or explicitly, the majority of these studies interpret populism as an attitudinal construct that influences political behaviour. At the same time, comprehensive theoretical and/or measurement studies in support of this rationale remain largely absent. This chapter provides some initial empirical insights into the potential interpretation of populism as an attitudinal construct.<sup>3</sup>

To capture populism at the individual level, scholars have gradually, yet systematically, converged around a battery of survey items (or at least parts of it) that was originally designed by Hawkins and Riding (2010). Even in light of this convergence, scholars have yet to reach a common standard to measure what some refer to as *populist attitudes* (e.g. Akkerman, Mudde, and Zaslove 2014). Oftentimes, the lack of a coherent results in the measurement of populist demand by means of different items and different combinations of items, which in turn makes both within- and between-country comparisons of populism complex, if not impossible. An additional concern that arises pertains to the cross-national validity and reliability of survey items and batteries as a whole, which often remain assumed, rather than empirically tested. In this chapter, we seek to remedy this empirical caveat by (i) providing an initial comparison and overview of instruments to measure populist attitudes in different contexts, and (ii) evaluating the items that measure populist attitudes through psychometric assessment. We test items from different data sets across the Americas (LAPOP, CCES, UCEP and UDP) and Europe (LIVEWHAT, SNES, WoPo) and use item response theory (IRT) models for our empirical analysis.

In this chapter, we proceed as follows. We introduce the reader to the growing literature of individual-level interpretations of populism (such as populist latent attitudes), as well as to the items and scales that scholars have developed. Following this, we discuss why additional empirical tests of the scales and items are essential. We then introduce the data sets and our methodological approach, followed by the results section in which we present our findings for each of the data sets separately. We conclude with a more comparative discussion of our results, which indicate (i) current measures of populist attitudes often fail to provide information about respondents at either extreme (low and high level) of the populism scale, and (ii) we identify instances in which certain items might be redundant because they fail to provide any additional information about a respondent's level of populism. Based on these findings, we discuss the implications for further research in the broader area of populist attitudes. These are the foundation for the ensuing scale development chapter by Silva et al. (Chapter 7 in this book).

### ***Populist attitudes: conceptualisation and importance***

The idea to conceive populism not only as a supply side but also as a demand side phenomenon arguably dates back to Axelrod (1967) who discovers what he refers to as a “weak cleavage” in American public opinion that runs along the lines of “1890s populism” (Axelrod 1967, 51).<sup>4</sup> Few others, however, continued to pursue that path of research in the immediate aftermath (see Dryzek and Berejikian (1993) and Farrell and Laughlin (1976)), despite a diverse set of populist expressions in the twentieth century (ideologies, leaders, parties, movements, policies, etc.). Yet, with an ever more extensive amount of research covering populism over the last two to three decades now, it does not come as a surprise that ultimately, interest in populism as an individual-level phenomenon would be rediscovered. Starting with Hawkins and Riding (2010), Stanley (2011) and Hawkins, Riding, and Mudde (2012), scholars of populism and public opinion began to invest in understanding what they commonly refer to as *populist attitudes*.

In line with one of this book's central premises, most of the studies that examine populist attitudes typically share the conception that there is a set of ideas that underlie the larger notion of populist proclivities. These populist attitudes cannot be observed and measured directly as they are said to be latent (cf. Introductory chapter by Hawkins and Rovira Kaltwasser in this book). Populist attitudes can then be described as a “tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly and Chaiken 1993, 1), where the so-called “entity” is congruent with the ideas of populism. Specifically, these ideas include anti-elitism, the inherent belief in a general or popular will (a *volonté générale*), and a Manichean worldview (Hawkins 2009; Rooduijn 2014).

### ***Measuring populist attitudes***

Based on this common theoretical understanding, initial empirical studies examined the question of how populist attitudes can be measured in the first place. In the process, researchers developed a series of items and scales now available to the wider research community. Hawkins and Riding (2010) were among the first to develop a set of Likert-scale items specifically developed to capture populist attitudes. They developed six of these items based on populism as an ideational construct.<sup>5</sup> Originally, these items were included in the 2008 AmericasBarometer (LAPOP 2008) and fielded in 24 American countries. The authors further refined this measurement by (i) putting the principal focus

on capturing the Manichean view of politics, and (ii) limiting the scale to four items they eventually fielded in the 2008 CCES and UCEP surveys (cf. also Hawkins, Riding, and Mudde 2012).<sup>6</sup> At about the same time, Stanley (2011) designed eight Likert-scale items to measure populism in Slovakia. These accounted for the homogeneity of the people and the elite (two items), the antagonistic nature of political life (two items), the attitudes towards democracy (two items), and the moral dimension of politics (two items). After this, Akkerman, Mudde, and Zaslove (2014) proposed eight items that served as a supposed direct measurement of populist attitudes. To advance the scale, they further built on the original set of items by Hawkins and Riding (2010), but added further items aimed at measuring the Manichean dimension of populism as well as some of its antipodes, such as pluralism (three items) and elitism (three items). They apply their proposed measure of populist attitudes to the Netherlands to also examine whether the demand for populism is met by the supply (Melendez and Rovira Kaltwasser 2017).

### **Empirical inferences based on populist attitudes**

Beyond developing a valid measure to capture populist attitudes, scholars drew on their scales to examine how widespread these attitudes actually are. Most of the previously discussed studies conclude that populist attitudes are prevalent and pervasive among the public (e.g. Akkerman, Mudde, and Zaslove 2014; Hawkins and Riding 2010; Hawkins, Riding, and Mudde 2012; Stanley 2011). Nonetheless, without questioning the general spread of populist attitudes, other studies conclude that there is a fair amount of variance between individuals with regards to their levels of populist attitudes. For instance, in trying to explain the individual differences, Elchardus and Spruyt (2016) find correlations between feelings of relative deprivation and populist attitudes, which – according to them – “is in the first place a consequence of declinism” (p. 15). Further disagreement exists regarding the role and impact of populism. Various scholars have looked at the (potential) influence of populist attitudes on individual voting behaviour. Stanley (2011) finds that populist attitudes are common among the Slovakian electorate but their influence on vote choice is rather limited, in particular in comparison to other factors, such as nationalism or economic preferences. Additional empirical results suggest differently and, in fact, reveal that populist attitudes closely interact with more substantive policy concerns, both for left- and right-wing populist party supporters (Van Hauwaert and van Kessel 2017).

What unites these studies is their objective to complement research on populism as a party-level phenomenon and to understand whether and how demand-side populism has contributed to the pervasiveness of contemporary populism. Based on our review, we can make three central observations. First, one of the more general indications from the studies populist attitudes is the “widespread hypothesis”, that is, the claim that populist attitudes are considerably common among the public. Second, scholars have started to conduct inferential research (particularly regarding vote choice), but until this point, they still draw different conclusions. These differences particularly relate to the role populism plays on the individual level. We suspect, however, that rather than being substantial, these differences may arise from the potential lack of a unified and well-rounded populism measurement. Third, and related to the previous point, we observe that many items used to measure populist attitudes are similar in some respect and have followed Hawkins and Riding (2010) in their formulations. Yet, there exist considerable differences across the items developed and the scales constructed (see also Table 6.1). Given the cross-continental context in which these items now have been used, it becomes increasingly important

to pay particular attention to issues of scale construction, scale development and overall measurement evaluation.

With a wide range of options available to researchers in terms of items and scales (to measure both single ideas of populism, such as anti-elitism, as well as the overall concept), none of these items and scales have thus far been held accountable to (i) whether or not they actually measure populist attitudes, and (ii) if so, to what extent they measure such attitudes (for an exception, see Van Hauwaert, Schimpf, and Azevedo 2017). It is only such an analysis that would allow us to identify the populist items that work best in the field and form the most complete and parsimonious measurement of populist attitudes. A comprehensive assessment of the accuracy and validity of the measurement of populist attitudes, however, is essential to any empirical study that uses these populist items to draw inferences and/or parallels with other political phenomena. An initial test of eight such populist items across nine countries in Europe, for instance, suggests that existing scales can be limited in the extent to which they measure populist attitudes (Van Hauwaert, Schimpf, and Azevedo 2017). Particularly, they can struggle to capture extreme values of populist attitudes, both at the upper and the lower levels of the construct. Additionally, the results of this initial study also suggest that existing populist scales contain several superfluous items that measure almost identical components of populism, while other components are left unmeasured. We argue these results make the critical assessment of the available measures even more necessary.

Table 6.1 provides an extensive overview of the most important survey items scholars have used to measure populism. From this, we can make four important observations. First, almost all measurements include some form of the “good vs. evil” and the “will of the people” questions. These provide an indication of a Manichean worldview and the belief in a general will, respectively. Second, researchers commonly include some kind of reference to the distance between the people and the elite(s), which serves as an indicator of (anti) elitism. Third, many of the questions strike us as context specific, referring to unique institutions or practices. Since we can observe similar formulations across the Americas, but different ones in Europe, we should interpret context as continental (world regions), not national (countries). Last, the majority of items that refer to elites operationalize them as “politicians” or some form of “the legislature” – this despite the fact that theoretical accounts do not typically restrict the meaning of elites to politicians, but also include economic or cultural elites.

Altogether, while populist attitudes and their measurement have gained more attention as of late, most studies limit their attention to substantial rather than measurement concerns. In the absence of proper measurement studies, (i) most existing studies suffer from limited external validity, (ii) most proposed populist measurements remain largely unproven, and (iii) most substantive conclusions drawn from these studies are at risk of being unsubstantiated or even misinformed. This is unfortunate, particularly seeing how the number of different items that propose to measure populist attitudes is increasing, despite a general trend toward the use of some form of the original Hawkins and Riding (2010) items (see Table 6.1). With that in mind, we set out to answer some important questions, namely (i) which items work “best” or are most informative to comprehensively and parsimoniously capture populist attitudes in certain contexts, and (ii) whether different items work equally well independent of context. While we are not able to provide conclusive evidence for both of these questions, we do provide a number of initial insights that will not only contribute to the broader populist measurement debate, but will also lay the foundation for the scale development approach proposed in the following chapter.

Table 6.1 Question wordings throughout the existing literature

	CCES 2008	CCES 2012	UCEP 2008	LAPOP 2008	LAPOP 2010	UDNP 2013	UDP 2015	SNES 2010	WoPo 2011	LIVEWHAT
Distance between people and elite/s (anti-elitism)										
The people should govern directly and not through elected representatives. How much do you agree or disagree?				pop107	pop107					
The biggest obstacle to progress in our country is the dominant class or oligarchy that takes advantage of the people. How much do you agree or disagree with that view?				<b>pop112</b> (1) (2)						
The political differences between the people and the elite are larger than the differences among the people.		bya415				P69c	P41_C		POP3	populism_3
Politicians (elected officials) talk too much and take too little action.	byu324	bya414				P69e	P41_E		POP5	populism_5
Our country would run better if decisions were left up to successful business people.	byu326	bya410								
Our country would run better if decisions were left up to non-elected, independent experts	byu327	bya411				P70a	P42_B			
Ordinary people can't be trusted to make the right choices about our nation's problems.		bya404				P69g	P41_G	a11g		
Politicians should lead the people, not follow them.		bya412				P70b	P42_A			
The particular interests of the political class negatively affect the welfare of the people										populism_7
Politicians always end up agreeing when it comes to protecting their privileges.										populism_8
Most politicians are basically honest people.								a11a		
Ordinary people are divided by very different values.								a11b		
Improvement of ordinary people's life is prevented by elite which is not controlled.								a11c		
Not all the politicians are the same, some really do care what people want.								a11d		
People who belong to political elite are divided by different values.								a11f		
Best government is which decides about everything alone and guarantees order and stability.								a11i		

A Manichean worldview									
Politics is ultimately a struggle between good and evil.	byu320	bya401	pop_good			a11h	POP6		
In today's world there is a struggle between good and evil, and people must choose between one of the two.				pop109 (2)					
What people call "compromise" in politics is really just selling out on one's principles.	byu325				P69f	P41_F	POP7	populism_6	
The inherent belief in a general or popular will (volonté générale)									
The politicians in Congress need to follow the will of the people.	byu321	bya403	pop_foll		P69a	P41_A	POP1	populism_1	
Our presidents/prime ministers must follow the will of the people because what the people want is always right.				pop106 (1) (2)					
Politicians in Congress have to follow the will of society						P41_H			
The power of a few special interests prevents our country from making progress.	byu322	bya405	pop_spec						
The people, not the politicians, should make our most important policy decisions.	byu323	bya407	pop_peop		P69b	P41_B	POP2	populism_2	
Once the people decide what is right, we must prevent opposition from a minority.				pop110 (2)					
Those who disagree with the majority represent a threat to the interests of the country.				pop113	pop113				
I would rather be represented by an ordinary citizen than an experienced politician.		bya408			P69d	P41_D	POP4	populism_4	
Democracy is about achieving compromise among differing viewpoints.		bya402	pop_comp		P70c	P42_C	a11e		
Diversity limits my freedom.		bya409							
Freedom depends on diversity.			pop_div						
It is important to listen to groups with different opinions.		bya413							
In a democracy it is important to hear from all groups					P70d	P41_D			

(continued)

Table 6.1 (continued)

	<i>CCES</i> 2008	<i>CCES</i> 2012	<i>UCEP</i> 2008	<i>LAPOP</i> 2008	<i>LAPOP</i> 2010	<i>UDNP</i> 2013	<i>UDP</i> 2015	<i>SNES</i> 2010	<i>WoPo</i> 2011	<i>LIVEWHAT</i>
When our opposition presents new and challenging positions, there is something we can learn by listening.			pop_list							
Interest groups have too much influence over political decisions.									POP8	
It is necessary for the progress of this country that our presidents/prime ministers limit the voice and vote of opposition parties. How much do you agree or disagree with that view?				pop101 (2)	pop101					
When the Congress hinders the work of our government, our presidents/prime ministers should govern without the Congress. How much do you agree or disagree with that view?				pop102	pop102					
When the Supreme Court/Constitutional Tribunal hinders the work of our government, it should not be paid attention to by our presidents/prime ministers. How much do you agree or disagree with that view?				pop103	pop103					

\* Items to measure populism as a style: pop101, pop102, pop103 (all from LAPOP 2008)

\* Items to measure pluralism: pop\_comp, pop\_div, pop\_list (all from UCEP 2008)

\* Items to measure stealth democracy: byu324, byu325, byu326, byu327 (CCES 2008); bya402, bya404, bya408, bya409, bya410, bya411, bya412, bya413, bya414, bya415 (CCES 2012)

\* Bold text indicates the items used for a populist scale in the literature

(1) Not asked in the USA

## Previous tests, results, and methods

In light of these unanswered questions, we examine and discuss how some of the scales of populist attitudes have been constructed and developed. In a first step, we focus on the methods that have been applied in the studies cited above. We then propose an alternative, and in our opinion more suitable, method to comprehensively analyse the characteristics of both survey items and item batteries that supposedly measure populist attitudes. The majority of studies of populist attitudes rely on Principal Component Analysis (PCA) or – to a lesser extent – Confirmatory Factor Analysis (CFA). However, we propose to use Item Response Theory (IRT) and particularly a Graded Response Model (GRM).<sup>7</sup> In what follows, we briefly discuss why IRT can be helpful when evaluating and developing particular items and scales.

Most studies that examine populist attitudes rely on PCA to determine whether their item batteries, or a subset, measure the same underlying concept (cf. Akkerman, Mudde, and Zaslove 2012; Hawkins 2010; Hawkins, Riding, and Mudde 2012). However, this particular technique serves the purpose of data reduction for subsequent analyses but cannot display how well the individual items and their sum, the scale, capture the latent construct (Bartholomew and Knott 1999). Formulated differently, PCA conceptualises constructs as causally determined by the items, as opposed to the other way around (Comrey 1988; Edwards and Bagozzi 2000). Elchardus and Spruyt (2016), however, apply CFA to their four items. In contrast to PCA, CFA is a commonly applied method for scale development that identifies the underlying dimensions of an instrument (Brown 2006).

In this chapter, however, we propose an IRT framework to comprehensively analyse the item and scale characteristics of proposed populist measurements. For an empirical application, see Van Hauwaert and van Kessel (2018). In general, IRT evaluates the relationship between the responses individuals give to specific items and the underlying latent construct. Particularly, IRT allows us to model the probability of endorsement of a particular response to a survey item, and further assumes this probability is contingent on an individual's level of the latent variable (Hambleton and Jones 1993; Sibley and Houkamau 2013). Compared to CFA (and PCA), IRT has a number of notable advantages. Perhaps most importantly, IRT allows both the precision of measurement (what is typically called, the *information*) and the measurement error to differ across different levels of a latent construct. This means IRT models can specify at which point or range of a latent variable a particular scale provides an accurate measurement and at which point or range it does not. Differently put, IRT has the possibility to capture the scale's accuracy at each level of the latent construct.<sup>8</sup> And while CFA and IRT model techniques can, in fact, be very similar, IRT models provide the additional benefit to determine the contribution of each item in terms of the information that they provide about the latent construct. Thus, the application of IRT models can be particularly useful to maximise the internal validity of a scale, while simultaneously allowing researchers to formulate parsimonious scales. In sum, considering that two important ambitions of this chapter are to identify the usefulness of the plethora of individual items that are currently being proposed as populist measurement and to make strides towards a unified measurement of populist attitudes (cf. also the following chapter), we refer to use IRT over CFA for our analyses.

## Data and methods

One of our principal goals throughout this chapter is to assess different items and instruments used across different contexts with IRT. We use a number of existing data sets



that all include some proposed measurement of populist attitudes. This has two specific advantages. First, such a harmonized approach allows us to cross-validate our results and increase the generalizability of our findings. Second, since our included data sets cover such a broad range of countries across both Europe and the Americas (North and South), it also gives us the opportunity to empirically analyse some of the patterns of variance between and within countries and regions.

We include analyses for the 2012 Cooperative Congressional Election Study (CCES), the 2010 Slovakian National Election Study (SNES), the 2011 Work and Politics (WoPo) survey, the 2014 LIVEWHAT survey and the 2015 *Encuesta Nacional* by the Universidad Diego Portales (UDP).<sup>9</sup> Together, these data sets allow us to cover a broad geographical scope, including Western Europe (LIVEWHAT, WoPo), Eastern Europe (LIVEWHAT, SNES), North America (CCES) and South America (UDP). We take a theory-driven approach and only perform our IRT analyses on the items that have been used as indicators of ideational populism. This means we exclude items that indicate pluralism, stealth democracy or even populism *as a style* from these analyses. For some background and contextual characteristics of the data sets and their data collection, we refer to Table 6.A1 in the Appendix.

## **Empirical results**

An important drawback from harmonising the analysis of such a large number of data sets is that comparisons can become challenging. Particularly, considering the same survey items do not always measure the theorised concepts of *populism* or *populist attitudes*, it becomes difficult to use more integrated methods (for example, Differential Item Functioning (DIF) analysis). Thus, we restrict our analyses to single data sets, providing a more reflective rather than empirically driven discussion about the comparisons of items in the concluding section.<sup>10</sup>

In what follows, we provide readers with a condensed IRT analysis and discussion for each data set. More specifically, as part of each separate analysis, we display two key parts of information that are generally part of the IRT paradigm side-by-side: a test information curve (scale-level) and a set of item information trace lines (item-level). The former allows us to gain insights in the overall informative qualities of a proposed populism scale from a specific data set. The latter provides – for each survey item – an indication of its contribution to the overall information of the populist instrument. Together, these provide detailed and unique insights into different populism scales and indicators.<sup>11</sup>

### ***Measuring populist attitudes in the USA: 2012 CCES***

The 2012 CCES survey includes a core set of four items that are typically used as indicators for populist attitudes. These items refer to the Manichean worldview (1) and to the so-called *volonté générale* (3). From Table 6.1, we see that the 2012 populist indicators are repeated from the 2008 CCES wave and the 2008 UCEP survey.<sup>12</sup> Using this 2008 data, Hawkins and Riding (2010) find a clear populism construct; yet, they also show sizeable correlations between populism and stealth democracy. Even though both CCES waves include various stealth democracy indicators, we are solely interested in how a latent populism variable and its items perform.

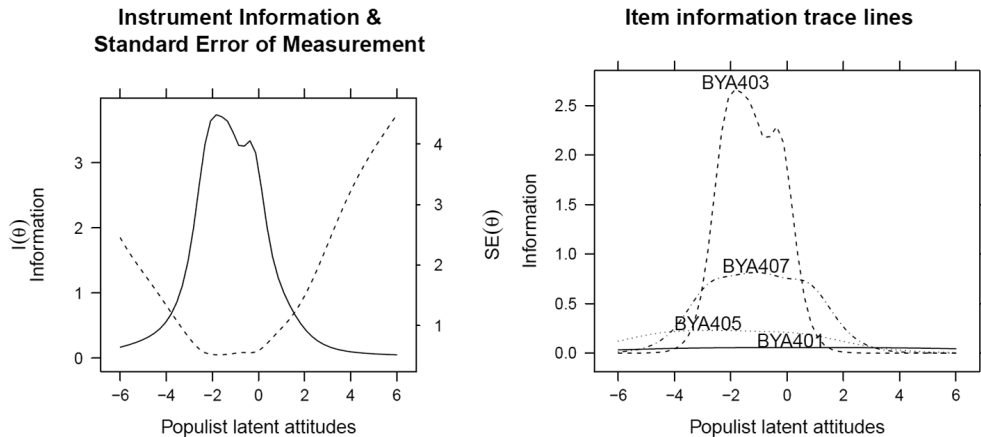


Figure 6.1 CCES test information curve and item information trace lines

Figure 6.1 shows the results of the IRT analysis for the four items that are included in the 2012 CCES wave. The left-side panel of Figure 6.2 indicates that the aggregate scale is only really informative alongside the  $[-2 ; 0]$  interval, meaning the scale only appropriately discriminates those who are moderately averse to populism. Furthermore, the scale is clearly not able to discriminate individuals with more extreme populist attitudes, at either extreme of the scale. When we examine the individual item contributions, we must conclude that three out of four items are limited in the information they provide. Only item “BYA403” (will of the people) is informative. Even more, the similarity between the information curve on the left and the individual trace line of item “BYA403” stands out. While this is not conclusive evidence, it raises questions about the necessity of the other CCES items. Put more extremely, perhaps rather than using these four items as a scale, scholars could just rely on the single item as a proxy of populism – at least in this specific context.

### ***Measuring populist attitudes in Slovakia: 2010 SNES***

The 2010 SNES survey includes eight items that are designed to measure populism, and are loosely based on Hawkins and Riding’s (2010) operationalisation of populism. Unlike other studies, Stanley (2011) does not generate a scale but individually models the items and argues that each item serves as a proxy for a specific component of populism. But, would a scale using these items provide a good measurement of populist attitudes? A number of items are unique, yet three items find overlap with other surveys, namely a11h (good vs. evil), a11e (democracy) and a11g (trust people).

Figure 6.2 shows a relatively wide test information curve, providing high levels of information in the  $[-1; 3]$  interval, i.e. the items capture a relatively broad range of the latent variable. Furthermore, we can also observe that the SNES scale would have difficulties accurately measuring those at the lower end of the scale (i.e. low levels of populist

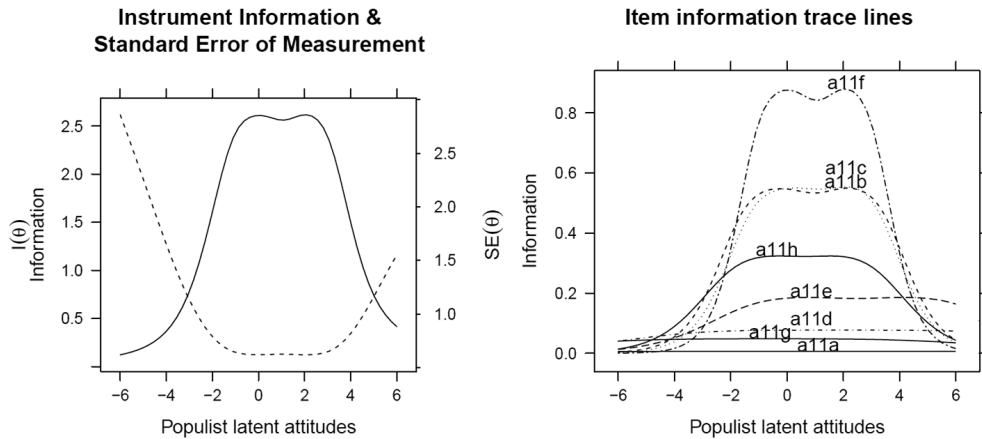


Figure 6.2 2010 SNES test information curve and item information trace lines

attitudes), as indicated by the left tail of the error measurement curve. At the higher end of the scale, the measurement does not appear to have such a problem, or at least not to the same extent.

To some extent, this observation is also reflected in the individual item information lines. Seeing how they are all located right of the middle, the items in the Slovakian sample appear better suited to measure higher values of populism than they do lower values. When we compare the curves with one another, we can divide them in three arbitrary groups. A first group provides little to no information in the measurement of populism (five items). A second group provides good levels of information and measure a relatively broad scope of our construct. This includes the variables a11h (good vs. evil), a11b (ordinary people) and a11c (ordinary people). Third, item a11f (elite values) clearly stands out in that it provides – by far – most information out of all items. From a theoretical IRT perspective, we would expect some items to have a higher peak than others; however, lower peaked items should then cover a broader scope of the latent construct in order to be equally valuable (i.e. provide the same information). As we can observe, this is clearly not the case for the less informative SNES items.

Furthermore, our empirical analysis allows for an important theoretical and more substantive insight. That is, the items that are generally most informative also broadly capture the three key attributes of populism, namely anti-elitism, a Manichean divide and an affinity to the common will. While we remain cautious in drawing any comparative conclusions, it indicates this scale allows for a particularly well-matched connection between theory and empirics – at least in this particular context. As we will reflect on other data sets, it will become clear this is certainly not always the case.

### ***Measuring populist attitudes in the Netherlands: 2011 WoPo***

The items, used by Akkerman, Mudde, and Zaslove (2014), were part of an online survey that was fielded in the Netherlands in November 2011. Regarding the populism items, this survey relies directly on the original work by Hawkins and Riding (2010), but it also adds to it. Namely, it includes two fully novel items (POP3 and POP4) and two items suggested by Hawkins and Riding (2010) for the purposes of future research (POP5

and POP7). Akkerman, Mudde, and Zaslove (2014) identify six relevant items for their proposed populism scale (POP1–5, POP7).<sup>13</sup> Together, these six items operationalise the three central aspects of populism [see Table 6.1] and the scale has quickly developed into an important reference point for subsequent attempts to measure populism.

Figure 6.3 shows that the 6-item scale that was used in Akkerman, Mudde, and Zaslove (2014) covers a relatively broad range of the information of the latent populism scale  $[-3: 2]$ . While it is not perfect as to the information it provides, we can argue the scale is at least informative in the Dutch context, and perhaps even more informative than some of its counterparts discussed below. Much like other scales evaluated in this chapter, the left-hand panel of Figure 6.3 indicates the scale measures respondents better at the lower end of the scale, compared to the upper end of the scale. Furthermore, there is a subset of items the works better in comparison to the others. The right-hand panel of Figure 6.3 indicates that items 3, 5, and 7 stand out as the most informative. Interestingly, whereas item 5 is the most informative, it captures slightly less of the range of the latent populism scale with all other items better equipped to measure high levels of populism. Out of all the items, Item 1, which aims at measuring the extent to which respondents agree with the statement that the politicians should follow the will of people, is the least informative. Interestingly, similar items turned out to be more informative relative to all other items in other country contexts (e.g. UCEP 2008 and UDP 2015). This result could indicate that, after all, populism varies in its meaning by context, which would have important consequences for its measurement and some of the generalisations drawn from this survey.

### *Measuring populist attitudes across Europe: LIVEWHAT*

The LIVEWHAT survey provides a cross-national measurement of populist attitudes across nine European countries (France, Germany, Greece, Italy, Poland, Spain, Sweden, Switzerland, and the UK). Its populist items are largely drawn from previous studies by Hawkins and Riding (2010) and Akkerman, Mudde, and Zaslove (2014) and have been used in comparative empirical studies such as Van Hauwaert and van Kessel (2018). They data set also offers two unique items that arguably measure European populist attitudes (populism\_7 and populism\_8).

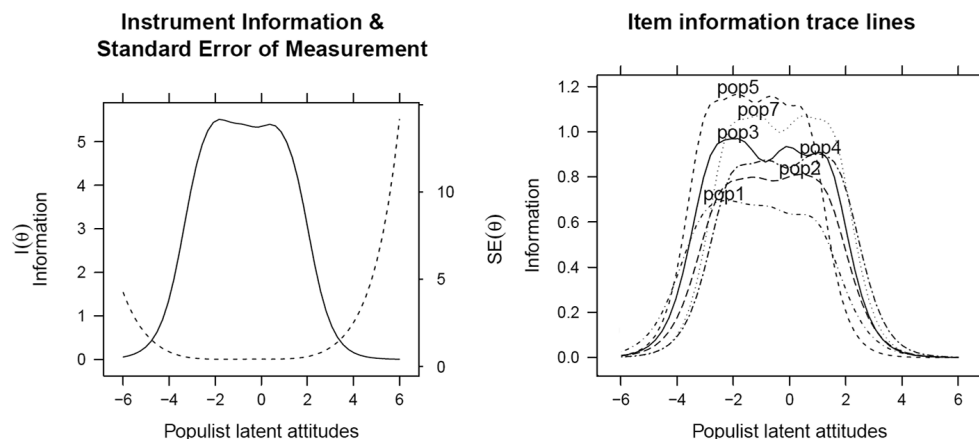


Figure 6.3 2011 WoPo test information curve and item information trace lines

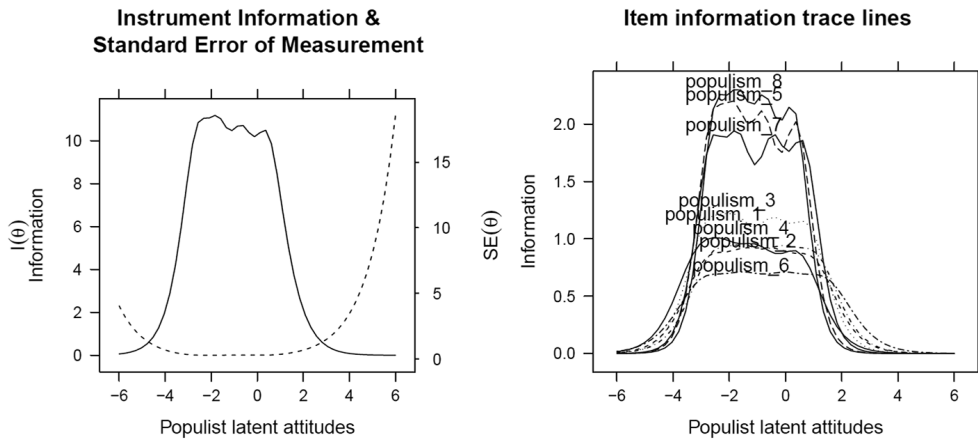


Figure 6.4 LIVEWHAT test information curve and item information trace lines

The LIVEWHAT test information curve indicates a  $[-3; 1]$  interval that can be considered to have high information. This means that within this particular interval of our latent construct, the proposed populism scale from these eight items is most reliable to measure populism. The steep increase of the measurement error curve on the right side of the graph indicates the LIVEWHAT scale has particular problems measuring high levels of populist attitudes, or capturing those with a high affinity towards populism. While this problem also appears to be somewhat present when levels of populist attitudes are very low, this is true to a lesser extent.

Much like the IRT-estimated populism scale in 2010 SNES survey, there are three populist items that stand out when it comes to providing information, namely populism\_8 (elite action), populism\_5 (elite interests) and populism\_7 (elite privileges). Two of those are unique items, while populism\_5 is a recurring item from the CCES and UDP surveys [cf. supra]. Contrary to – for example – the SNES scale, we observe that our three most informative items are most successful in measuring a particular theoretical component of populism, namely anti-elitism. The informative contributions of the other items remain both moderate and nearly identical. Not only do all items cover a similar scope of the latent construct (i.e. they all cover the same range of populist attitudes), but a number of them provide similar amounts of information as well. This could be a first indicator of a number of superfluous items in the proposed scale and a lack of parsimony in this particular eight-item scale. Interestingly, populism\_6 and populism\_3 are not that informative in a broader European context, whereas the same items proved to be quite informative in the Dutch context (2011 WoPo – POP3 and POP7, respectively). The item “politicians (or elected officials) talk too much and take too little action”, however, does appear to be quite informative both across Europe (populism\_5) and in the Netherlands (POP5).

### ***Measuring populist attitudes in Chile: 2015 UDP***

The 2015 UDP data set includes six items that are theorised to tap into the notion of populist attitudes (see e.g. Hawkins et al. 2016; Mélenz and Rovira Kaltwasser 2017). Looking at the items, we observe that the designed measurement of populist attitudes

includes one of the most common anchors, namely the “will of the people” item. Additionally, we find important overlap between the UDP items and the LIVEWHAT (five items), as well as the 2012 CCES (two items). For researchers interested in populism and populist attitudes, Chile provides an interesting case because it has now fielded three different surveys waves that include these items.

As a whole, Figure 6.5 illustrates that an IRT estimated populism scale in Chile is most informative in the  $[-1; 3]$  interval. While comparisons across data sets are difficult, this would constitute the broadest “informative range” of our included data sets. However, the high-peaked information curve indicates this particular scale has difficulty measuring extreme values of populist attitudes. Likewise, we do observe our test information curve is left-skewed, meaning it does better (albeit slightly) at measuring an affinity toward populism than an aversion to populism.

Figure 6.5 further illustrates that the individual information levels amongst the six different items differ considerably. Again, comparison across data sets is difficult, but this particular scale does not include completely uninformative items and the overlap between items appears more limited, indicating different items capture unique parts of the latent construct. Based on the amount of information provided by each item, item P41\_A (will of the people) is the most informative item. This is perhaps not surprising, considering it is typically one of the core indicators of populist attitudes. Furthermore, a number of items, particularly those with lower levels of information (P41\_F and P41\_D), cover a larger scope of the populist construct (i.e. they have wider tails). This contributes to their uniqueness and thereby also usefulness as part of the scale.

## Conclusions and comparative reflections

Over the past years, populism research has grown exponentially and its scope continues to broaden. Yet, at the same time, several aspects of the study of populism remain

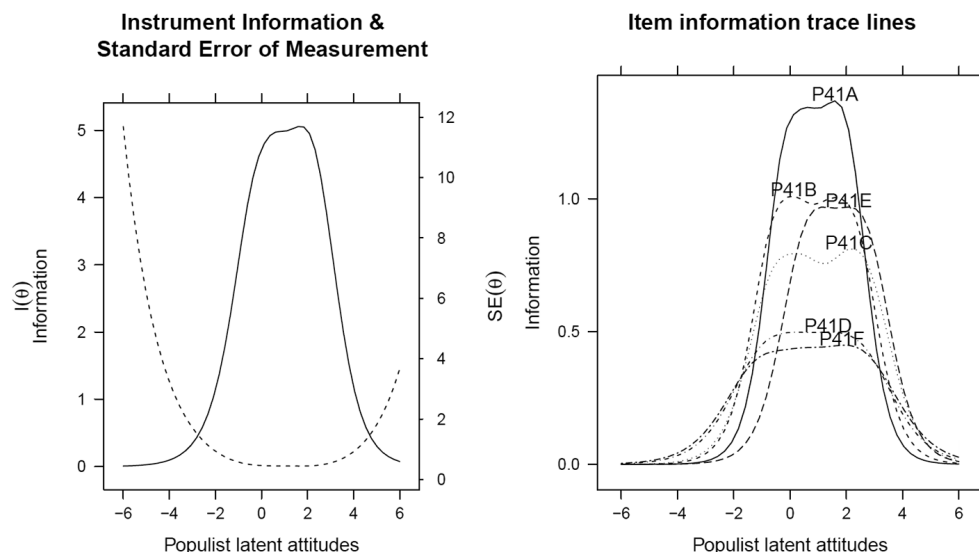


Figure 6.5 2015 UDP test information curve and item information trace lines

underdeveloped, most notably the study of its demand-side and its translation as a latent attitudinal construct. While scholars primarily explore how widespread populist attitudes are and what consequences these attitudes have for political behaviour, we know very little about the intrinsic characteristics of populist attitudes. While measurement studies should precede more substantive research questions, the accurate, parsimonious and comprehensive measurement of populist attitudes is currently still in its infancy. Consequently, much of the research using populist attitudes draw conclusions from a relatively unexplored measurement tool.

With that in mind, this chapter sought to shed some light on several of the measurements (instruments) of populist attitudes the literature uses or proposes, exactly to avoid any misinformed or incorrect conclusions. In an effort to offer a comprehensive empirical analysis of existing populist scales, we use multiple data sets from different contexts in Europe and across the Americas and we propose a methodological paradigm (IRT) that provides us with a number of unique possibilities. We examined how scales as a whole fared in these different contexts, but we also examined single items within these scales to assess the levels of information each survey item contributes to the overall populism scale. Particularly this latter aspect is something current literature often foregoes, but can be extremely informative for scale development, item repetition and general survey design.

On the basis of our results, we can draw several preliminary conclusions and provide an initial comparative reflection on the interpretation of populism as an attitude.<sup>14</sup> First, a notable majority of the populism scales are limited in their ability to measure populist attitudes at the extremes, and this both for very low and very high levels of populism. Concretely, this means that regardless of the scale or the context, most proposed populist measurements have difficulty to accurately capture or discriminate against those individuals who are either quite averse to populism or quite prone to it. There are, however, some exceptions to this observation, particularly the items used by Stanley (2011) as part of the 2010 SNES data set. His scale differs from others in that it provides higher levels of information for those respondents who are located at the higher end of the latent populism construct. In its specific context (Slovakia), we thus suggest this particular set of items fares better in comparison to other scales in different contexts. That being said, without further empirical analyses and more in-depth scale comparisons based on anchor items (so-called *equating*), this conclusion is preliminary and must remain subject to further empirical scrutiny.

Second, we also observed that across most scales, there exists great heterogeneity with regard to the information provided by individual items. Even though item inclusion and formulation should be a theoretical process, careful empirical analysis teaches us there is certainly some kind of disconnect between theory and empirics. This may result from two possibilities that, again, have to be considered within the specific context in which the respective surveys have been fielded. On the one hand, it is quite possible that items with low levels of information are not formulated well enough to properly measure the latent construct. Upon analysis, one must question the contribution of such items throughout the scale development process. Considering there are quite a few items like this included in the proposed populism scales, we would urge future scale development efforts to account for this. On the other hand, even though there is currently little evidence that populism is a context-specific construct, the role of specific dimensions of populist attitudes (e.g. the will of the people) differs or can even depend on the context (e.g. country) certain items are

fielded in. A brief cross-country measurement invariance test confirms this (see Van Hauwaert and van Kessel 2018), which in itself hints at some heterogeneity in the measurement of populist attitudes. Combined, these two preliminary observations indicate that certain items can be considered a better fit than others.

Much in line with this last point, we must also consider how well a proposed measurement scale covers the concept of populism. Perhaps the most striking similarity between the graphs throughout this chapter is the relative inability of proposed populism scales to measure the populist construct from one extreme to the other. Part of this can be attributed to what we have referred to as low-information and identical items being included in scales [cf. *infra*]. However, a more theoretically driven explanation of this could be the construct's primary focus on certain aspects of populism, rather than on all (or at least multiple) of its dimensions. While our IRT analysis provides more detailed insights into this argument, this is already clear from Table 6.1. Most items indicate anti-elitism and the *volonté générale*, while only few items (often with relatively limited information) gauge the Manichean component of populism.

These observations have ample implications for researchers who intend to use any of the data sets for their own analyses or for those who seek to field their own survey and are looking for existing scales or individual items to draw from. For empirical analyses, researchers should take into account that their populist scale can be contextually limited and is unlikely to measure the full range of the latent construct within the context of analysis. While this is – by itself – not necessarily problematic, it does limit the inferences one can draw from the analyses, particularly when looking to generalise findings beyond the specific context. It is something we encounter all too often, particularly with studies that take the USA or the Netherlands as their primary focus.

For those who initiate their own research, different pitfalls must be carefully considered. The verbatim duplication of certain scales into different context does not necessarily guarantee accurate measurement of populist attitudes and thus, ideally, should be preceded by extensive pre-tests if possible. Furthermore, even when relying on only the most informative items, it may result in identical information across respondents. In other words, since we indicated most populist items (and really all populist scales) have difficulties measuring extremes, it is likely that even “good” (highly informative) items provide similar information levels across a similar, yet limited range of the latent construct. Therefore, when developing items or scales, one might consider adjusting the (number of) items based on their informative contributions.

Finally, we repeat that the research on populist attitudes is still in its infancy. As such, we need additional (exploratory) studies that test new (sets of) items or new item formation to find items and scales that are more informative and cover a larger scope of the latent construct than those we have tested here. Further, this would allow us to test for items that can be used in more than one context. For these types of studies, we encourage researchers to also include at least two (preferably even more) of the items that have been used in the above-cited studies as anchor items. In other words, we would recommend future studies with “new” (and better) items to measure populism should include a number of “old” items. This makes newly developed and tested scales comparable (in terms of validity) to existing ones and allows for a better evaluation of scales and their often contextual insights. In line with this suggestion, Silva et al. (Chapter 7 in this book) take such an explorative approach as they set out to analyse how a wide variety of potential populist items measures the phenomenon across different contexts.



## Notes

- 1 An earlier version of this chapter has been presented at the 2016 ECPR General Conference in Prague. We would like to thank Joost van Spanje and Robert Huber for their feedback on previous version of this chapter. We recognize all those who have generously contributed data to this chapter: Ben Stanley, Olga Gyarfassova, Agnes Akkerman, Cristóbal Rovira Kaltwasser, and Kirk Hawkins. Christian H. Schimpf acknowledges the support by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation. Flavio Azevedo acknowledges the support by the University of Cologne's Center for Comparative Politics and the Cologne Graduate School. We would also like to thank the editors of this book for their combined efforts to continuously improve this chapter.
- 2 For a more detailed account of the ideational interpretation of populism, as well as its implications for the study of populism, we refer to Hawkins (2009; 2010), Mudde and Rovira Kaltwasser (2013) and Rovira Kaltwasser (2014).
- 3 We recognize that the interpretation of populism as an attitudinal construct, rather than just an opinion, remains contested in the recent literature. We argue, however that (i) this (theoretical) debate does not fall within the scope of this chapter, and (ii) the different terminology – while one might agree or disagree – does not make a difference for the substantial conclusions we intend to draw in this chapter. For a more detailed discussion of this contentious formulation, we refer to some of our work elsewhere (e.g. Van Hauwaert, Schimpf, and Azevedo 2017). For a more general discussion of the distinction between attitudes and opinions, we refer to Bergman (1998) and Oskamp and Schultz (2014).
- 4 The literature has also advanced some less complete measurements of populism. Elchardus and Spruyt (2016) used only four items that measure the people-centrist component (two items) and to some extent, the anti-elitist view of populism (two items). Ford, Goodwin, and Cutts (2012) limit the scope of their populism measurement to popular hostilities and discontent with major parties.
- 5 The other three items were constructed differently in order to capture populism as a “political-institutional style” (cf. also Barr 2009; Roberts 2003; Weyland 2001).
- 6 While reducing their populism measurement, the authors also complement the populism scale in two particular ways. In the 2008 UCEP survey they include three items that serve as indicators of pluralism, a concept closely related to populism. In the 2008 CCES survey, they include four items that serve as indicators of stealth democracy, a concept that has been found to encompass populism (cf. Hibbing and Theiss-Moore 2002).
- 7 For a more detailed discussion of IRT in the field of political science, we refer to Jackman (2008). For a more detailed analysis of the GRM, we refer to Samejima (1997). We rely on a GRM because of the ordered polytomous data structure of most populist items (cf. Samejima 1997).
- 8 Whereas most of the political science literature refers to reliability, the IRT literature uses the concept of information to refer to the measurement precision of an instrument. Altogether, reliability can be thought of as a simplified version of IRT's information (Wainer and Thissen 1996).
- 9 In the appendix (Figures 6.A1 to 6.A5), we also include empirical analyses of the 2013 UDNP survey in Chile, the 2008 LAPOP survey in both Latin America and the USA, the 2008 UCEP survey in Utah and the 2008 CCES survey in the USA.
- 10 As part of the IRT paradigm, a more comparative approach or so-called *equating*, that is the comparison of scales based on similar items, may be possible in some instances, but certainly not for all data sets here. In order to apply equating correctly, the different data sets would require a number of stable anchor items between them.
- 11 Table 6.A1 in the Appendix provides information about the scale-level information of the populist measurements from the different data sources included in this chapter.
- 12 The primary difference between the populist items in the UCEP and the CCES surveys is that the UCEP survey relies on 4-point answer scales, whereas the UCEP survey gives respondents 5-point Likert-type scales with labels at the extremes. See also Table 6.1.
- 13 In their survey, Akkerman, Mudde, and Zaslove (2014) added two additional items. The wordings of the items were a) “Politics is ultimately a struggle between good and evil” and b) “Interest groups have too much influence over political decisions.” However, the first item was found to

relate to a separate elitism scale, rather than the populism scale. The second item was found not to be related to any of the proposed theoretical constructs and thus, Akkerman, Mudde, and Zaslove (2014, 1334) eliminated it from their analyses. For reasons of comparability, we therefore focus our analyses on the six items that were used to construct the final populism scale in the original paper.

14 Note that all comparative insights should be taken with the appropriate grain of salt. As a consequence of the large variety of populism items and the subsequent absence of an appropriate set of anchor items, we were not able to compare all scales in a single harmonized analysis. In other words, our conclusions are based on the assumption that despite operational differences, all scales essentially measure the same construct.

## References

- Akkerman, Agnes, Cas Mudde, and Andrej Zaslove. 2014. "How Populist Are the People? Measuring Populist Attitudes in Voters." *Comparative Political Studies* 47(9): 1324–1353.
- Albertazzi, Daniele, and Duncan McDonnell, eds. 2008. *Twenty-First Century Populism: The Spectre of Western European Democracy*. Houndmills and Basingstoke and Hampshire [England] and New York: Palgrave Macmillan.
- Axelrod, Robert. 1967. "The Structure of Public Opinion on Policy Issues." *Public Opinion Quarterly* 31(1): 51–60.
- Barr, Robert R. 2009. "Populists, Outsiders and Anti-Establishment Politics." *Party Politics* 15(1): 29–48.
- Bartholomew, David J., and Martin Knott. 1999. *Latent Variable Models and Factor Analysis*. 3rd edn. (Wiley series in probability and statistics). London: Arnold. Available at: <http://site.ebrary.com/lib/academiccompleteitles/home.action>.
- Bergman, Manfred Max. 1998. "A Theoretical Note on the Differences between Attitudes, Opinion, and Values", *Revue Suisse de Science Politique*, 4(2): 81–93
- Brown, T.A. 2006. *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Canovan, Margaret. 2005. *The People*. (Key concepts). Cambridge: Polity.
- Comrey, Andrew L. 1988. "Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology." *Journal of Consulting and Clinical Psychology* 56(5): 754–61.
- De La Torre, Carlos, ed. 2015. *The Promise and Perils of Populism: Global Perspectives*. Lexington, KY: Univ. Press of Kentucky.
- Dryzek, John S., and Jeffrey Berejikian. 1993. "Reconstructive Democratic Theory." *The American Political Science Review*, 87(1): 48–60.
- Eagly, Alice H., and Shelly Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth and TX: Harcourt Brace Jovanovich College Publishers.
- Edwards, Jeffrey E., and Richard P. Bagozzi. 2000. "On the Nature and Direction of Relationships between Constructs and Measures." *Psychological Methods* 5(2): 155–174.
- Elchardus, Mark, and Bram Spruyt, B. 2016. "Populism, Persistent Republicanism and Declinism: An Empirical Analysis of Populism as a Thin Ideology." *Government and Opposition* 51(01): 111–33.
- Farrell, James J., and Patrick R. Laughlin. 1976. "A Scale to Measure Populist Attitudes." *The Journal of Psychology*, 94(1): 33–38.
- Ford, Robert, Matthew J. Goodwin, and David Cutts. 2012. "Strategic Eurosceptics and polite xenophobes: Support for the United Kingdom Independence Party (UKIP) in the 2009 European Parliament elections." *European Journal of Political Research*, 51(2): 204–34.
- Hambleton, Ronald K., and Russell W. Jones. 1993. "An NCME Instructional Module on: Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development." *Educational Measurement: Issues and Practice* 12(3): 38–47.
- Hawkins, Kirk A., Scott Riding, and Cas Mudde. 2012. "Measuring Populist Attitudes." *IPSA Committee on Concepts and Methods Working Paper Series*. Accessed December 15, 2014. Available at: [www.concepts-method.org/](http://www.concepts-method.org/).
- Hawkins, Kirk A. 2009. "Is Chavez Populist? Measuring Populist Discourse in Comparative Perspective." *Comparative Political Studies* 42(8): 1040–67.

- Hawkins, Kirk A. 2010. *Venezuela's Chavismo and Populism in Comparative Perspective*. Cambridge: Cambridge University Press.
- Hawkins, Kirk A., and Scott Riding. 2010. Populist Attitudes and their Correlates among Citizens: Survey Evidence from the Americas. Paper prepared for the ECPR Workshop "Disassembling Populism (and Putting It Back Together Again): Collaborative Empirical Research on Interactions among Populism's Attributes," March 22–25, 2010, Muenster, Germany.
- Hibbing, John R., and Elizabeth Theiss-Morse. 2002. *Stealth Democracy. Americans' Beliefs About How Government Should Work*. Cambridge, UK: Cambridge University Press.
- Jackman, Simon. 2008. "Measurement." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 119–151. Oxford: Oxford University Press (The Oxford handbooks of political science).
- Kriesi, Hans-Peter, and Takis Pappas eds. 2015. *European Populism in the Shadow of the Great Recession*. Colchester: ECPR Press (Studies in European political science).
- Latin American Public Opinion Project (LAPOP). 2008. *The AmericasBarometer 2008*. Accessed October 10, 2015. Available at: <http://vanderbilt.edu/lapop/raw-data.php>.
- Mudde, Cas. 2004. "The Populist Zeitgeist." *Government and Opposition* 39(4): 542–63.
- . 2007. *Populist Radical Right Parties in Europe*. Cambridge and UK and New York: Cambridge University Press.
- Mudde, Cas, and Cristóbal Rovira Kaltwasser. 2013. "Exclusionary vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America." *Government and Opposition* 48(2): 147–74.
- eds. 2012. *Populism in Europe and the Americas: Threat or corrective for democracy?* Cambridge: Cambridge University Press.
- Oskamp, Stuart, and P. Wesley Schultz. 2014. *Attitudes and Opinions* (3rd edition). New York, NY: Psychology Press
- Roberts, Kenneth M. 2003. "Social Correlates of Party System Demise and Populist Resurgence in Venezuela." *Latin American Politics and Society* 45(3): 35–57.
- Rooduijn, Matthijs. 2014. "The Nucleus of Populism: In Search of the Lowest Common Denominator." *Government and Opposition* 49(04): 573–599.
- Rooduijn, Matthijs, Sarah de Lange, and Wouter van der Brug. 2014. "A Populist Zeitgeist? Programmatic Contagion by Populist Parties in Western Europe." *Party Politics* 20(4): 563–575.
- Rovira Kaltwasser, Cristóbal. 2014. "The Responses of Populism to Dahl's Democratic Dilemmas." *Political Studies* 62(3): 470–487.
- Samejima, Fumiko. 1997. "Graded Response Model." In *Handbook of Modern Item Response Theory*, edited by Wim J. Van der Linden and Ronald K. Hambleton, 85–100. New York: Springer.
- Sibley, Chris G., and Carla Houkamau. 2013. "The Multi-Dimensional Model of Maori Identity and Cultural Engagement: Item Response Theory Analysis of Scale Properties." *Cultural Diversity & Ethnic Minority Psychology* 19(1): 97–110.
- Stanley, Ben. 2011. "Populism, nationalism, or national populism? An analysis of Slovak voting behaviour at the 2010 parliamentary election." *Communist and Post-Communist Studies* 44(4): 257–270.
- Van Hauwaert, Steven M., and Stijn van Kessel. 2018. "Beyond Protest and Discontent: A Cross-National Analysis of the Effect of Populist Attitudes and Issue Positions on Populist Party Support." *European Journal of Political Research*. Advance online publication. doi: 10.1111/1475-6765.12216.
- Van Hauwaert, Steven M., Christian H. Schimpf, and Flavio Azevedo. 2017. "Measuring Populism as a Demand-Side Phenomenon: A Psychometric Evaluation of Populist Attitudes across Europe." Working Paper.
- Wainer, Howard, and David Thissen. 1996. "How Is Reliability Related to the Quality of Test Scores? What Is the Effect of Local Dependence on Reliability?." *Educational Measurement: Issues and Practice* 15(1): 22–29.
- Weyland, Kurt. 2001. "Clarifying a Contested Concept: Populism in the Study of Latin American Politics." *Comparative Politics* 34(1): 1–22.