



Democracy Measurement and Metrology

Clara Monteiro Vieira¹ and Elisabeth Costa Monteiro¹

¹ Postgraduate Programme in Metrology, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro – RJ, Brazil

beth@puc-rio.br

Abstract. Providing reliable and comparable results of democracy measurement is essential for scientific research and decision-making processes. While previous studies have discussed strategies for developing and evaluating democracy measurement systems, there is no in-depth analysis scrutinizing the existence of essential elements of metrology in this framework. Considering the alternatives described in the literature to meet metrological requirements regarding measurements associated with the complex measurands from the Human and Social Sciences, this work investigates the possible advances or yet-to-be-tackled features to provide reliable and comparable measurement results of democracy. Through analysing the most relevant democracy measuring systems available and their validation approaches, this study reveals that while metrology is not explicitly mentioned, diverse essential elements of the Science of Measurement are embodied in the concerns surrounding the extant methods used to measure democracy. Rasch's theory, valuable for assessing measurement validity in the Human and Social Sciences, also holds the potential for overcoming metrological challenges associated with this field but is practically absent in democracy measurements. Adopting the Rasch model in the future could lead to theoretical advances and yield objectivity and intersubjectivity, thus providing metrological traceability and accurate results of democracy measurements.

1. Introduction

Fundamental precepts of Metrology apply to measurement practices across all fields of knowledge. An essential requirement is that of metrological traceability, which ensures the achievement of comparable measurement results for the same property, as each result “can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (JCGM 200:2012, section 2.41) [1]. While providing metrological traceability is naturally challenging, it becomes particularly more complex in the Human and Social Sciences, in which the intricate features of their measurands and measurement processes pose unique difficulties [2]-[14].

Subjective perceptions play a critical role in measuring psychosocial phenomena, as they can be the object of measurement or even a part of the measurement system [2], [6], [10], [11]. The complexity of the measurement process is enhanced by the numerous variables that may affect it, as they are particularly hard to assess –an issue that is shared, to some extent, with other domains, such as chemistry, biology, and quantum physics [14], [15].

Likewise, the measured properties within the Human and Social Sciences are usually characterized by significant definitional uncertainty. Many concepts addressed in that field of science present highly abstract definitions [2]-[4], that may also prove to be quite unstable, as they are primarily influenced by



particular socio-historical contexts in which the concepts are used [2], [8]. Given their very abstract nature, such concepts' definitions are also typically difficult to operationalize; thus, even when a certain definitional consensus is achieved, it often remains unclear how to empirically assess (and, therefore, measure) the concept at hand [3], [4].

Democracy measurement deals with those challenges. Despite many fluctuations in the way democracy has been defined throughout the millenniums [16], at the core of the concept, there is substantial agreement over its general meaning as “rule by the people” [4], [17]-[19]. Still, the definition of democracy in more specific terms remains the subject of much debate, and consequently, so does its measurement [4], [18]-[21].

Measuring democracy may offer valuable insights for both academic and decision-making activities. Moreover, given its effect on public opinion, empirical assessments stipulating variations on the level of democracy across space and time may significantly impact local, national, and international political relations. Thus, providing meaningful and comparable measures of democracy is of utmost importance. For that purpose, fundamental metrological directions applicable to measurement in all fields of scientific knowledge should be considered [9], [13]-[15].

Since the second half of the twentieth century [22], multiple approaches have been proposed to measure the democratic quality of a given country at a given time [4], [17]-[19], [21], [23]-[37]. So far, several studies have also delved into the quality of the measurements performed by such democracy indices [21], [36]-[63]. Whether those affairs adequately tackle fundamental requirements of Measurement Science is a question that remains unanswered; in fact, the literature on democracy measurement lacks any study with explicit reference to the Metrological discipline.

This study aims to address the mentioned knowledge gap by delving into the literature on extant democracy measurement systems and the methods used to assess the quality of their results and investigating the successes and limitations in aligning the measurement of democracy with the fundamental requirements of Measurement Science.

Following this introduction, the paper's second section outlines the various approaches to measuring democracy. The methods used to evaluate the quality of democracy measurement results are discussed in the third section, along with insights drawn from the literature on Metrology applied to Human and Social Sciences. Finally, the fourth section presents the conclusions from the analysis.

2. Democracy measuring systems

Some of the most prominent currently available democracy indices are briefly examined in this section [17], [19], [23]-[35]. While not an exhaustive list of democracy measurement systems, this selection of indices is sufficient to illustrate differences between those measuring approaches. Table 1 elaborates on such differences, as each index is characterized according to relevant parameters.

The definition of democracy's key observable features, conditioning the choice of indicators, is a major point of debate [4], [18]-[21]. A common way to distinguish between these varying conceptual approaches is to assess the breadth of elements they encompass as defining features of a democratic regime [4], [20], [21]. Accordingly, indices guided by “minimalist” conceptions take few observable variables into account, whereas those aligned with “maximalist” approaches consider a broader range of aspects as indicators of democratic quality (see Table 1). Meanwhile, indices such as Democracy Barometer (DB) have been described as falling between minimalist and maximalist conceptual approaches [24] –being depicted as “medium” in Table 1. On the other hand, certain democracy measurement systems, like the Polity 2 index, received contrasting characterizations, with its conceptual framework being labelled as “minimalist” by some [20], [38], and “maximalist” by others [21]; thereby exposing substantial theoretical disagreements on what democracy should include and which variables underlie each indicator.

Table 1. Characterization of democracy measuring systems in terms of relevant methodological issues

Index (democracy measuring system)	Operationalized Concept of Democracy ^a	Aggregation Methods ^a	Data collection method					Interpretation given to Results in terms of Scale-Type ^a
			factual data	mass surveys	in-house coders	country experts	extant indices	
Boix-Miller-Rosato (BMR) index [23]	Minimalist	Multiplicative ^b			X			Ordinal (2 categories)
Democracy Barometer (DB) [24]-[26]	Medium	Additive and Multiplicative	X	X			X	Interval [18]
Democracy-Dictatorship (DD) index [27], [28]	Minimalist	Multiplicative ^b	X		X			Ordinal (2 categories)
Freedom House (FH) status of freedom ^c [29]	Maximalist [20], [21]	Additive			X	X		Ordinal (3 categories)
Lexical Index of Electoral Democracy (LIED) [4], [17]	Minimalist	Multiplicative ^b	X		X			Ordinal (7 categories)
Polity 2 index [30]	Minimalist for some [20], [38] Maximalist for others [21]	Additive			X			Discreet distribution from -10 to 10
Unified Democracy Score (UDS) [31]	Maximalist [21]	Bayesian latent variable model					X	Interval
Vanhanen's index [32]	Minimalist [20], [21]	Multiplicative	X		X			Interval [18]
electoral democracy	Maximalist	various	X		X	X	X	Interval
participatory democracy	Maximalist	various	X		X	X	X	Interval
deliberative democracy	Maximalist	various	X		X	X	X	Interval
egalitarian democracy	Maximalist	various	X		X	X	X	Interval
liberal democracy	Maximalist	various	X		X	X	X	Interval

^a Along this column, cells where references are not mentioned have information based on the sources indicated in the corresponding row of the first column – i.e., relative to each democracy index.

^b The scoring method used by this index is classified as ‘multiplicative’ because indicators are treated as necessary conditions for achieving a given score, even though multiplication is not involved.

^c Despite focusing on "freedom" rather than democracy *per se*, the main index provided by the Freedom House organization (as part of their "Freedom in the World" annual reports) has been incorporated in this overview due to its widespread recognition as a measure of democracy.

^d The V-Dem project offers five separate democracy indices (electoral, liberal, participatory, egalitarian, and deliberative), whereas the electoral democracy index is constitutive of all the other four indices.

Differences among democracy measurement systems further extend to the aggregation function, the procedure used to combine several indicators into a single index. Analogous to a measurement model, which is defined in the International Vocabulary of Metrology (VIM 3) as the “mathematical relation among all quantities known to be involved in a measurement” (JCGM 200:2012, section 2.48) [1], an aggregation function relies on a theoretical understanding of the measurand and its relation with other

properties. The literature on the aggregation strategies applied to democracy measurement commonly distinguishes between two theoretical approaches underpinning different mathematical procedures [18], [19], [21], [34], [35], [63]. On the one hand, additive aggregation techniques, like sums, averages, and weighted averages, are aligned with the idea that indicators are mutually compensating and, therefore, partially substitutable in democracy measurement. On the other hand, if indicators are understood as interdependent or necessary conditions for democracy, multiplicative aggregation procedures are justified. Some democracy indices rely on both strategies: for instance, DB applies an additive approach at lower aggregation levels and a multiplicative formula at higher levels [26], while V-Dem combines both techniques within a single function, obtaining their main democracy indices through an equally weighted average between the results from an additive and a multiplicative approach [19]. Item Response Theory and Bayesian factor analysis are additionally employed in the construction of V-Dem's indices [19]. Likewise, UDS uses a Bayesian latent variable model to aggregate several democracy indices into a single index, assuming they all provide approximations of the same underlying quantity [31].

Table 1, moreover, illustrates the different methods for data collection on democracy measurement. Some indicators use 'factual' data, such as the share of the population with the right to vote. Others present ratings based on value judgments, either provided by consulted experts or by members of the measurement project (referred to as 'in-house coders' [18]). Also centred on subjective data, a less frequent approach is that of mass surveys aiming to capture citizens' behaviours and opinions that might signal democratic or non-democratic attributes in their daily life experiences [18], [22], [25]. Moreover, several democracy measurement systems incorporate data from other available indices [18], [19], [26], [31].

Finally, democracy measurement systems report results in varying ways. To capture these different manners of conveying information, Table 1 employs the classic distinction of scale types proposed by Stevens [64], who has differentiated between nominal, ordinal, interval, and ratio scales, each corresponding to different forms of representing and thus interpreting observations. In a nominal scale, observed events are classified into types with no inherent hierarchy among them, whereas an ordinal scale has categories arranged in order, indicating variations in a given underlying property [5], [64]. Interval scales take a step further by preserving a consistent measurement unit across the scale and thus providing meaningful information on the differences between measurand values. Ratio scales additionally present absolute zeros that indicate the absence of the quantity being measured. Achieving measurement in a ratio scale is deemed likely beyond current capabilities for measuring most psychosocial phenomena [5]; therefore, efforts to enhance measurement in the Humanities and Social Sciences usually aim at providing information on the interval level. Nonetheless, the assumption that a particular dataset falls under an interval-level scale is often taken for granted [5].

Table 1 shows the scale types on which each democracy measurement system assumes, either explicitly or implicitly, to present its results but does not go into the merits of the validity of these assumptions. Some democracy measurement systems, like LIED and V-Dem, engage in notable discussions on the appropriate method for conveying information on specific types of scale [4], [17], [19]. The assumption of scale type for Polity 2's measurement results, on the other hand, remains ambiguous since the scores are given in the form of a discrete distribution of integers from -10 to +10 [30], without there being a clear indication as to whether these numbers should be interpreted as categories on an ordinal scale or as equidistant threshold values along an interval scale. Recent changes in the reporting methods of some democracy measurement projects are also worth noting, as is the case with DB, which has recently abandoned its fixed 0 to 100 scale, leaving it to researchers to decide the scaling criteria according to their specific purposes [26].

3. Measurement Science and Validity of Democracy Measurement

For promoting the quality of measurement aimed at properties across all fields of knowledge, recent studies in Measurement Science suggest object-relatedness (objectivity) and subject-independence

(intersubjectivity) as two fundamental ideals that ought to be pursued [8], [9], [13]. Object-relatedness comprises the extent to which the information obtained through measurement reflects only variations on the measured property [8], [9], [13]. Enhancing objectivity thus requires a solid theory about the measurand and a reduced influence from other phenomena, minimizing definitional and instrumental uncertainty [8], [9], [13]. Intersubjectivity, on the other hand, strives for consistent interpretation of measurement results by individuals across various locations and times. Such a goal requests metrological traceability of measurement results to a shared reference scale [8], [9], [13].

Intimately related to the matter of subject-independence of measurement results, concerns about comparability frequently arise in the literature on democracy measurement, with different measuring systems being said to yield conflicting empirical findings [22], [27], [53], [60], [61]. Likewise, conceptual disagreements on the way democracy is defined have been outlined [18]-[21], [38], [63], [65], and it has been emphasized the need for democracy measurement systems to rely on a definition of democracy whose applicability across space and time is explicitly justified [22].

Concerns about objectivity in democracy measurement are also evidenced as scholars draw attention to potential bias in the way democracy has been defined, which is likely to echo international asymmetries of power [22], [66]-[68]. In a similar vein, there is a growing call to move beyond strictly institutional definitions and embrace more inclusive and socially oriented conceptions of democracy [22], [65], [69]-[71]. Researchers have also questioned the feasibility of representing a particular operationalized concept of democracy as a singular variable, defined along a single (unidimensional) measurement scale [20], [36], [38], [72]-[76]. Discussions regarding the impacts of methodological choices over measurement error [20], [77], alongside attempts to investigate and minimize those errors [18], [21], [31], [51], [52], [58], [59], are likewise noticeable in the literature on democracy measurement.

When assessing the quality of measurement of psychosocial quantities, a commonly used concept is that of measurement validity, which states whether a measuring instrument, such as questionnaires or indicators, effectively measures the property it purports to measure [39], [40], [56]. [39] identifies four main traditions of approaches to measurement validity in Political Science, for each of which examples are given in the field of democracy measurement. The four traditions are the case-based method, the pragmatic tradition, the structural equation modelling with latent variables (SEM-L), and the levels of measurement approach (LoM).

The case-based method involves conducting in-depth case studies to gather more knowledge and evaluate whether scores accurately capture the realities they represent [39]. Studies assessing the validity of democracy scores through this approach include [41]-[48]. Although this method is valuable for understanding specific cases and contributing to relevant conceptual and methodological insights, it cannot, alone, enable systematic comparisons across various contexts [39].

The pragmatic tradition relies on straightforward statistical techniques such as correlation and regression analysis [39]. Within this tradition, indicators' performances are evaluated based on immediate application purposes and regardless of any general measurement model [39]. In the field of democracy measurement, methods aligned with the pragmatic tradition have been used by [48], [49], [58], [60]. Although fruitful for exploratory studies, the pragmatic tradition for assessing measurement validity has drawn great scepticism, highlighting the lack of attention to the connections between the measured property and indication values [39]. Such a characteristic undermines object-relatedness and subject-independence of measurement results, as it limits the possibility of meaningful comparisons across different measurement contexts.

Studies evaluating democracy measurement based on the Structural Equation Modelling with Latent Variables (SEM-L) approach [51]-[53], [56]-[58], on the other hand, apply sophisticated statistical models for aggregating indicators and assessing measurement error [39]. SEM-L plays a central role in the history of democracy measurement since the extensive use of that strategy following the work of Bollen [57], from 1980, paved the way for the methodological sophistication of democracy measurement practice [22]. Nonetheless, studies applying this kind of approach have often been criticized for relying

on untested assumptions concerning, for instance, the dimensionality of the operationalized concept and the sort of distribution the measured property might present beyond the specific measurement sample [4], [39].

Finally, the Levels of Measurement tradition (LoM) brings up concerns about the scale type in which a given data can be reliably interpreted, as this limits the statistical techniques and mathematical operations that can be meaningfully applied to the data [20], [39]. Studies developed under this tradition typically attempt to transform data for incorporating higher levels of measurement (scale types), thus broadening the range of applicable statistical techniques [39]. The Levels of Measurement tradition provides a path to developing strategies based on criteria that can be aligned with the requirements used for measurements in the physical sciences. Methods used by this tradition include the Guttman scale analysis, Item Response Theory (IRT), and the Rasch Measurement Theory.

Guttman scale analysis seeks to provide ordinal measurement based on nominal (qualitative) observations [39], [78], [79]. Following the strategy based on ‘ideal-types’ proposed by the German sociologist Max Weber [2], Guttman developed a model embodying the conditions of “perfect” measurement [13], [78], [79], posing unidimensionality as a central requisite. Accordingly, in a Guttman scale, a series of observable attributes, typically assessed by dichotomous indicators, is hierarchically distributed according to variations in a single underlying quantity [4], [39], [78]. This arrangement follows a cumulative pattern, meaning that an individual displaying any of these attributes is expected to display all lower-ranked (less “difficult”) attributes as well [4], [39], [78]. In the field of democracy measurement, researchers have utilized the Guttman scale for both the construction and evaluation of measurement systems [36], [37], [54]. Additionally, a review of democracy measurement studies has underscored the importance of the Guttman model in empirically testing the unidimensionality hypothesis of a resultant measurement scale [20].

The Guttman scale analysis operates under a deterministic measurement model, establishing a direct link between the property being measured (measurand) and the resulting measurement outcomes [4], [7]. Deterministic approaches, however, are often deemed less suitable in complex measurement contexts, such as those found in the Human and Social Sciences, prompting researchers to opt for probabilistic strategies instead [4], [7]. Rasch Measurement models and certain approaches in Item Response Theory (IRT) stand out as alternative tools in this scenario. By modelling the output of measurement as a probability distribution rather than a singleton, probabilistic approaches are useful for estimating meaningful differences along the measurement scale [7]. Thereby, while Guttman scale analysis is deployed for producing ordinal measures; IRT and Rasch probabilistic models are used to attain measurements on the interval-level [4], [7], [39].

Although the Item Response Theory and the Rasch measurement approach share similarities, they were developed independently from one another and have philosophical and methodological differences that are worth noting [81]. Primarily focused on describing idiosyncrasies of the data and explaining its variance, IRT methods used for providing interval-level data typically seek a model that best fits the data, incorporating one or more parameters designed to reflect characteristics of the sample [5], [11], [12], [82], [83]. In contrast, the Rasch Measurement approach tests the extent to which the data fits the model [5], [11], [12], [82], [83]. Featuring a probabilistic realization of the Guttman scale, the Rasch Model incorporates unidimensionality as a fundamental requirement [5], [84]. Local independence is another important requisite, according to which the response to any item should not exert substantial influence on the response to another [5], [9].

The basic form of the Rasch probabilistic model relies on two parameters, one for each item (indicator) of the measuring instrument and the other for the individual instance of the quantity under measurement. More sophisticated forms of that model, like the multi-facets Rasch model, incorporate additional parameters considering only those factors that are expected to influence measurement results in a systematic (measurable) way [5]. Even in those cases, the requirement of unidimensionality is maintained, since estimates for every parameter are placed along the same measurement scale –thus indicating magnitudes of the same underlying quantity– and this condition is empirically tested [5].

Overall scores are considered sufficient statistics for estimating the parameters' values as long as the data fits the model [5], [11]. Moreover, a unique feature of the Rasch model is that it allows for parameter separability, meaning that item parameters can be estimated independently from the parameters related to the individuals under measurement and vice-versa [5], [7]-[9], [85], [86]. This allows for measurement results to remain invariant, within a range of measurement uncertainty, across different measurement contexts [5], [8], [9], [79], thus allowing for intersubjectivity of measurement results [8].

With a measurement model that is not contingent on the specific characteristics of the sample to which it is applied, Rasch analysis allows for psychosocial measurement to meet the same requirements applied to the measurement of physical quantities, demanding that measurement results related to the same measured property be comparable independently of the measurement object and the measuring instrument that is used. For that reason, the Rasch Measurement Theory has been appointed by recent studies in the field of Measurement Science as an ideal infrastructure for supporting metrological traceability in the Social Sciences [8], [10], [12], [13], [87]-[89].

In the literature applying methods identified with the LoM tradition to assess the validity of democracy measurement, IRT methods have been adopted [19], [35], [52]. In contrast, the use of the Rasch model is practically absent, except for a study that focuses on assessing the validity of "political trust" measures [90], performed on an available database of widely used cross-national surveys (World Values Survey, Afrobarometer, Arabbarometer, Asian Barometer, Eurobarometer, European Social Survey, and Latinobarometro). The data provided by those cross-national surveys also comprise part of the sources employed by the Democracy Barometer measuring system [26]. Despite not delving into metrological concepts, such as the potential of Rasch modelling to reach intersubjectivity by providing metrological traceability to measurement results, [90] evaluates unidimensionality, equivalence, and item hierarchy of political trust, revealing the lack of cross-national correspondence of political trust measurement results. These outcomes corroborated with predominant views on the theoretical literature, indicating the non-unidimensionality of 'political trust' and contradicted political trust measurement practices, which typically assumed the unidimensionality of the construct. Accordingly, [90] points to the need for higher consistency and robustness in data analysis, demonstrating some of the valuable elements of the promising feature of the Rasch Theory approach.

4. Conclusion

Considering the existing alternatives described in the literature to meet metrological requirements regarding measurements associated with measurands from the complex field of Human and Social Sciences, this paper explored the literature on democracy measurement to investigate possible advances or yet-to-be-tackled elements to provide reliable and comparable democracy measurement results.

Although with no explicit reference to metrology as such, this study evidenced fundamental concerns with elements from the Science of Measurement manifest throughout the discussions concerning the measurement approaches of democracy. Several researches conducted on democracy measurement have highlighted problems of comparability. The attention to issues such as random error and systematic bias is also present, with various scholars advocating for more accurate measurement results of democracy and a less biased comprehension of that measured concept. Different strategies have been used to assess and improve the quality of democracy measuring systems. Still, important aspects remain to be tackled for aligning those approaches with fundamental metrology precepts.

The Rasch measurement model is a standout tool when determining measurement validity in the humanities and social sciences. It offers measurements in a probabilistic manner, which is more suitable for assessing psychosocial phenomena. Additionally, it allows for comparable results when measuring the same property, regardless of the sample or the measurement system used.

Although it has not been used for measuring democracy, a preliminary study using Rasch's method regarding the property of 'political trust' has already identified some promising contributions, demonstrating its potential for investigating theoretical conceptions.

Further application of the Rasch model can contribute to theoretical advances in the concept of democracy and is required to step forward towards providing objectivity and intersubjectivity, hence, metrological traceability in measuring democracy.

Acknowledgments

This study was financed in part by CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) Finance Code 001 and FAPERJ (*Fundação Carlos Chagas Filho Amparo Pesquisa do Estado do Rio de Janeiro*).

References

- [1] JCGM 200:2012 2012 *International Vocabulary of Metrology – basic and general concepts and associated terms, 3rd edition* (Paris: Joint Committee for Guides in Metrology)
- [2] Weber M, Faber-Kaiser M 1971 *Sobre la teoría de las ciencias sociales* **73** (Barcelona: Península)
- [3] Plümper T, Neumayer E 2012 *Model uncertainty and robustness tests: Towards a new logic of statistical inference*
- [4] Gerring J, Pemstein D, Skaaning S-E 2021 *Sociological Methods & Research* **50** 778-811
- [5] Bond T G and Fox C M 2020 *Aplicação do modelo de Rasch* (Hogrefe)
- [6] Pendrill L 2014 *NCSLi Measure* **9** 24-35
- [7] Mari L and Wilson M 2014 *Measurement* **51** 315-327
- [8] Maul A, Mari L and Wilson M 2019 *Measurement* **131** 764-770
- [9] Mari L, Wilson M and Maul A 2023 *Measurement across the sciences: Developing a shared concept system for measurement* (Springer Nature)
- [10] Pendrill L 2020 *Quality Assured Measurement: Unification Across Social and Physical Sciences* (Springer)
- [11] Salzberger T, Cano S, Abetz-Webb L, Afolalu E, Chrea C, Weitkunat R, Rose J 2021 *Measurement* **181** 109593
- [12] Fisher W P Jr and Cano S J 2022 *Person-Centered Outcome Metrology: Principles and Applications for High Stakes Decision Making* (Springer Nature)
- [13] Monteiro Vieira C and Costa Monteiro E 2023 *Acta IMEKO* **12** 1-6
- [14] Costa Monteiro E 2017 *Measurement: Interdisciplinary Research and Perspectives* **15** 34-36
- [15] Costa Monteiro E and Summers R 2022 *Measurement: Sensors* **24** 100574
- [16] Dunn J 2018 *Setting the people free: The story of democracy* (Princeton University Press)
- [17] Skaaning S-E, Gerring J, Bartusevičius H A 2015 *Comparative Political Studies* **48** 1491-1525
- [18] Coppedge M, Gerring J, Lindberg S I, Skaaning S E and Teorell J 2017 *V-Dem working paper* **45**
- [19] Coppedge M *et al.* 2023 V-Dem methodology v13 *Varieties of Democracy (V-Dem) Project*
- [20] Munck G L and Verkuilen J 2002 *Comparative political studies* **35** 5-34
- [21] Gründler Klaus, Krieger Tommy 2021 *European Journal of Political Economy* **70** 102047
- [22] Giebler H, Ruth S P and Tanneberg D 2018 *Politics and Governance* **6** 1-10
- [23] Boix C, Miller M and Rosato S 2013 *Comparative political studies* **46** 1523-1554
- [24] Bühlmann M, Merkel W and Wessels B 2008 *The quality of democracy: Democracy barometer for established democracies* (Zurich: NCCR Democracy)
- [25] Bühlmann M, Merkel W, Müller L and Weßels B 2012 *European Political Science* **11** 519-536
- [26] Engler S *et al.* 2020 *Democracy Barometer Codebook version 7* (Aarau: Zentrum der Demokratie)
- [27] Cheibub J A, Gandhi J, Vreeland J R 2010 *Public choice* 67-101
- [28] Bjørnskov C and Rode M 2020 *The Review of International Organizations* **15** 531-551
- [29] House Freedom 2020 methodology *Freedom in the World*
- [30] Marshall, M 2020 *Polity5: Dataset Users' Manual* (Center of Systematic Peace)
- [31] Pemstein D, Meserve S A and Melton J 2010 *Political Analysis* **18**
- [32] Vanhanen T 2000 *Journal of peace research* **37** 251-265
- [33] Lindberg S I, Coppedge M, Gerring J and Teorell J 2014 *Journal of Democracy* **25** 159-169
- [34] Teorell J, Coppedge M, Skaaning S E and Lindberg S I 2016 *V-Dem Working Paper* **25**

- [35] Coppedge M, Lindberg S, Skaaning S-E, Teorell J 2016 *International Political Science Review* **37** 580-593
- [36] Coppedge M, Reinicke W H 1990 *Studies in Comparative International Development* **25** 51-72
- [37] Baker P J and Koesel K J 2001 *Annual meeting of the American Political Science Association* (San Francisco)
- [38] Gleditsch K S and Ward M D 1997 *The Journal of Conflict Resolution* **41** 361–383
- [39] Seawright J, Collier D 2014 *Comparative Political Studies* **47** 111-138
- [40] Adcock R and Collier D 2001 *American political science review* **95**(3) 529-546
- [41] Yashar D J 1997 *Demanding democracy: Reform and reaction in Costa Rica and Guatemala, 1870s–1950s* (Cambridge University Press)
- [42] Mahoney J 2001 *The legacies of liberalism: Path dependence and political regimes in Central America* (Baltimore: Johns Hopkins University Press)
- [43] Collier D 1999 *Newsletter of the Organized Section for Comparative Politics of the American Political Science Association* **10** 1-2, 4-6
- [44] Rueschemeyer D, Stevens E H, Stevens J D 1992 *Capitalist development and democracy* (University of Chicago Press)
- [45] Bowman K, Lehoucq F and Mahoney J 2005 *Comparative Political Studies* **38** 939-970
- [46] O'Donnell G 1996 *Journal of Democracy* **7** 34-51
- [47] Collier D and Levitsky S 1997 *World Politics* **4** 430-451
- [48] Mainwaring S, Brinks D, Pérez-Liñán A 2001 *Studies in Comparative International Development* **36** 37-65
- [49] Przeworski A, Alvarez M E, Cheibub J A, Limongi F 2000 *Democracy and development: Political institutions and well-being in the World, 1950–1990* (Cambridge University Press)
- [50] Casper G and Tufis C 2003 *Political Analysis* **11** 196-203
- [51] Shen C and Williamson J B 2005 *International Journal of Comparative Sociology* **46** 327-345
- [52] Treier S and Jackman S 2008 *American Journal of Political Science* **52** 201-217
- [53] Elff M and Ziaja S 2018 *Politics and Governance* **6**(1) 92–104
- [54] Møller J and Skaaning S E 2010 *International Political Science Review* **31**(3) 261-283
- [55] Coppedge M and Gerring J 2011 *Perspectives on Politics* **9** 247-267
- [56] Bollen K A 1989 *Structural equations with latent variables* (New York: John Wiley)
- [57] Bollen K A 1980 *American Sociological Review* **45**(3) 370–390
- [58] Bollen K A 1993 *American Journal of Political Science* **37** 1207-1230
- [59] Bollen K A and Paxton P 2000 *Comparative Political Studies* **33** 58-86
- [60] Elkins Z 2000 *American Journal of Political Science* **44** 287-294
- [61] Vaccaro A 2021 *European Political Science* **20** 666-684
- [62] Casper G and Tufis C 2003 *Political Analysis* **11** 196-203
- [63] Boese V A 2019 *International Area Studies Review* **22**(2) 95-127
- [64] Stevens S S 1946 *Science* **103**(2684) 677-680
- [65] Landman T 2018 *Politics and Governance* **6** 48–59
- [66] Koelble T A and Lipuma E 2008 *Democratisation* **15** 1-28
- [67] Grugel J 2003 *Government and Opposition* **38** 238-264
- [68] Brooks H, Ngwane T and Runciman C 2020 *The Sociological Review* **68** 17-32
- [69] Fuchs D and Roller E 2018 *Politics and Governance* **6**(1) 22–32
- [70] Mayne Q and Geißel B 2018 *Politics and Governance* **6**(1) 33–47
- [71] Fleuß D, Helbig K and Schaal G S 2018 *Politics and Governance* **6**(1) 11–21
- [72] Coppedge M, Alvarez A and Maldonado C 2008 *The Journal of Politics* **70**(3) 632–647
- [73] Fishman R M 2016 *Annual Review of Political Science* **19**(1) 289–309
- [74] Lauth H-J 2015 *The matrix of democracy: A three-dimensional approach to measuring the quality of democracy and regime transformations* (Würzburg: Universität Würzburg)
- [75] Coppedge M and Reinicke W H 1991 *On measuring democracy: Its consequences and*



- concomitants*, ed A Inkeles (New Brunswick: Transaction) chapter 3 pp 47-68
- [76] Lauth H-J and Schlenkrich O 2018 *Politics and Governance* **6**(1) 78–91
- [77] Skaaning S-E 2018 *Politics and Governance* **6** 105–116
- [78] Guttman L 1944 *American sociological review* **9** 139-150
- [79] Engelhard Jr G 2013 *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences* (Routledge)
- [80] Duncan O D 1984 *Notes on Social Measurement: Historical and Critical* (Russell Sage Foundation)
- [81] Andrich D 2004 Controversy and the Rasch model: a characteristic of incompatible paradigms?. *Medical care* I7-I16
- [82] McGrane J A and Maul A 2020 *Measurement* **152** 107346
- [83] Aryadoust V, Ng L Y and Sayama H 2021 *Language Testing* **38**(1) 6-40
- [84] Wright B D and Stone M 1999 *Measurement essentials*
- [85] Tutz G 2023 *Invariance of Comparisons: Separation of Item and Person Parameters beyond Rasch Models*
- [86] Lord F 2010 *Rasch Measurement Transactions* **24** 1289-1290
- [87] Cano S J, Melin J, Fisher W P, Stenner A J, Pendrill L R and EMPIR NeuroMet 15HLT04 Consortium 2018 *Journal of Physics: Conference Series* **1065** 072033
- [88] Pendrill L and Petersson N 2016 *Measurement Science and Technology* **27** 094003
- [89] Fisher Jr W P and Massengill P J 2023 *Explanatory Models, Unit Standards, and Personalized Learning in Educational Measurement: Selected Papers by A Jackson Stenner* (Springer Nature)
- [90] Van Der Meer T W G Ouattara E 2019 *Quality & Quantity* **53** 2983-3002