Quality & Quantity (2019) 53:2983–3002
https://doi.org/10.1007/s11135-019-00913-6

Putting 'political' back in political trust: an IRT test of the unidimensionality and cross-national equivalence of political trust measures

T. W. G. van der Meer[1] · **E. Ouattara**[1]

Published online: 14 August 2019

Abstract
Much research intro political trust—its causes, correlates and trends—builds on the twin assumptions that trust in a wide range of political institutions is ultimately an expression of (1) a singular and (2) a cross-nationally equivalent underlying attitude. Yet, the widespread assumptions of unidimensionality and cross-national equivalence of political trust is at odds with the dominant conceptual understanding of political trust as a relational concept, driven by subjects, objects, and their interplay. This paper employs Rasch modelling as a direct, strict test of unidimensionality, equivalence and item hierarchy. We test the fit of the Rasch model on political trust items in seven widely used, cross-national surveys (World Values Survey, Afrobarometer, Arabbarometer, Asian Barometer, Eurobarometer, European Social Survey, and Latinobarometro), covering 161 national surveys in 119 countries across the globe. We find that the unidimensional specification of the Rasch model does *not* fit the standard political trust question batteries. Political trust is not cross-nationally equivalent; trust in specific political institutions is more than a mere indicator of an underlying attitude. This conclusion does not impede cross-national research into political trust; rather it illustrates the need for consistent robustness checks across a range of objects of political trust. Our findings open up new venues for substantive research questions on specific objects of political trust and their relationships.

Keywords Political trust · Trust in government · Item response theory · Rasch modelling · Comparative analysis · Measurement equivalence

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11135-019-00913-6) contains supplementary material, which is available to authorized users.

✉ T. W. G. van der Meer
t.w.g.vandermeer@uva.nl

[1] Department of Political Science, University of Amsterdam, Amsterdam, The Netherlands

1 Introduction

For decades, political trust research has been considered a necessary requirement for democratic resilience (Zmerli and Van der Meer 2017; Citrin and Stoker 2018). Cross-national research has therefore focused on political trust and its causes (e.g. Esaiasson 2011; Hakhverdian and Mayne 2012; Bargsted et al. 2017), correlates (e.g., Voogd and Dassonneville 2018; Choi and Kwon 2019), and supposed crisis (e.g., Norris 2011; Van Erkel and Van der Meer 2016). This empirical literature on political trust builds on widely shared assumptions that trust is unidimensional as well as cross-nationally equivalent. According to these assumptions political trust—conceptualized narrowly as trust in government, parliament, and political parties, and more broadly to include the civil service, the police, and the judicial system—is an expression of a single underlying attitude (cf. Fischer et al. 2010: 162). The twin assumptions of unidimensionality and cross-national equivalence allowed scholars to rely on single political trust indicators or political trust scales as more or less interchangeable indicators of a singular underlying concept.

Two strands of research have tested the presumed data-theoretical structure of political trust in a particularly stringent way. The first is the factor analytical approach, that focused on the measurement equivalence of trust in various objects (e.g., Breustedt 2018; Marien 2017; Schneider 2017), showing uniformly that political trust items relate to a single underlying factor. The second line of research departed from Item Response Theory, finding that political trust items form a unidimensional, hierarchical scale in which trust in one object is a precondition to trust in another object (e.g., Harteveld et al. 2013; Newton and Zmerli 2011; Zmerli and Newton 2017).

In light of this evidence, the notion of political trust in a wide range of political institutions as the expression of a single underlying attitude has become highly prevalent in the political trust literature. Nevertheless, this unidimensional interpretation of political trust is problematic, conceptually as well as empirically.

Conceptually, it conflicts with the common understanding of political trust as a fundamentally relational concept (Van der Meer 2017; Citrin and Stoker 2018): *A* trusts *B* to do *X* (Hardin 1992; cf. Bauer and Freitag 2018), in which *A* is the citizen and *B* is a political institution (Norris 2011; Zmerli and Van der Meer 2017). This relational approach implies that trust hinges on the subject (i.e., the citizen that does or does not trust), the object (i.e., the political institution that is or is not trusted), and their performance. Trust in a range of objects can hardly be considered relational if these trust objects would be little more than links in a chain and would primarily be expressions of a single underlying attitude. A strong unidimensional interpretation of political trust suggests that citizens do not substantially distinguish between objects such as government, parliament, and the justice system. To the extent that objects of political trust are interchangeable, object-specific benchmarks, expectations, and performances do not factor into subjects' attitudes. Political trust would then be reduced to its subject: *A* trusts. Hence, the prevalent assumptions of unidimensionality and equivalence of political trust are at odds with the equally prevalent understanding in the literature of political trust as a fundamentally relational concept.

Moreover, empirical evidence for the unidimensionality of political trust is indeterminate. One research line finds that trust in various political objects is a cross-nationally equivalent dimension with scalar invariance, whereas another finds that these trust items are hierarchically related (implying that trust in one object functions as a stepping stone to trust in others). Yet, neither is sufficient evidence to support the assumption of unidimensionality in for instance cross-national or longitudinal comparisons, or in explanatory

analyses. Crucially, the two traditions do not agree about the structure of the underlying political trust scale. This paper will therefore test to what extent the underlying dimension is metric *and* hierarchical via Rasch modeling.

The Rasch model (see Andrich 1988; Wright and Mok 2004) was developed in the tradition of Item Response Theory (van Schuur 2003). While it is vastly more difficult to meet the demands of the Rasch model, there are some distinct advantages to its use (Meijer et al. 1990; van Schuur 2003). The Rasch scale is a parametric test: Even though the original items may be dichotomous or ordinal, the resulting item and subject positions on the Rasch scale are metric (Meijer et al. 1990). Moreover, a fitting Rasch scale is more robust, as it is better equipped to compare scales that only include a subset of items or respondents (Glas and Verhelst 1995). Finally, and perhaps most importantly, because the ordering of items (and subjects) is invariant across the scale, the Rasch scale allows an absolute comparison of groups over time or across different sections of the underlying scale (van Schuur 2003). Rasch therefore allows us to test the unidimensionality and cross-national equivalence of the hierarchical model simultaneously (Annoni and Charron 2019).

Rather than a showcase and detailed introduction to the Rasch model, this paper aims to answer a question with rather substantive implications: Do political trust survey measures meet the demands of the Rasch model? We test the unidimensionality of political trust across seven widely used, cross-national data sets spanning the globe, and under a variety of methodological choices. Substantively, a strong Rasch model would imply that trust scholars should not be particularly interested in specific objects of trust, and need to reconsider our understanding of political trust as a relational concept.

2 The unidimensionality of political trust

Despite its widespread use, there is no undisputed understanding of the concept of political trust (Hardin 2000; Warren 2017). Commonly, even if mostly implicitly, political trust is understood to be relational—requiring a subject who trusts (truster) as well as a political object that is trusted (trustee)—as well as situational—applied to specific circumstances or tasks. *A trusts B to do X* (Hardin 1992). To the extent that trust is an evaluation of this relationship, scholars have proposed various criteria on which these objects are evaluated, such as their competence, motivation, accountability or reliability (van der Meer 2017). Our measures and measurement models of political trust do not cover these theoretical notions very deeply. Political trust is conventionally measured in survey research in a range of objects, such as government, parliament, and political parties (Zmerli and van der Meer 2017). Most measurement thereby incorporates the relational aspect (*A trusts B*), but not the situational aspect (*to do X*).

The last decade has seen an increased focus by political trust scholars into the dimensionality of people's answers to survey items on political trust in a range of political objects.[1] In the factor analytical tradition scholars have focused on the measurement equivalence of trust in various objects (e.g., Breustedt 2018; Marien 2017). This measurement equivalence is a necessary precondition for meaningful comparison (Davidov et al. 2014; Stegmueller 2011). Explaining group differences in political trust only becomes possible

[1] Empirical support for the existence of a single, latent dimension has also been found in studies employing less stringent techniques, including exploratory factor analyses (e.g., Hooghe 2011; Lu 2014) and reliability analyses (e.g., Van Erkel and Van Der Meer 2016; Choi and Kwon 2019).

if political trust items are understood similarly across these groups. Hence, scholars have looked into the measurement equivalence of scales measuring trust in a range of political objects across time (Poznyak et al. 2014; Smets et al. 2013; Turper and Aarts 2017), country (André 2014; Ariely 2015; Breustedt 2018; Marien 2017; Schneider 2017), time and country (Coromina and Davidov 2013; Marien 2011), subgroups (André 2014; De Vroome et al. 2013; Turper and Aarts 2017), and even levels of analysis (e.g., Ruelens et al. 2018). Measurement equivalence first requires that political trust has a similar structure across groups (configural equivalence), second that the factor loadings—that indicate how well individual trust items relate to this factor—are equal across groups (metric invariance), and ultimately that intercepts have a similar meaning across groups (scalar invariance). In this tradition there is no hierarchical relationship between items themselves. Studies in this field without exception found evidence for metric and even partial scalar invariance. That would imply that items measuring trust in a range of political objects relate similarly to an underlying structure that we label political trust, enabling a meaningful comparison of the underlying structure across time, space and/or subgroups (Breustedt 2018).

A second line of research departed from a different tradition, that of Item Response Theory. More specifically, scholars have tested to what extent political trust items form a hierarchical scale in which trust in one object is a (probabilistic) precondition or precursor to trust in another (e.g., Zmerli and Newton 2017). Mokken scale analysis (Gillespie et al. 1987; Sijtsma et al. 1990) has been the most popular way to test whether political trust items load on this single, underlying, hierarchical political trust dimension, benefiting from differences across items and respondents in means and variances. Mokken scale analysis has been implemented on political trust items in countries such as Italy (Quaranta 2014), the Netherlands (van Elsas 2015), Sweden (Naurin 2011) and the United Kingdom (Rose 2014) as well as in cross-national studies covering Europe (Harteveld et al. 2013; de Vries and van Kersbergen 2007) and the globe (Newton and Zmerli 2011; Zmerli and Newton 2017). All of these studies found evidence that the answer patterns in trust in various political objects indicate strong Mokken scales.[2] In other words, without exception these studies find strong evidence for an underlying, hierarchical structure that we label political trust.

Despite incidental pushback (e.g. Fischer et al. 2010), there is strong evidence to treat political trust measures as unidimensional and cross-nationally equivalent, as well as evidence for the existence of a hierarchical political trust scale. Yet, various shortcomings and irregularities cast the evidence for cross-national equivalence and unidimensionality in a new light.

First, there is an ongoing question of delineation of political trust by its objects. Some opt for a rather narrow operationalization, focusing primarily on parliament and government as its object (Ariely 2015; Van Erkel and Van Der Meer 2016), whereas others take up a broader definition including a.o. the police and the justice system (Marien 2017), the civil service (e.g., Zmerli and Newton 2017) and international institutions (e.g., André 2014; van Elsas 2015). Few studies actively report specific items that do not meet the demands of the scale (but see Breustedt 2018 who argues that the civil service hinders scalar invariance in her study of cross-national equivalence).

Second, while most studies find evidence for a singular political trust scale or factor, the number of dimensions is not uncontested. On the one hand, Zmerli and Newton (2017) analyzed the World Values Survey and show that political trust can also be understood to

[2] In methodological terms: The size of Loevinger's coefficient *H* (higher than .5) indicates a strong scale (Van Schuur 2003).

belong to an even broader, overarching dimension of societal trust. On the other hand, others show evidence for multiple scales within the subset of political trust items. In her analysis of the World Values Survey data, Breustedt (2018) concludes that political trust is best understood as two factors that she labels representative and implementing.[3]

Third, most of these scales have been tested in European countries. Only few studies actively incorporate democracies (let alone: non-democracies) outside of Europe. Studies that do encompass a wider set of countries were more likely to reach inconsistent findings. Zmerli and Newton (2017), for instance, find strong Mokken scales of political trust for all democracies in the World Values Survey except India, where they only find a weak scale after eliminating two trust items. Breustedt (2018) finds configural invariance of her two-factor model in only 19 of the 32 countries under study, and full invariance in even fewer.

Fourth, the measurement equivalence literature does not generally find the strongest level of invariance, i.e., full scalar invariance. At best, they tend to find partial scalar invariance (e.g., Marien 2017; Poznyak et al. 2014). Partial scalar invariance is a sufficient condition for meaningful comparisons across groups (Steenkamp and Baumgartner 1998), acknowledging that full scalar variance is in most instances an untenable demand (Byrne et al. 1989). Yet, other studies do not even find partial scalar invariance, particularly across countries (e.g., Ariely 2015; Breustedt 2018).

Fifth, while Mokken analyses consistently find strong, hierarchical political trust scales, the ordering of items from easy to difficult (i.e., most to least trusted) is not constant across countries (Zmerli and Newton 2017). Rather, there are vast differences. In Latin American countries such as Argentina and Chile, government is one of the most trusted political object, whereas particularly in European countries government belongs to the least trusted objects. While political parties are generally the least trusted object, this is not the case in Colombia and Mexico where the civil service is trusted even less. There might be country specific reasons—cultural, institutional, or otherwise—why specific objects are trusted more or less. Yet, although the trust *item* might itself still be cross-nationally equivalent, too much object-specific cross-national variance would imply that the underlying trust *scale* is not.

3 The Rasch model

Item Response Theory models strive for a "calibration of both items and persons onto a latent variable scale that represents a construct" (Engelhard Jr. 2008). In other words, in the IRT tradition both the truster (the subject who trusts) and the trustee (the item that is trusted, for practical purposes the trustworthiness of the institution) is positioned on the underlying dimension. In the IRT tradition three main approaches have been put forward: the Guttman, Mokken, and Rasch models (see Engelhard Jr. 2008 for a historical overview). All estimate subject and object positions simultaneously, acknowledging that the difficulty of items might differ. The Mokken and Rasch models are probabilistic, whereas the Guttman model is deterministic. The Mokken and Rasch models differ on various criteria (cf. Engelhard Jr. 2008; Gillespie et al. 1987; van Schuur 2003). Most notably, the Rasch model has stricter demands. Whereas the Mokken scale is a non-parametric test

[3] Although it is noteworthy that trust in government is part of the representative factor and not part of the implementing institutions, that encompass the police and the courts.

(Sijtsma et al. 1990), the Rasch model is a parametric model with a strict demand of double monotonicity.

A first requirement is monotone homogeneity: subjects who score higher on the latent trait (i.e., political trust) should also be more likely to dominate each individual item (i.e., have a high value on each political trust object). This has several implications. Subjects that dominate a difficult item (a less trustworthy object) should be even more likely to dominate all the easier items (the more trustworthy objects), but not necessarily inversely. An answer pattern in which a respondent trusts the hardest item (the averagely least trusted object) but not the easiest (the most trusted object) constitutes an error. In other words, the easier items function as stepping stones to the more difficult ones. In Mokken scale analysis, monotone homogeneity is a sufficient precondition to measure subject positions (Sijtsma et al. 1990; van Schuur 2003).

Yet, invariant measurement requires double monotonicity: the items that relate to the underlying scale should discriminate equally well albeit at different points of the underlying scale. The difference between 'easy' and 'difficult' items—reflecting the likelihood of respondents to dominate each item—should be consistent at all points of the scale (van Schuur 2003). Formally, the item response functions should not intercut but rather run parallel. Studies to date have almost without exception focused on the first demand (monotone homogeneity), but not on the more stringent second one.

The Rasch model originates from psychometrics and is rarely used in fields such as behavioral political science. As an IRT approach it stands out by being "grounded in the theory of fundamental measurement" (Annoni and Charron 2019): "First, the calibration of measuring instruments must be independent of those objects that happen to be used for calibration. Second, the measurement of objects must be independent of the instrument that happens to be used for the measuring" (Wright 1968). In this theory of fundamental measurement, measurement is thus not affected by the specific set of be subjects (e.g., respondents) or the specific of objects (e.g., items) to infer the measurement model (Wright and Mok 2004). "An important property of the Rasch model is that, under mild regularity assumptions, consistent item parameter estimates can be obtained from a sample of any subgroup of the population where the model holds. So item parameter estimates obtained using different samples from different subgroups (say, gender or ethnic subgroups) of the population should, apart from random fluctuations, be equal" (Glas and Verhelst 1995).

Double monotonicity is a firm requirement in the Rasch model. Compared to Mokken scale analysis, the assumptions behind the Rasch model are therefore more strict and rigid (Gillespie et al. 1987; van Schuur 2003). It is a parametric model, based on a log transformation of the item and person scores. We can sum up its main assumptions as follows: (1) sum scores are sufficient parameters to calculate the underlying person and item positions on the latent variable of interest; (2) the items constitute a single latent dimension; (3) local stochastic independence: "the response behavior of a person on an arbitrarily selected item g" does not depend on "his or her response on previous items, nor will it affect response behavior on subsequent items" (Meijer et al. 1990); (4) singular monotonicity and double monotonicity.

Figure 1 is a graphical representation of a unidimensional Rasch model of political trust items. Respondents' likelihood to dominate each of the six trust items (A–F) increases as respondents are more trusting on the underlying political trust dimension. Respondents who dominate the more difficult item are more likely to also dominate all easier items. The ordering of items by difficulty (likelihood to dominate) is constant, regardless of one's position on the underlying dimension. Finally, all items discriminate equally well, as is evidenced by the parallel trace lines.

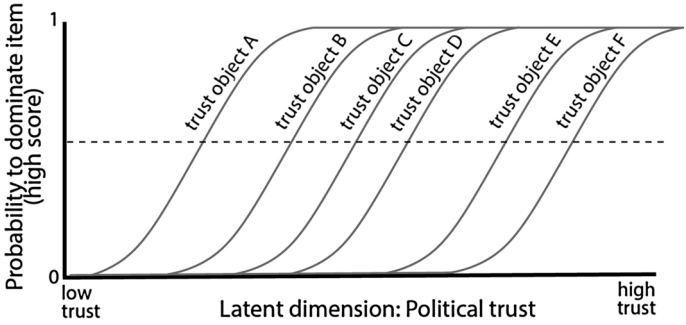

Fig. 1 The unidimensional Rasch model

There are different Rasch models based on sets of items with varying response category composition. For polytomous items, the most common models are the Rating Scale Model (RSM) and the Partial Credit Rasch Model (PCM) (Andrich 2016). The RSM is developed for a set of items with a constant set of response categories (such as Likert scales) that are also used similarly across items (Ostini and Nering 2006). The PCM has been developed to estimate Rasch parameters for items with varying response categories, and thus has more relaxed constraints (Masters 2016). In this paper we employ the RSM in line with the format of the data. For specific models, however, we also employ the PCM.

The Rasch rating scale model presents the probability to dominate category (k) of item (i) by person (n) as a natural logarithmic function of that person's position (B_n), the difficulty/trustworthiness of the object (D_i) and the threshold of the category (F_k) (Bond and Fox 2007: 281).

$$P_{nik} = \frac{e^{(B_n - D_i - F_k)}}{1 + e^{(B_n - D_i - F_k)}}$$

Like most IRT models (Cai and Thissen 2016), there is no single, absolute goodness-of-fit statistic for the Rasch model. Instead, there are multiple tests to assess whether the Rasch model holds. On the one hand, there are overall fit measures such as M_2, a limited-information test statistic that "is found to be especially effective for testing the fit of unidimensional IRT models" (Cai 2016). Model fit may be assessed via traditional parameters such as RMSEA and CFI. On the other hand, there are tests of specific assumptions of the Rasch model. Most notably, assumptions 1 (sufficiency of the sum scores) and 4 (monotone homogeneity and double monotonicity) can be tested by using Andersen's conditional LR test (Glas and Verhelst 1995: 70; Meijer et al. 1990: 289). The test is based on the premise that subgroups within the dataset have homogeneous item parameters (subgroup homogeneity). It evaluates differences between CML estimates of the item parameters in different subgroups based on score levels or other external criteria.[4] Assumptions 2 (unidimensionality) and 3 (local stochastic independence) can be tested through the use of the Martin-Löf LR test (Glas and Verhelst 1995: 70, 87). The Martin-Löf test assesses whether two sets of items—i.e., those items that respondents find easy and those that respondents

[4] For the purpose of this analysis we create subgroups based on the median sum score of individuals.

find hard to dominate—tap into the same latent dimensions and thereby meet the demands of the Rasch model.[5]

4 Data and methods

4.1 Data

We test the unidimensionality of political trust on the most recent waves of a globe spanning set of cross-national and longitudinal surveys that contain question batteries on trust in various institutions: the World Values Survey, Afrobarometer, Arabbarometer, Asiabarometer, Eurobarometer, European Social Survey, and the Latinobarometro.[6] All data sets are used in political trust research (see Zmerli and van der Meer 2017 for an overview). Together, they cover (1) a wide variety of democratic and non-democratic countries, (2) similar but different sets of (political) trust objects, and (3) various answer categories that range from a dichotomy in the Eurobarometer (tend to trust/distrust) to a ten point scale in the European Social Survey. This allows us to test the unidimensionality of political trust under a range of substantive and methodological conditions.

We cannot extensively discuss all data sets. Our main analyses therefore focus on one of these data sets, the sixth round of the World Values Survey 2010–2014 that includes a diverse set of countries across the globe. As a robustness check, we test the fit of the Rasch model on the six region-specific data sets, with generally the same results (see the overview in the section on Robustness checks, or a more extensive discussion in electronic supplementary material).

Two of the most extensive studies on the unidimensionality—the study on equivalence by Breustedt (2018) and the study on the hierarchical structure by Zmerli and Newton (2017)—analyze the sixth wave of the WVS. To mirror these studies, we limit our analyses of the WVS data to liberal democracies. Consequently, the sample of our main analyses covers 35,042 individuals in 23 countries with a Polity IV democracy score of at least 8 (10 representing a full democracy).[7] An additional 2917 individuals are dropped from our sample after we employ list-wise deletion across all 6 political trust items (though this did not affect our outcomes substantially, see below). In total our final sample size consists of 32,125 individuals.

4.2 Political trust items

The World Values Survey includes a lengthy question battery measuring confidence[8] in a range of societal and political objects. Six items with more or less political objects are available across all 23 countries in the data set:

[5] For the purpose of this analysis, we split items based on the median sum score of the items.

[6] Unfortunately, this selection of data sets does not encompass the United States and Canada, that tend to employ a different set of questions to measure trust in government (Dalton 2017).

[7] These countries are Argentina, Australia, Brazil, Chile, Cyprus, Estonia, Germany, India, Japan, Mexico, the Netherlands, Peru, Philippines, Poland, Romania, Slovenia, South Africa, South Korea, Spain, Sweden, Taiwan, United States, Uruguay. In line with Zmerli and Newton (2017) we did not include Ghana, New Zealand, and Trinidad and Tobago. We plan to do so in the next phase of our work.

[8] Confidence and trust are used rather interchangeably in the political trust literature (cf. Zmerli and Van der Meer 2017).

1. Confidence in the Police.
2. Confidence in the Court System.
3. Confidence in the Civil Service.
4. Confidence in the National Government.
5. Confidence in the National Parliament.
6. Confidence in Political Parties.

They are measured on a four point scale, with answer categories 1 ('a great deal'), 2 ('quite a lot'), 3 ('not very much') and 4 ('none at all'). We reverse coded these values from 0 to 3 such that value 0 represents ('none at all') and 3 ('a great deal'). While our main analysis relies on the full information (estimating a Rasch model for polytomous items), we checked the robustness of our findings after dichotomizing the trust items between 2 and 3.[9] Dichotomization did not affect our conclusions (see our online appendix).

While there are missing values on each trust item (see Table 1), we ran our main analyses on the subset of respondents without any missing value. Yet, we ran robustness checks on the WVS data by modeling item and person parameters for subgroups with different patterns of missing responses. Doing so made no substantive changes on the item locations, item fit, and global fit of the Rasch model (see our online appendix).[10]

4.3 Estimating the Rasch model

Lacking an absolute goodness-of-fit statistic for the Rasch model, we employ two strategies simultaneously. The first strategy is the estimation of the global fit of the Rasch model using the m_2 statistic in the mirt package (Chalmers 2012). It relies on ML estimators, and uses fit indices that may be interpreted similarly to more widely known SEM models. Its main downside is a lack of a priori specifications of the model in line with the assumptions of the Rasch model.

The second strategy is a test for violations of the key assumptions at the heart of the Rasch model. For that purpose we employ Andersen's conditional LR test and the Martin-Löf LR test (Glas and Verhelst 1995: 70, 87). Rasch parameters are estimated through Conditional Maximum Likelihood (CML) available through the eRm package provided by Mair et al. (2018).

Given the predefined structured response categories to the trust items, we use the Rating Scale Model. Yet, as a check on the Anderson LR test specifically, we also checked the robustness of the outcomes to the less stringent Partial Credit Model. Findings did not differ substantively.

[9] We do not know to what extent the differences between the four answer categories (item steps) are meaningful to the respondents. There are indications that the small differences between related answer categories are not very meaningful at the micro-level. However, this particularly relates to the 10 point scales used in for instance the European Social Survey. A higher degree of noise arguably makes it less likely to find fitting Rasch scales.

[10] There were extremely small differences between the list-wise deleted models and the models in which patterns of responses were incorporated as sub-groups. None of them changed the relative position of item locations, item fits, or the results of the Andersen LR test. The Martin-Löf test does not allow for missing values.

Table 1 Missing cases per item

Item	Number of missing cases	Percentage of missing cases
Confidence in the police	645	1.84
Confidence in parliament	1353	3.86
Confidence in the civil service	1579	4.51
Confidence in government	979	2.79
Confidence in political parties	1242	3.54
Confidence in the courts	1172	3.35
Sex	3	0.01
Age	21	0.06
Education	1332	3.80
List-wise missing on political trust	2917	8.32

All data sets used in this study are accessible without charge via their own web sites. The code we used to estimate all our models are deposited (see electronic supplementary material).

5 Results

5.1 Item hierarchy and single monotony

First, we assess the hierarchy of items from least difficult (most trusted institution) to most difficult (least trusted institution) in each country. Table 2 provides an overview of this hierarchical ordering, based on their means.

In the full sample, the more politicized institutions—parliament, government, and particularly political parties—are the least trusted institutions, whereas the more impartial institutions—the civil service, the courts, and particularly the police—are trusted the most.

Yet, the hierarchy differs substantially from country to country. In fact, very few countries conform to the ordering of items in the pooled data set. While it is difficult to find strict patterns, there are some indications that (Western) Europe differs from Latin America. The police are, for instance, trusted relatively less in countries such as Argentina, Mexico, and Uruguay. By contrast, government ranks better than average in countries such as Argentina and Chile.

This ordering of objects from most to least trusted at the group level does not tell us, whether these hierarchical relationships also structure individual's attitudes. For that purpose, the right panel in Table 2 provides the overall fit of the Mokken scale (Loevinger's H). In line with the conclusions of Zmerli and Newton (2017) we find that these six trust items form a strong Mokken scale, both in the pooled data set (H = 0.57) and in all separate countries save one (India). The scale fits particularly well in Japan and South Korea (H = 0.65). Within each country except for India, the political trust scales thus meet the demand of single monotony. Yet, as the hierarchy of items differs across countries, we cannot automatically conclude that these scales are cross-nationally equivalent.

Table 2 Item difficulty and Loevinger H-coefficient, by country (polytomous items)

Item difficulty (polytomous items), ordered by full sample							Loevinger H-coefficient	
Countries	Police	Courts	Civil service	National government	Parliament	Political parties	H (scale)	SE
Full sample	1	2	3	4	5	6	0.57	0.00
Argentina	4	5	1	2	3	6	0.51	0.02
Australia	1	2	3	5	4	6	0.49	0.02
Brazil	2	3	1	4	5	6	0.50	0.01
Chile	1	3	4	2	5	6	0.55	0.02
Cyprus	2	1	3	5	4	6	0.57	0.02
Germany	1	2	3	4	5	6	0.55	0.01
Estonia	1	3	2	4	5	6	0.58	0.02
Spain	1	2	3	5	4	6	0.48	0.02
India	2	3	1	5	4	6	0.21	0.01
Japan	2	1	3	4	5	6	0.65	0.01
Korea	2	1	3	4	5	6	0.65	0.02
Mexico	4	2	1	5	3	6	0.55	0.01
Netherlands	2	1	3	5	4	6	0.65	0.01
Peru	1	4	2	3	5	6	0.59	0.02
Philippines	2	1	3	5	4	6	0.52	0.02
Poland	1	2	3	4	5	6	0.61	0.02
Romania	1	2	3	4	5	6	0.59	0.02
Sweden	2	1	5	4	3	6	0.58	0.02
Slovenia	1	2	3	4	5	6	0.62	0.02
Taiwan	1	3	2	4	5	6	0.60	0.02
United States	1	2	3	4	5	6	0.58	0.01
Uruguay	4	1	2	5	6	3	0.50	0.02
South Africa	2	4	1	3	5	6	0.61	0.01

5.2 Global fit, Andersen LR test, and Martin-Löf test

Next, we turn to the test of the Rasch model. Table 3 presents the global fit indices for the unidimensional Rating Scale Model: The M_2 statistic, RMSEA, TLI and CFI. The global fit indicators strongly suggest that these scales do not fit the Rasch model. The non-significant M2-based p values indicate a poor fit. Likewise, the confidence intervals of the RMSEA are all above the widely accepted cut-off of 0.05 for close fit and 0.08 for adequate fit. The TLI and CFI values are well below commonly accepted cut-offs of 0.95 to 0.90. In other words, the Rasch Rating Scale Model is unlikely to have generated the observed data.

These global fit indicators only provide an estimate of closeness of fit to the data. We may use different tests of basically the very same model to assess whether specific demands of the Rasch model are met. We first focus on the Andersen LR Test to assess sub-scale invariance/homogeneity (see Table 4). The RSM model could not be estimated on Brazil, India, Mexico, South Africa, and Uruguay, as the minimization algorithm did

Table 3 WVS (2010–2014) global fit indicators for rating scale model

Country	M_2	*df*	*p*	RMSEA	RMSEA_5	RMSEA_95	TLI	CFI
India	196.465	12	0	0.099	0.087	0.111	−0.776	0.000
Peru	199.920	12	0	0.116	0.102	0.131	0.873	0.492
Mexico	331.858	12	0	0.117	0.106	0.128	0.834	0.337
Brazil	299.846	12	0	0.130	0.117	0.142	0.671	0.000
Taiwan, Republic of China	244.621	12	0	0.133	0.119	0.148	0.643	0.000
Chile	219.670	12	0	0.136	0.120	0.152	0.621	0.000
South Africa	804.245	12	0	0.143	0.135	0.151	0.715	0.000
Philippines	318.961	12	0	0.146	0.133	0.160	0.422	0.000
Spain	300.971	12	0	0.149	0.135	0.164	0.560	0.000
Cyprus	312.282	12	0	0.162	0.146	0.177	0.281	0.000
Poland	302.199	12	0	0.174	0.157	0.191	0.485	0.000
Argentina	358.214	12	0	0.174	0.159	0.190	0.458	0.000
Slovenia	392.313	12	0	0.179	0.164	0.194	0.590	0.000
Australia	553.898	12	0	0.179	0.167	0.192	0.296	0.000
Sweden	381.103	12	0	0.179	0.164	0.195	−0.081	0.000
Uruguay	336.469	12	0	0.181	0.164	0.198	0.270	0.000
Germany	839.380	12	0	0.191	0.181	0.203	−0.024	0.000
Romania	741.546	12	0	0.212	0.199	0.225	0.437	0.000
South Korea	666.249	12	0	0.214	0.200	0.228	0.236	0.000
Estonia	812.649	12	0	0.218	0.205	0.230	−0.183	0.000
Netherlands	1.012.109	12	0	0.221	0.210	0.233	0.242	0.000
United states	1.360.736	12	0	0.230	0.220	0.240	−0.136	0.000
Japan	1.666.991	12	0	0.268	0.257	0.279	−0.357	0.000

Fit indicators of the rating scale model sorted by the M2-based RMSEA

not converge.[11] In the Andersen LR test the response patterns need to meet a number of criteria. Among them is the condition that each item has an equal number of response categories and that variation exists in the response pattern for each sub-sample. For example, in the event that everyone in the low-trusting group answers 0 to the item trust in parliament, that item cannot be used for the Andersen LR test. Table 4 lists the items that meet these pre-conditions. It shows that the full set of trust items do not meet the demands of the Rasch scale: The significant *p* values in Table 3 indicate an important difference between item location parameters (i.e., the trustworthiness of these institutions) within sub-samples of highly-trusting individuals and others with low-political trust.[12] As a check, we

[11] A closer analysis of the item trace lines suggests that in these countries response patterns conditioned on the latent score on certain items were very closely related, making it difficult for the algorithm to converge on a solution of the location parameters.

[12] The non-significant *p* values in Slovenia, South Korea, and Poland indicate a fit of the Rasch model, but only two-item scales consisting of confidence in the police and the courts in Slovenia and Poland and confidence in the police and the civil service in South Korea. In the Philippines and Taiwan no likelihood value converged.

Table 4 WVS (2010–2014) Andersen likelihood ratio test for rating scale model (RSM)

Country	LR	Chi^2 *df*	*p* value	Items kept
Slovenia	3.796	3	0.284	Police, courts
South Korea	5.988	3	0.112	Police, civserv
Poland	6.123	3	0.106	Police, courts
Japan	11.543	3	0.009	Police, parties
Netherlands	15.746	3	0.001	Police, civserv
Sweden	25.546	3	0.000	Civserv, courts
Spain	41.980	6	0.000	Police, parl, civserv, parties, courts
Argentina	59.883	6	0.000	Police, civserv, gov, parties, courts
Chile	54.379	3	0.000	Police, parties
Estonia	66.030	4	0.000	Police, gov, courts
Australia	127.345	5	0.000	Police, parl, civserv, courts
Cyprus	118.860	7	0.000	Police, parl, civserv, gov, parties, courts
Germany	187.965	7	0.000	Police, parl, civserv, gov, parties, courts
Peru	119.425	5	0.000	Police, civserv, gov, courts
Philippines	13,783,097.215	7	0.000	Police, parl, civserv, gov, parties, courts
Romania	133.021	7	0.000	Police, parl, civserv, gov, parties, courts
Taiwan, Republic of China	18,511,187.548	6	0.000	Police, civserv, gov, parties, courts
United states	123.877	5	0.000	Police, civserv, gov, courts

Andersen likelihood ratio test with median split: location parameters estimated with CML in eRm package

estimated a PCM instead of the RSM; this did not affect our conclusion (see online appendix). Finally, the Martin-Löf test leads to the same conclusion (see Table 5): If we split the set of items in two, the two resulting scales should be treated as significantly different scales.

6 Robustness

Up to this point, we performed multiple tests (m_2, Andersen LR, Martin-Löf) in various permutations (RSM/PCM, full range or dichotomized data, with or without the worst fitting items of police and parties, various ways to deal with missing data, bootstrapping) to understand the empirical fit of the very same unidimensional, hierarchical model. The outcomes are not affected by these permutations: The unidimensional Rasch model does not fit the political trust question battery in any democratic country in the World Values Survey (2010–2014). While we do find strong hierarchical (Mokken) scales within most countries, these scales are not at all cross-nationally equivalent.

The strongest robustness check entails fitting the Rasch model on other survey data that tend to differ in wording, number and type of political trust objects, number of answer categories, and most crucially context. Employing a similar setup as in our main analysis (e.g., original range of answer categories; listwise deletion of missing values), we attempted to fit the unidimensional Rasch model on the most recent waves of the Afrobarometer, Arabbarometer, Asian and South Asian barometer, Eurobarometer, European Social Survey, and

Table 5 WVS (2010–2014) Martin-Löf test for RSM/PCM

Country	LR	*df*	*p* value
Argentina	256.510	80	0.0000
Australia	895.766	80	0.0000
Brazil	567.910	80	0.0000
Chile	387.858	80	0.0000
Cyprus	395.890	80	0.0000
Germany	1181.433	80	0.0000
Estonia	1068.763	80	0.0000
Spain	433.183	80	0.0000
India	255.821	80	0.0000
Japan	1343.498	80	0.0000
South Korea	724.723	80	0.0000
Mexico	864.098	80	0.0000
Netherlands	1287.098	80	0.0000
Peru	439.198	80	0.0000
Philippines	436.874	80	0.0000
Poland	413.286	80	0.0000
Romania	1197.510	80	0.0000
Sweden	247.994	80	0.0000
Slovenia	497.759	80	0.0000
Taiwan	360.970	80	0.0000
United States	1221.255	80	0.0000
Uruguay	325.143	80	0.0000
South Africa	767.049	80	0.0000

Median splits and two item groups

Latinobarometro. In these robustness tests we did not limit the sample of countries to those with a sufficiently high Polity IV score, but include full, partial and non-democracies. We ended up with 138 national surveys in 119 unique countries and territories in these 6 surveys. The item hierarchies tend to be more similar within these geographic regions/(sub-) continents than across regions, but with the exception of the Arab region they are nevertheless not at all identical.

The *Afrobarometer (wave 6, 2014–2016)* contains a political trust battery with a somewhat different set of political objects than the other data sets. It includes trust in government, parliament, the president, courts, the police, the national election committee, the tax department, and the ruling party. The hierarchy of these trust objects varies vastly across countries. For instance, trust in the president is highest in some dictatorships with a strong repression record (Burundi, Cameroon, Sudan, Uganda, Zimbabwe). Although we find moderate to strong Mokken scales in all countries, the Rasch model does not fit in any (see online appendix Table A1).

The *Arabbarometer (2016–2017)* covers the same set of trust objects that were central in our main analyses, except for the civil service. We find moderate to strong Mokken scales in all seven countries under study; moreover, the item hierarchies are identical across the Arab countries in this survey. More remarkably, the global fit indices suggest an acceptable fit of the Rasch model for two countries, Tunisia and Palestine (see online

appendix Table A2). While Palestine does not meet the specific assumptions assessed by the Andersen LR test, we find a consistently acceptable fit in Tunisia: The item location parameters for high and low trusters does not differ significantly. Note, however, that the Tunisian case did not fit the Rasch model as part of the previous survey, the Afrobarometer.

The *Asian and South Asian barometers (2010–2016)* covers a rather broad set of trust objects, even though not all objects were questioned in all countries (see the online report for more details). It includes trust in the civil service, the courts, local governments, national government, the national electoral commission, parliament, political parties, the police, and the presidency/Prime Minister. We find weak to strong Mokken scales in all countries, even though the item hierarchies strongly differ across countries. In none of the countries, we find a fitting Rasch model. Some countries (Philippines, Mongolia, Indonesia) have an acceptable RMSEA, but then other fit measures are out of bounds (see appendix Table A3).

The *Eurobarometer (wave 87.3 in 2017)* stands out because the trust items are dichotomous.[13] The question battery includes courts, national government, national parliament, the police, the public administration, local government and political parties. Although there are consistently strong to very strong Mokken scales, the Rasch model does not fit the Eurobarometer data (see appendix Table A4).[14] Dichotomous measures might reduce the complexity for respondents, but do not evidently create better fitting Rasch scales.

The *European Social Survey (wave 8, 2016–2017)* stands out at the other end of the spectrum, for having eleven-point scale trust items, ranging from 0 (no trust at all) to 10 (complete trust). We find consistently strong Mokken scales in each country, and the item hierarchies tend to be rather similar across countries. Moreover, the values of the RMSEA, the TLI and the CFI suggest that in 6 countries the data may adequately fit the Rasch model, i.e., Slovenia, Portugal, Hungary, Spain, France, and Israel (see appendix Table A5). Although stricter and more commonly accepted cut-off values to TLI and CFI ($>.90$) suggest the Rasch model only fits responses in Slovenia and Hungary, we consider it worthwhile to assess whether the assumptions of the Rasch model are met in all six countries. The Andersen LR test reveals that none of the six countries were close to an acceptable p value for the Andersen LR test. Political trust items with more response categories thus do not necessarily yield better measurement. Rather, these long ranged answer categories may lead to more violations of the sufficiency of the sum-score for person placement on the latent scale.

Finally, the *Latinobarómetro (2017)* covers trust in the electoral system, government, the justice system, parliament, political parties, and the police along a four point scale. While Mokken scales tend to be (very) strong, the global fit of the Rasch model does not fit in most Latin countries (see appendix Table A6). In four countries—Colombia, Peru, Chile, and Paraguay—the Rasch model fits on multiple fit indicators, but not all. Particularly, the CFI values remain below the commonly accepted cut-off of 0.95. Additional analyses—particularly the Andersen LR test to assess sub-sample invariance—show that the Rasch model does not meet the assumptions behind the Rasch scale in these four countries.

[13] Consequently, fewer items are excluded in the Andersen LR test, as fewer cell combinations are empty.

[14] In approximately 7 of the 35 countries, there was a mismatch between the M2-based fit indices and the Andersen LR test. While the former indicated a poor fit, the latter suggested that the scales in these 7 countries met the assumptions tested by the LR test. Closer analysis in these 7 countries reveal important violations of the local independence assumption (see online appendix for further details).

All in all, the conclusion should be that we find a fitting Rasch model only in 1 of 161 national surveys.

7 Conclusion

The vast and expanding empirical literature on the causes, correlates, and crises of political trust has built on the twin assumptions that political trust is unidimensional and cross-nationally equivalent. Support for these twin assumptions have been based on confirmatory factor analyses as well as Mokken scale analyses (e.g., Breustedt 2018; Marien 2017; Zmerli and Newton 2017). If these twin assumptions are valid, this would suggest that political trust is hardly relational (driven by characteristics of subjects, objects, and their interaction), as objects of political trust would to a considerable extent be interchangeable. Rather, political trust would primarily be an outcome of subjects' structural and cultural tendencies to be trusting or distrusting. Methodologically, these assumptions imply that studies directly speak to each other even when they focus on different trust objects (such as government, parliament, or various scales), and that the structure of underlying attitudes is similar across countries. In this article, we questioned how realistic these twin assumptions are.

The conclusion of our analyses is rather unequivocal and sobering: We find no evidence that the political trust items meet the demands for a unidimensional Rasch model. Despite the strong evidence for partial scalar invariance in the measurement equivalence literature, and despite the strong evidence for monotonous homogeneity in studies applying Mokken scale analysis to these political trust items, the Rasch model does not hold. We studied seven cross-national data sets across the globe, containing 161 national surveys in 119 countries and territories, and found that the Rasch model fit the data in only 1 of those surveys.

While the Rasch model does not hold, the political trust items do form strongly homogenous Mokken scales in almost all national surveys we studied. While they do not meet the demand of double monotonicity, they do meet the demand of monotonous homogeneity. Non-metric, country-specific, hierarchical scales of political trust items can thus still be used, even though there is no measurement invariance across all levels of the scale.

Moreover, these political trust scales are not equivalent across countries. The hierarchical structure of the political trust objects varies in important ways (see also Zmerli and Newton 2017, p. 119). Generally, impartial institutions—the civil service, the courts, and particularly the police—are trusted more than the representative institutions—parliament, government, and particularly political parties. Yet, political parties are the most trusted in China; parliament is trusted more than courts in most countries across the African continent; and government is among the most trusted institutions in Argentina and Chile. On the one hand, this lack of a consistent hierarchy of trust objects emphasizes the different conceptualizations of trust in democratic and authoritarian countries (see Zmerli and van der Meer 2017). On the other hand, we cannot conclude that political trust scales have cross-nationally equivalent structure if the hierarchy of the items making up this scale differs so substantially across countries.

The unidimensional and cross-nationally equivalent nature of political trust was an awkward assumption anyway. It is directly at odds with the equally dominant conceptual understanding of political trust as a relational and at least partially evaluative concept. In this relational, evaluative approach trust is the result of the subject (truster), object (trustee),

and their ties (Bauer and Freitag 2018; Hardin 1992; van der Meer 2017). If a Rasch model had fit the political trust items, trust would not be object-driven. The objects' (procedural) performance and the subjects' norms and benchmarks towards each of those objects would not have been a relevant factor; impartial institutions such as the courts and representative institutions such as parliament would be evaluated by the same yardstick.

But that is not the case. Rather, the outcomes of our analyses show that the various objects of political trust cannot be reduced to mere links in a chain. These institutions have unique features that factor in the trust that people have in them. Trust may 'operate' somewhat '*similarly* across these institutions' (cf. Fischer et al. 2010: 162) (hence the fitting Mokken scales), but not *identically* (hence the non-fitting Rasch scales). We cannot safely assume that political trust has a uniform, unidimensional structure across countries, that political trust items are to some extent interchangeable, or that political trust scales are cross-nationally equivalent.

What does this conclusion mean for the substantive literature on political trust?

First, the outcomes of this article do not imply that we need to stop asking our current sets of research questions. Cross-national analyses of political trust remain feasible and just as relevant as they are now. But these analyses cannot be done under a self-evident assumption of unidimensionality and/or cross-national equivalence. Rather, they require more tests to assess the robustness of conclusions on multiple measures of political trust, to assess whether relationships are similar across multiple objects of trust.

Second, our conclusions imply that some substantive research questions have gotten insufficient attention under the assumption of unidimensionality. The rejection of the Rasch model suggests that individual objects of political trust are sufficiently distinct to analyze in their own right. Even though the underlying hierarchical structure is not cross-nationally equivalent, the specific trust items continue to be focal points for scholarly inquiry. Trust objects have some unique meaning to respondents, as they function as more than mere indicators of an underlying scale. These differences between the political institutions do not merely show up in levels of trust respondents assign to them, but also in their variance structure. In this perspective the focus on (respondents with) answer patterns that do *not* conform to the unidimensional scale offers information that is substantively relevant (cf. Loner 2016).

Third, the different item hierarchies across countries and across different levels of political trust are interesting objects for further study. Theorizing about these interdependencies is quite common in systems of multilevel governance (e.g., Harteveld et al. 2013; Muñoz 2017, who study the relationship between trust in national and European political institutions), but not so much for objects at the national level such as impartial and representative institutions. One may, for instance, consider the relationship between civil society and political parties under varying conditions of political clientelism. Similarly, populist messages might explain why in some countries government is trusted more than courts, whereas in many liberal democracies the impartial institutions are a stepping stone to trust in representative institutions.

The lack of a fitting Rasch scale reflects the lack of firm theorizing about the underlying measurement model. Political trust scholars have often been limited by the availability of data. The conventional measure of political trust consists of a question battery on trust or confidence in a range of objects. Yet, these political trust measures have not been developed and optimized a priori from a firm conceptual footing with the aim to be unidimensional, hierarchical, and equivalent across subgroups. More detailed data-theoretical models are needed that do not merely isolate objects of trust but also take the circumstantial and evaluative nature of political trust into account (Van der Meer 2017; Bauer and Freitag

2018) and distinguish by type of trust (Fischer et al. 2010; Warren 2017). Open-ended and close-ended, qualitative and quantitative probing strategies allow us to test the empirical validity of these models, by comparing them to respondents' interpretations across countries. Ultimately, it might prove to be as fruitful as costly to develop more refined measures of political trust that balance conceptual suggestions (such as the circumstantial and evaluative nature of political trust) with the need for a firm empirical foundation behind measurement models.

Acknowledgements This work was supported by the Dutch Science Foundation NWO (Grant Number 452-16-001, 2016).

References

André, S.: Does trust mean the same for migrants and natives? Testing measurement models of political trust with multi-group confirmatory factor analysis. Soc. Indic. Res. **115**, 963–982 (2014)

Andrich, D.: Rasch models for measurement. SAGE Publications Inc, Newbury Park (1988)

Andrich, D.: Rasch rating-scale model. In: van der Linden, W.J. (ed.) Handbook of item response theory, vol. 1. Chapman and Hall, Boca Raton (2016)

Annoni, P., Charron, N.: Measurement assessment in cross-country comparative analysis: Rasch modelling on a measure of institutional quality. Soc. Indic. Res. **141**, 31–60 (2019). https://doi.org/10.1007/s11205-017-1821-z

Ariely, G.: Trusting the press and political trust: a conditional relationship. J. Elections Public Opin. Parties **25**, 351–367 (2015). https://doi.org/10.1080/17457289.2014.997739

Bargsted, M., Somma, N.M., Castillo, J.C.: Political trust in Latin America. In: Zmerli, S., Van der Meer, T. (eds.) Handbook on political trust, pp. 217–395. Edward Elgar Publishing, Cheltenham (2017)

Bauer, P.C., Freitag, M.: Measuring trust. In: Uslaner, E.M. (ed.) The oxford handbook of social and political trust, pp. 15–36. Oxford University Press, Oxford (2018)

Bond, T.G., Fox, C.M.: Applying the Rasch model: fundamental measurement in the human sciences. Routledge, New York (2007)

Breustedt, W.: Testing the measurement invariance of political trust across the globe. A multiple group confirmatory factor analysis. Methods Data Anal **12**, 7–46 (2018). **10.12758/mda.2017.06**

Byrne, B.M., Shavelson, R.J., Muthén, B.: Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. Psychol. Bull. **105**, 456–466 (1989)

Cai, L.: Two-tier item factor analysis modeling. In: Van der Linden, W.J. (ed.) Handbook of item response theory, vol. 1, pp. 421–435. Chapman and Hall, Monterey (2016)

Cai, L., Thissen, D.: Nominal categories models. In: Van der Linden, W.J. (ed.) Handbook of item response theory, vol. 1, pp. 79–102. Chapman and Hall, Monterey (2016)

Chalmers, P.: mirt: a multidimensional item response theory package for the R environment. J. Stat. Softw. **48**, 1–29 (2012)

Choi, Y.-T., Kwon, G.-H.: New forms of citizen participation using SNS: an empirical approach. Qual. Quant. **53**, 1–17 (2019). https://doi.org/10.1007/s11135-018-0720-y

Citrin, J., Stoker, L.: Political trust in a cynical age. Annu. Rev. Political Sci. **21**, 2.1–2.22 (2018)

Coromina, L., Davidov, E.: Evaluating measurement invariance for social and political trust in Western Europe over four measurement time points (2002–2008). ASK Res. Methods **22**, 37–54 (2013)

Dalton, R.J.: Political trust in North America. In: Zmerli, S., van der Meer, T.W.G. (eds.) Handbook on political trust, pp. 375–394. Edward Elgar Publishing, Cheltenham (2017)

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., Billiet, J.: measurement equivalence in cross-national research. Annu. Rev. Sociol. **40**, 55–75 (2014). https://doi.org/10.1146/annurev-soc-071913-043137

De Vries, C.E., van Kersbergen, K.: Interests, identity and political allegiance in the European Union. Acta Polit. **42**, 307–328 (2007). https://doi.org/10.1057/palgrave.ap.5500184

De Vroome, T., Hooghe, M., Marien, S.: The origins of generalized and political trust among immigrant minorities and the majority population in the Netherlands. Eur. Sociol. Rev. **29**, 1336–1350 (2013)
Engelhard Jr., G.: Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. Meas. Interdiscip. Res. Perspect. **6**, 155–189 (2008). https://doi.org/10.1080/15366360802197792
Esaiasson, P.: Electoral losers revisited: how citizens react to defeat at the ballot box. Elect. Stud. **30**(1), 102–113 (2011)
Fischer, J., Van Heerde, J., Tucker, A.: Does one trust judgement fit all? Linking theory and empirics. Br. J. Politics Int. Relat. **12**(2), 161–188 (2010)
Gillespie, M., Tenvergert, E.M., Kingma, J.: Using Mokken scale analysis to develop unidimensional scales: do the six abortion items in the NORC GSS form one or two scales? Qual. Quant. **21**, 393–408 (1987)
Glas, C.A.W., Verhelst, N.D.: Testing the Rasch model. In: Fischer, G.H., Molenaar, I.W. (eds.) Rasch models: foundations, recent developments, and applications, pp. 69–95. Springer, New York (1995)
Hakhverdian, A., Mayne, Q.: Institutional trust, education, and corruption: a micro–macro interactive approach. J. Politics **74**(3), 739–750 (2012)
Hardin, R.: The street-level epistemology of trust. Analyse Kritik **14**, 152–176 (1992)
Hardin, R.: Do we want trust in government. In: Warren, M.E. (ed.) Democracy and trust. Cambridge University Press, Cambridge (2000)
Harteveld, E., van der Meer, T.W.G., De Vries, C.E.: In Europe we trust? Exploring three logics of trust in the European Union. Eur. Union Politics **14**, 542–565 (2013)
Hooghe, M.: Why there is basically only one form of political trust. Br. J. Politics Int. Relat. **13**, 269–275 (2011). https://doi.org/10.1111/j.1467-856X.2010.00447.x
Loner, E.: A new way of looking at old things. An application of Guttman errors analysis to the study of environmental concern. Qual. Quant. **50**, 823–847 (2016). https://doi.org/10.1007/s11135-015-0177-1
Lu, P.: A comparative analysis of political confidence in the BRICS countries. Jpn. J. Political Sci. **15**, 417–441 (2014). https://doi.org/10.1017/S1468109914000176
Mair, P., Hatzinger, R., Maier, M.J.: eRm: extended Rasch modeling. R package version 0.16-2 (2018). http://CRAN.R-project.org/package=eRm. Accessed 22 Nov 2018
Marien, S.: Measuring political trust across time and space. In: Hooghe, M., Zmerli, S. (eds.) Political trust: why context matters, pp. 13–46. ECPR Press, Colchester (2011)
Marien, S.: The measurement equivalence of political trust. In: Zmerli, S., van der Meer, T.W.G. (eds.) Handbook on political trust, pp. 89–103. Edward Elgar Publishing, Cheltenham (2017)
Masters, G.N.: Partial credit model. In: Van der Linden, W.J. (ed.) Handbook of item response theory, vol. One, pp. 137–154. Chapman and Hall, Boca Raton (2016)
Meijer, R.R., Sijtsma, K., Smid, N.G.: Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. Appl. Psychol. Meas. **14**, 283–298 (1990)
Muñoz, J.: Political trust and multilevel government. In: Zmerli, S., van der Meer, T.W.G. (eds.) Handbook on political trust, pp. 69–88. Edward Elgar Publishing, Cheltenham (2017)
Naurin, E.: Election promises, party behaviour and voter perceptions. Palgrave Macmillan, London (2011)
Newton, K., Zmerli, S.: Three forms of trust and their association. Eur. Political Sci. Rev. **3**, 169–200 (2011)
Norris, P.: Democratic deficit: critical citizens revisited. Cambridge University Press, Cambridge (2011)
Ostini, R., Nering, M.L.: Polytomous item response theory models. SAGE Publications Inc, Thousand Oaks (2006)
Poznyak, D., Meuleman, B., Abts, K., Bishop, G.F.: Trust in American government: longitudinal measurement equivalence in the ANES, 1964–2008. Soc. Indic. Res. **118**, 741–758 (2014)
Quaranta, M.: The 'normalisation' of the protester: changes in political action in Italy (1981–2009). South Eur. Soc. Politics **19**, 25–50 (2014)
Rose, J.: The public understanding of political integrity. Palgrave Macmillan, London (2014)
Ruelens, A., Meuleman, B., Nicaise, I.: Examining measurement isomorphism of multilevel constructs: the case of political trust. Soc. Indic. Res. **140**(3), 907–927 (2018). https://doi.org/10.1007/s11205-017-1799-6
Schneider, I.: Can we trust measures of political trust? Assessing measurement equivalence in diverse regime types. Soc. Indic. Res. **133**, 963–984 (2017). https://doi.org/10.1007/s11205-016-1400-8
Sijtsma, K., Debets, P., Molenaar, I.W.: Mokken scale analysis for polychotomous items: theory, a computer program and an empirical application. Qual. Quant. **24**, 173–188 (1990)
Smets, A., Hooghe, M., Quintelier, E.: The scale validity of trust in political institutions measurements over time in Belgium. An analysis of the European Social Survey, 2002–2010. In: European Survey Research Association (ESRA), Date: 2013/07/15–2013/07/19. Ljubljana: Centre for Citizenship and Democracy (2013)
Steenkamp, J.-B.E., Baumgartner, H.: Assessing measurement invariance in cross-national consumer research. J. Consum. Res. **25**, 78–90 (1998)

Stegmueller, D.: Apples and oranges? The problem of equivalence in comparative research. Political Anal. **19**, 471–487 (2011)
Turper, S., Aarts, K.: Political trust and sophistication: taking measurement seriously. Soc. Indic. Res. **130**, 415–434 (2017). https://doi.org/10.1007/s11205-015-1182-4
Van der Meer, T.W.G.: Political trust and the "crisis of democracy." Oxford Research Encyclopedia of Politics (2017). https://doi.org/10.1093/acrefore/9780190228637.013.77
Van Elsas, E.: Political trust as a rational attitude: a comparison of the nature of political trust across different levels of education. Political Stud. **63**, 1158–1178 (2015). https://doi.org/10.1111/1467-9248.12148
Van Erkel, P.F.A., Van Der Meer, T.W.G.: Macroeconomic performance, political trust and the Great Recession: a multilevel analysis of the effects of within-country fluctuations in macroeconomic performance on political trust in 15 EU countries, 1999–2011. Eur. J. Political Res. **55**, 177–197 (2016). https://doi.org/10.1111/1475-6765.12115
Van Schuur, W.H.: Mokken scale analysis: between the Guttman scale and parametric item response theory. Political Anal. **11**, 139–163 (2003)
Voogd, R., Dassonneville, R.: Are the Supporters of Populist Parties Loyal Voters? Dissatisfaction and Stable Voting for Populist Parties. Gov. Oppos. (2018). https://doi.org/10.1017/gov.2018.24
Warren, M.E.: What kinds of trust does a democracy need? Trust from the perspective of democratic theory. In: Zmerli, S., van der Meer, T.W.G. (eds.) Handbook on political trust, pp. 33–52. Edward Elgar Publishing, Cheltenham (2017)
Wright, B.D.: Sample-free test calibration and person measurement. In: Proceedings of the 1967 Invitational Conference on Testing Problems, pp. 85–101 (1968)
Wright, B.D., Mok, M.M.C.: An overview of the family of Rasch measurement models. In: Smith, E.V., Smith, R.M. (eds.) Introduction to Rasch measurement: theory, models and applications, pp. 1–24. JAM Press, Maple Grove (2004)
Zmerli, S., Newton, K.: Objects of political and social trust: scales and hierarchies. In: Zmerli, S., van der Meer, T.W.G. (eds.) Handbook on political trust, pp. 67–94. Edward Elgar Publishing, Cheltenham (2017)
Zmerli, S., van der Meer, T.W.G. (eds.): Handbook on political trust. Edward Elgar Publishing, Cheltenham (2017)