



V-Dem Comparisons and Contrasts with Other Measurement Projects

Michael Coppedge, John Gerring,
Staffan I. Lindberg, Svend-Erik Skaaning,
Jan Teorell

April 2017

Working Paper

SERIES 2017:45

THE VARIETIES OF DEMOCRACY INSTITUTE



UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a new approach to the conceptualization and measurement of democracy. It is co-hosted by the University of Gothenburg and University of Notre Dame. With a V-Dem Institute at University of Gothenburg that comprises almost ten staff members, and a project team across the world with four Principal Investigators, fifteen Project Managers, 30+ Regional Managers, 170 Country Coordinators, Research Assistants, and 2,500 Country Experts, the V-Dem project is one of the largest-ever social science research-oriented data collection programs.

Please address comments and/or queries for information to:

V-Dem Institute
Department of Political Science
University of Gothenburg
Sprängkullsgatan 19, PO Box 711
SE 40530 Gothenburg
Sweden
E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at www.v-dem.net.

Copyright © 2017 by authors. All rights reserved.

V-Dem Comparisons and Contrasts with Other Measurement Projects*

Michael Coppedge
Professor of Political Science
University of Notre Dame

John Gerring
Professor of Political Science
University of Texas at Austin

Staffan I. Lindberg
Professor of Political Science
Director, V-Dem Institute
University of Gothenburg

Svend-Erik Skaaning
Professor of Political Science
Aarhus University

Jan Teorell
Professor of Political Science
Lund University

* This research project was supported by Riksbankens Jubileumsfond, Grant M13-0559:1, PI: Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden; by Knut and Alice Wallenberg Foundation to Wallenberg Academy Fellow Staffan I. Lindberg, Grant 2013.0166, V-Dem Institute, University of Gothenburg, Sweden; as well as by internal grants from the Vice-Chancellor's office, the Dean of the College of Social Sciences, and the Department of Political Science at University of Gothenburg. We performed simulations and other computational tasks using resources provided by the Notre Dame Center for Research Computing (CRC) through the High Performance Computing section and the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre in Sweden, SNIC 2016/1-382 and 2017/1-68. We specifically acknowledge the assistance of In-Saeng Suh at CRC and Johan Raber at SNIC in facilitating our use of their respective systems.

Introduction

In the wake of the Cold War democracy has gained the status of a mantra.¹ However, no consensus has emerged about how to conceptualize and measure this key concept. Skeptics may wonder whether such comparisons are even possible. Distinguishing the most democratic countries from the least democratic ones is fairly easy: Almost everyone agrees that Switzerland is democratic and North Korea is not. It has proven to be much harder to make finer distinctions: Is Switzerland more democratic than the United States? Is Russia less democratic today than it was last year? Has Venezuela become more democratic in some respects and at the same time less democratic in others? Yet, if we cannot measure democracy in some fashion we cannot mark its progress and setbacks, explain processes of transition, reveal the consequences of those transitions, and affect their future course.

For policymakers, activists, academics, and citizens around the world the conceptualization and measurement of democracy matters. Billions of dollars in foreign aid spent every year for promoting democracy and governance in the developing world are contingent upon judgments about a polity's current status, its recent history, future prospects, and the likely causal effect of particular forms of assistance. A large body of social science work deals with these same issues. The needs of democracy promoters and social scientists are convergent. We all need better ways to measure democracy.

In the first section of this document we critically review the field of democracy indices. It is important to emphasize that problems identified with extant indices are not easily solved, and some of the issues we raise vis-à-vis other projects might also be raised in the context of the V-Dem project. Measuring an abstract and contested concept such as democracy is hard and some problems of conceptualization and measurement may *never* be solved definitively. The discussion in this section is thus not intended to debunk or otherwise de-legitimate the use of any of the indices discussed therein (which includes several indices developed by members of the V-Dem team). Instead, its purpose is to illustrate why the current crop of democracy indices is not sufficient to solve all our measurement needs.

In the second section we discuss in general terms how the *Varieties of Democracy* (V-Dem) project differs from extant indices and how the novel approach taken by V-Dem might assist the work of activists, professionals, and scholars. Appendix A addresses the possible uses of V-Dem

¹ This paper integrates material previously published in Coppedge et al. (2011) and Lindberg et al. (2014). The authors thank the many people who have generously provided comments and feedback as this document evolved. This includes Hauke Hartmann, Monte Marshall, Wolfgang Merkel, Arch Puddington, and Dag Tanneberg.

for program evaluation, Appendix B offers a glossary of key terms used in this document, and Appendix C clarifies search terms used for several analyses in Table 1.

I. Extant Indices

Many attempts have been made to measure democracy. It is somewhat complicated to identify these indices because they do not always refer explicitly to democracy. Nonetheless, we include an index in our survey if it is commonly viewed as representing democracy or some aspect of democracy, and if it features fairly broad country coverage. This includes the BNR index developed by Bernhard, Nordstrom & Reenock (2001); the Bertelsmann Transformation Index (“BTP”) directed by the Bertelsmann Stiftung (various years); the Democracy Barometer developed by Wolfgang Merkel & associates (Bühlmann, Merkel, Müller & Weßels 2012); the BMR index developed by Boix, Miller & Rosato (2013), the Contestation and Inclusiveness indices developed by Coppedge, Alvarez, & Maldonado (2008); the Political Rights, Civil Liberty, Nations in Transit, and Countries at the Crossroads indices, all sponsored by Freedom House (freedomhouse.org); the Economist Intelligence Unit (2010) index (“EIU”); the Unified Democracy Scores (“UDS”) developed by Pemstein, Meserve & Melton (2010); the Polity2 index from the Polity IV database (Marshall, Gurr & Jaggers 2014); the democracy-dictatorship (“DD”) index developed by Adam Przeworski & colleagues (Alvarez, Cheibub, Limongi & Przeworski 1996; Cheibub, Gandhi & Vreeland 2010); the Lexical index of electoral democracy developed by Skaaning, Gerring & Bartusevičius (2015); the Competition and Participation indices developed by Tatu Vanhanen (2000); and the Voice and Accountability index developed as part of the Worldwide Governance Indicators (“WGI”) (Kaufmann, Kraay & Mastruzzi 2010).²

In order to make comparisons in a systematic fashion we summarize key features of these indices, along with V-Dem, in Table 1. Note that an *index* is understood as a highly aggregated, composite measure of democracy such as Polity2, while an *indicator* is understood as a measure of a more specific, disaggregated attribute of democracy such as turnout. (See Appendix B for more precise definitions of these and other terms used in this document.)

² Other indices, not included in Table 1, may be briefly listed: the Political Regime Change [PRC] dataset (Gasiorowski 1996; updated by Reich 2002), the Democratization Dataset (Schneider and Schmitter (2004a), and indicators based on Altman and Pérez-Liñán (2002), Arat (1991), Bollen (1980, 2001), Bowman, Lehoucq, and Mahoney (2005), Hadenius (1992), Mainwaring, Brinks, and Pérez-Liñán (2001), and Moon et al. (2006).

In Table 1, data sources are categorized as (a) extant indices (democracy indices collected from other sources that become components in a broader index), (b) factual data (requiring little coder judgment and obtainable from primary or secondary sources), (c) mass surveys (of the general public or selected publics, e.g., business persons), (d) in-house coders (often research assistants or staff employed by the project), or (e) country experts (often academics or professionals working in some area closely related to democracy and governance), who code one or several countries according to expertise. For the latter, we note the number (N) of independent expert coders employed to code each country/year/indicator. (If the work of multiple coders is not conducted independently, or is subject to revision by others working on a project – as is the case for the BTI and Freedom House – the number of coders is counted as one.) A wide array of data sources is evident across democracy projects, though only one project (V-Dem) utilizes multiple (independent) coders.

With respect to each project, we count the total number of indicators (K) collected from various sources. These serve as ingredients of higher-level indices, so K provides a clue as to how disaggregated the base-level evidence is. Some projects work at a highly aggregated level, producing only a single index. At the other extreme, the EUI offers several dozen indicators and V-Dem offers several hundred. We also note whether these underlying indicators are publicly available, and hence replicable. Some democracy projects do not allow end-users to access the components they use to arrive at summary indices. Others, such as the Civil liberties and Political rights indices, are available only for the past decade.

Next, we indicate whether reliability analyses are conducted at the indicator level. This includes traditional inter-coder reliability tests (e.g., the Lexical index) as well as more complex measures such as discrimination parameters and overall error rates from a measurement model (e.g., Pemstein, Meserve & Melton 2010). Only three of the listed projects utilize reliability analyses.³ For each index, we note the scale type and the range (minimum and maximum values). If a theoretical range for an index is clearly defined we regard this – rather than the range of observable values – as defining the min and max. If an interval index takes the form of a normalized (standardized) scale around the mean, we report this as a “Z score” since the observed min and max are not very informative. A mix of binary, ordinal, and interval scales are on display in Table 1.

For each index, we also note whether an estimate of uncertainty (aka reliability or precision) is provided. This may be based on inter-coder agreement and estimates of coder-

³ Polity has done so, but only for a single year (1999); see below.

reliability, on reliability across alternate measures for a core concept, or on other features of the data. Table 1 demonstrates that most indices are not accompanied by an estimate of uncertainty. The coverage of each index is gauged according to the number of sovereign and semisovereign units included in the resulting dataset. This ranges from 29 (Nations in Transit, a regional index) to 200+ (effectively, all sovereign countries, and sometimes semisovereign units such as colonies as well). We also count the years across which each dataset extends, ranging from several decades to two centuries. All indices provide annual coverage, though not all indices are collected annually. Indeed, some projects do not have a schedule for regular (annual or semi-annual) updates, as noted.

The final section of Table 1 gauges the overall impact of the project by measuring the number of “hits” a project obtains from a search of selected key words, detailed in Appendix C. Readers should bear in mind that this approach suffers from false negatives and false positives. Moreover, these errors are unlikely to be evenly distributed across the various projects, being highly sensitive to the choice of keywords. Nonetheless, this is a well-established approach to measuring impact in areas where the outcome of interest is likely to be registered on the worldwide web or in publicly accessible databases. We measure impact on the general public by the number of hits in a Google search, as shown in the first column. Long-standing projects such as Freedom House and Polity lead the way. We measure academic impact by the number of hits in a Web of Science search. Here, Polity is the clear frontrunner – though DD, WGI, and Freedom House can also claim considerable impact on scholarly work.

Our goal in this table, as well as in the subsequent discussion, is not to provide a comprehensive survey of this extremely complex – and extremely crowded – field.⁴ It is, rather, to identify a few of the persistent problems affecting the conceptualization and measurement of democracy. Some of these problems apply to all extant indices, while others apply to a subset, as suggested by Table 1. We turn now to a focused discussion of six key issues: *definition*, *sources*, *disaggregation*, *coverage*, *discrimination*, *aggregation*, and general considerations pertaining to *validity and reliability*.

⁴ Detailed surveys covering the conceptualization and measurement of democracy can be found in Hadenius and Teorell (2005), Landman (2003), and Munck and Verkuilen (2002). See also Acuña-Alfaro (2005), Beetham (1994), Berg-Schlosser (2004a, 2004b), Bollen (1993), Bollen and Paxton (2000), Bowman, Lehoucq, and Mahoney (2005), Casper and Tufis (2003), Coppedge, Alvarez, and Maldonado (2008), Elkins (2000), Foweraker and Krznaric (2000), Gleditsch and Ward (1997), Lindberg (2006), McHenry (2000), Munck (2009), Munck and Verkuilen (2002), Pickel, Stark and Breustedt (2015), Seawright and Collier (2014), Treier and Jackman (2008), and Vermillion (2006). For work focused more generally on governance indicators, see Arndt and Oman (2006), Besancon (2003), Kurtz and Schrank (2007), Sudders and Nahem (2004), Thomas (2010), and USAID (1998).

Table 1: Democracy Indices Compared

	DATA SOURCES					INDICATORS			INDICES			COVERAGE			IMPACT	
	Extant indices	Factual data	Mass surveys	In-house coders	Country experts(N)	K	All data available	Reliability analysis	Scale	Range	Uncertainty	Countries	Years	Regular up-dates	Google	Google Scholar
Bernhard et al. <i>BNR</i>				✓		1	✓		Binary	0/1		124	1913-2010		1,850	134
Bertelsmann <i>BTI</i>					1	18	✓		Interval	0-10		129	2003-	✓	23,500	227
Boix et al. <i>BMR</i>				✓		1	✓		Binary	0/1		208	1800-2007		91	114
Coppedge et al. <i>Inclusiveness</i>	✓					6	✓		Interval	Z score	✓	197	1950-2000		2,160	161
<i>Contestation</i>	✓					6	✓		Interval	Z score	✓	197	1950-2000		1,460	161
EIU <i>EIU index</i>			✓		1	60			Interval	0-10		167	2006-	✓	5,700	156
Freedom House <i>Civil Liberties</i>					1	15			Ordinal	1-7		202	1972-	✓	200,000	1,560
<i>Countries@Crossroads</i>					1	17	✓		Ordinal	0-7		70	2004-2012		16,400	28
<i>Nations in Transit</i>					1	7	✓		Ordinal	1-7		29	1995-	✓	52,200	553
<i>Political Rights</i>					1	10			Ordinal	1-7		202	1972-	✓	167,000	1,560
Merkel et al. <i>Demo Barometer</i>	✓	✓	✓			105	✓		Interval	0-100		70	1990-	✓	5,730	173
Pemstein et al. <i>UDS</i>	✓					11	✓	✓	Interval	Z score	✓	198	1946-2012		851	262
Polity IV <i>Polity2</i>				✓		6	✓		Ordinal	-10-10		182	1800-	✓	78,900	4,856
Przeworski et al. <i>DD</i>		✓		✓		1	✓		Binary	0/1		199	1946-2008		5,810	1,317
Skaaning et al. <i>Lexical</i>		✓		✓		6	✓	✓	Ordinal	0-6		224	1800-	✓	332	11
Vanhanen <i>Competition</i>		✓		✓		1	✓		Interval	0-100		203	1810-	✓	2,690	580
<i>Participation</i>		✓		✓		1	✓		Interval	0-100		203	1810-	✓	3,100	580
WGI <i>Voice&Accountability</i>	✓		✓			32	✓		Interval	Z score	✓	215	1996-	✓	14,700	1,645
V-Dem <i>[various]</i>	✓	✓		✓	5	~400	✓	✓	Various	Various	✓	173	1900-	✓	2,200	9

Definition

Democracy, understood in a very general way, means *rule by the people*. This common understanding claims a long heritage stretching back to the Classical age (Dunn 2005). All usages of the term also presume sovereignty. A polity must enjoy some degree of self-government in order for democracy to be realized.

Beyond these core definitional elements there is great debate. The debate covers both descriptive and normative aspects, i.e., what political regimes are and what they ought to be.⁵ Since definitional consensus is necessary for obtaining consensus over measurement, the goal of arriving at a single universally accepted measure of democracy is impossible so long as this great debate remains unresolved. Let us explore some of the consequences.

The Polity2 index rates the United States as fully democratic throughout the twentieth century and much of the nineteenth century. This may be a fair conclusion if one disregards, for example, the composition of the electorate—from which women, most blacks, and many poor people were excluded—in one's definition of democracy (Keyssar 2000; Paxton 2000). Similar challenges could be levied against other indices that focus narrowly on the electoral properties of democracy without much attention to suffrage (e.g., DD).

Other indices include attributes that fall far from the core meaning of the term. For example, the Civil Liberties index includes questions pertaining to civilian control of the police, the absence of widespread violent crime, willingness to grant political asylum, the right to buy and sell land, and the distribution of state enterprise profits (Freedom House 2007). The authors of the index would doubtless point out that it measures civil liberty, not democracy per se. Nevertheless, the Civil Liberties index is incorporated by Freedom House into a broader index of Freedom (comprised of the Civil Liberties and Political Rights indices, equally weighted) that is frequently regarded as synonymous with democracy.

A third definitional problem is ignoring the multidimensional nature of democracy. Because the great debate about the definition of democracy is unresolved (and, we argue, unresolvable), there are competing conceptions of democracy (Cunningham 2002; Held 2006; Møller & Skaaning 2013). To take a simple example, the EIU index regards mandatory voting as reflecting negatively on the quality of democracy in a country. While this provision can be said to

⁵ Studies of the concept of democracy are legion. See, e.g., Beetham (1994, 1999), Collier and Levitsky (1997), Held (2006), Lively (1975), Merkel (2004), Munck (2016), Naess et al. (1956), Saward (2003), and Weale (1999). All emphasize the far-flung nature of the core concept.

infringe upon individual rights and in this respect may be considered undemocratic, it also enhances participation (turnout) and hence, one could think of it as improving the quality of representation (Lijphart 1997). Its status in enhancing rule by the people therefore hinges on one's conception of democracy. Most extant indices focus only on the competitive or liberal properties of democracy and therefore have little to say about the majoritarian, consensual, deliberative, or egalitarian properties.

To some extent, problems of mismatch between concepts and measures can be mitigated by the application of Bayesian latent variable models (Armstrong 2011; Pemstein, Meserve & Melton 2010) or factor analysis (Bollen 1993; Coppedge, Alvarez & Maldonado 2008). These methods allow one to combine information from many extant indices or from multiple components of a single index. In this way, one can construct a summary latent-variable index that reflects all of the qualitative distinctions found in any of the measures included in the analysis (assuming they lie on the same empirical dimension). However, these secondary analyses cannot make qualitative distinctions that are found outside of the measures they analyze. They are thus constrained by the limitations of extant measures. Also, when a latent variable model imposes unidimensionality on many different measures, it is often unclear what concepts the latent variable actually reflects. In combination, while such measures make it possible to compare country-years in terms of “better or worse”, they typically make substantive interpretations in terms of actual conditions difficult if not impossible.

Sources

Sources employed to provide coding for extant indices are often problematic. For example, the Political Rights and Civil Liberties indices rely heavily on secondary accounts such as the *New York Times* and Keesing's Contemporary Archives for coding in the 1970s and 1980s. These historical sources, while informative, do not cover every country in the world with equal depth and with equivalent sensibilities, introducing potential bias into the resulting indices. In later eras, these indices have relied much more on country expert coding. However, the change from one source of evidence to another—coupled with some possible changes in coding procedures—may compromise the continuity of the time-series. No systematic effort has been made to revise previous scores so that they are consistent with current coding criteria and expanded knowledge

of past regimes.⁶

Some indices such as the EIU rely heavily on polling data, which is available on a non-comparable and highly irregular basis for 100 or so nation-states. For other countries (about half of the population covered by the EIU) these data must be estimated by country experts or imputed. Procedures employed for this estimation are not made publicly available.⁷ Although surveys of citizens are important for ascertaining attitudes, they are not available for every country in the world, and in no country are they available on an annual basis. Moreover, use of such surveys severely limits the historical reach of any democracy index, since the origin of systematic surveying stretches back only a half-century (in the US and parts of Europe) and is much more recent in most countries. Other survey-based questions are of questionable relevance for understanding the quality of democracy in a polity. It is of course interesting to know whether citizens regard their country as democratic, whether they support democratic institutions and practices, and whether they subscribe to democratic norms such as tolerance. However, responses to these questions should not without great caution be regarded as accurate reflections of how democratic a country is. For example, a recent Pew Research Center Global Attitudes Project survey asked citizens of 18 Latin American countries whether they preferred “a democratic form of government” over “a leader with a strong hand” to solve their country’s problems (Pew Research Center 2014). There was no familiar pattern in their responses: Nicaragua, Panama, and Bolivia came out on top and the conventionally most democratic countries Chile, Uruguay, and Costa Rica were in the middle. The correlation with Freedom House’s ratings was -0.002, and with our Electoral Democracy Index, -0.067.⁸

All indices rely (at least in part or indirectly) on subjective coding by country experts or trained coders. Such measurement practices have been criticized because reliability problems are likely to arise owing to random and systematic measurement errors introduced as raters interpret the sources differently. However, some properties of democracy are very difficult to capture with

⁶ Gerardo Munck, personal communication (2010).

⁷ Reliance on survey data also raises even more difficult questions about validity, i.e., whether the indicators measure what they are supposed to measure. There is surprisingly little empirical support for the notion that respondents are able to assess their own regimes in a cross-nationally comparable way or that they tend to live under regimes that are congruent with their own values.

⁸ The Pew data are the percentage in 2013-14 choosing “a democratic form of government” minus the percentage choosing “a leader with a strong hand.” Freedom House data are from <https://freedomhouse.org/sites/default/files/Individual%20Country%20Ratings%20and%20Status%2C%201973-2015%20%28FINAL%29.xls>, accessed March 13, 2015. The V-Dem data are the Electoral Democracy Index values for 2012 from version 4 (March 2015). The correlation between the Freedom House and V-Dem scores for these countries was 0.91.

objective indicators, justifying the use of subjective questions (Schedler 2012). Detailed coding criteria, careful training, and the use of multiple coders may be used to counter the potential problems of subjective assessments.

While the Freedom House Political Rights and Civil Liberties ratings were based originally on the work of a single person (Raymond Gastil) the scores are now based on a multilayered process of analysis and evaluation by Freedom House staff and consultant regional experts. After analysts (one for each country) have suggested numerical scores for components, these are reviewed in a series of regional meetings by the analysts, regional experts, and an in-house team. In a subsequent step, a general cross-regional evaluation is carried out to ensure comparability and consistency in the scores. No information is provided about the extent to which scores undergo changes during this process. A very similar process is used for the BTI.

Judgments by experts or trained coders can be made fairly reliably if there are clear and concrete coding criteria (Munck & Verkuilen 2002). Unfortunately, this is not always the case. For example, the Nations in Transit expert survey poses five sub-questions to answer the question, “Is the country’s governmental system democratic?”:

1. Does the Constitution or other national legislation enshrine the principles of democratic government?
2. Is the government open to meaningful citizen participation in political processes and decision-making in practice?
3. Is there an effective system of checks and balances among legislative, executive, and judicial authority?
4. Does a freedom of information act or similar legislation ensure access to government information by citizens and the media?
5. Is the economy free of government domination?⁹

Quite aside from the debatable validity of equating democracy with separation of powers and a free-market economy, these are not easy questions to answer, and their difficulty stems from the rather general or ambiguous terms in which they are posed. One cannot judge whether the “principles of democratic government” are “enshrined” without specifying what those principles are. Are these principles enshrined if they are written on parchment but not practiced? What does it mean to be “open to meaningful citizen participation”? What is the basis for determining whether checks and balances are “effective”? What degrees and kinds of government regulation

⁹ Report on Methodology, downloaded from www.freedomhouse.hu/images/nit2009/methodology.pdf.

and ownership can be permitted before economic “freedom” is infringed upon?

Wherever the items in a questionnaire do not define these criteria, respondents must rely on their own implicit beliefs and assumptions. This creates a danger that coding decisions about particular topics—e.g., press freedom—will reflect the coder’s overall sense of how democratic a country is rather than an independent evaluation of the level of press freedom. It is the ambiguity of the questions underlying these surveys that foster this sort of premature aggregation. In this respect, “disaggregated” indices may actually be considerably less disaggregated than they appear, as discussed below.

An additional issue with the use of expert coders is that their coding is often not fully independent in the final product. As mentioned, Freedom House and the BTI employ panels that adjust scores originally determined by country experts. While this helps to achieve cross-country and cross-coder comparability, it also means that the resulting scores follow a more centralized process than it may appear and do not necessarily reflect the judgments of individual country experts who fill out the standardized questionnaire. Country expert judgments are, in effect, advisory rather than determinative. One helpful and important aspect of Freedom House’s methodology is the provision of narrative country reports where reasons for the judgments are detailed.

Disaggregation

A few approaches to measurement in the field of democracy are highly disaggregated, measuring democracy at the ground level (i.e., with specific questions about specific features of this ambient concept). This would include democracy assessments (aka audits), which provide detailed indicators for a single country¹⁰ and datasets like National Elections across Democracy and Autocracy (“NELDA” [Hyde & Marinov 2012]). Other ventures to measure democracy in a disaggregated fashion have been proposed but not fully implemented (e.g., Beetham, Bracking, Kearton & Weir 2001).

Among the extant indices reviewed in Table 1, only the BTI, the Democracy Barometer, the Freedom House indices (since 2006), and the EIU index can be regarded as highly and systematically disaggregated. They report scores for several properties of (liberal) democracy – such as electoral process, the functioning of government, civil liberties, and the rule of law.

¹⁰ E.g., Beetham (1994, 2004), Diamond and Morlino (2005), Landman and Carvalho (2008), Proyecto Estado de la Nación (2001).

Unfortunately, EIU and Freedom House (regarding Freedom in the World) are unwilling to divulge their raw data at the indicator level, so it is difficult to judge the accuracy and independence of these measures.

Most democracy indices approach the subject at a fairly abstract level, even at the point of data collection. This introduces problems of coding, for there are inevitably different ways to interpret a generally posed question about the state of democracy in a country, as suggested in the previous section.

Consider the Polity index, which is disaggregated into six indicators: competitiveness of participation, regulation of participation, competitiveness of executive recruitment, openness of executive recruitment, regulation of executive recruitment, and constraints on executive. Although each of these components is described at length in the Polity codebook (Marshall, Gurr & Jaggers 2014), it is difficult to say precisely how they would be coded in particular instances. Even in disaggregated form (e.g., Gates, Hegre, Jones, & Strand 2006), the Polity index is fairly abstract, and therefore open to diverse interpretations.

The two principal Freedom House indices—Civil Liberties and Political Rights—are similarly difficult to interpret. The Political Rights index is a weighted sum of (a) Electoral Process, (b) Pluralism and Participation, and (c) Functioning of Government. The Civil Liberties index comprises (a) Freedom of Expression, (b) Association and Organizational Rights, (c) Rule of Law, and (d) Personal Autonomy and Individual Rights. This represents a step towards disaggregation, yet intercorrelations among the seven components are extremely high—Pearson's $r = 0.86$ or higher. This by itself is not necessarily problematic; it is possible that all democratic (or nondemocratic) things go together. However, the high inter-correlations coupled with their ambiguous coding procedures suggest that these components may not be independent of one another. It is impossible to discard the possibility that country coders have a general idea of how democratic each country is, and that this idea is reflected in consistent scores across the multiple indicators.¹¹

¹¹ Naturally, V-Dem is not free from this concern. However, the specificity of the questions in the V-Dem questionnaire should encourage coders to bracket their general understandings of democracy, writ large, and focus instead on the question at-hand.

Coverage

Many democracy indices are limited in country coverage, as noted in Table 1. *Nations in Transit* covers only the post-communist states. *Countries at the Crossroads* covers seventy countries (beginning in 2004) deemed to be strategically important and at a critical juncture in their trajectory. The Democracy Barometer covers 70 countries, some of which are selected on the basis of data availability. The core of the enterprise is focused on measuring differences among the thirty most democratic polities in the world – countries where Freedom House, Polity, and most other indices do not provide meaningful variation. No indices, with the exception of V-Dem, include colonies prior to independence.

Indices are also generally limited in their temporal coverage, especially those that offer the most disaggregated data. The EIU begins in 2006, the BTI starts in 2003, the Democracy Barometer commences in 1990, the Political Rights and Civil Liberties indices stretch back to 1972, the DD embarks in 1946, and the BNR in 1913. Only a few democracy indices stretch back further in historical time—notably Vanhanen’s indices of Competition and Participation (1810–), and Polity, BMR, and the Skaaning, Gerring & Bartusevičius (2015) Lexical index – all starting in 1800. We suspect that the value of these indices stems partly from their fairly comprehensive historical coverage—though Polity excludes states with fewer than 500,000 inhabitants.

Discrimination

Many of the leading democracy indices are insensitive to gradations in the degree or quality of democracy. If one purpose of any measurement instrument is discrimination—being able to distinguish greater and lesser degrees of democracy in a precise and reliable way (Jackman 2008) — extant democracy indices fall short of the ideal.

At the extreme, binary measures such as DD, BMR, and BNR reduce democracy to a dummy variable. This allows one to divide up the world of polities into crisp sets – democracies and non-democracies (variously articulated as autocracies, dictatorships, or authoritarian regimes) – an approach that resonates with ordinary language and is useful for many purposes. However, dichotomous coding of regimes reduces discrimination to the very minimum, ignoring differences of degree within the two types. For example, the DD index recognizes no distinctions within the large category of countries that have competitive elections and occasional leadership turnover. Papua New Guinea and Sweden thus receive the same score, despite evident

differences in the quality of elections, civil liberties, and barriers to competition, which are not part of their definition. Thus, although binary indices serve an important and indispensable function they cannot be used to capture fine differences of degree across regimes or through time.

Ordinal and interval indices are more sensitive to quantitative gradations of democracy/autocracy because they have more ranks or levels. Freedom House scores democracy on a seven-point index (13 points if the Political Rights and Civil Liberties indices are combined). Polity provides a total of 21 points if the component Democracy and Autocracy indices are merged, creating the Polity2 index. Appearances, however, can be deceiving. Polity scores, for example, bunch up at a few thresholds (64% percent of the observations are either at the extremes of at or below -6 or at or above +6), suggesting that the scale does not discriminate between levels of democracy as well as it seems.

The Democracy Barometer and Unified Democracy Scores, as well as indices produced by EIU, Vanhanen (2000), and Coppedge, Alvarez & Maldonado (2008), are the most smoothly continuous. However, even when countries receive different scores, their scores may not be *significantly* different because of measurement error. The magnitude of the measurement uncertainty is usually unknown or unreported, but secondary analyses of these measures suggest that it is quite large (Treier & Jackman 2008; Pemstein, Meserve & Melton 2010).¹² To some extent, latent-variable models can improve the quantitative discrimination of the summary index. Such techniques also make it possible to assign confidence intervals to each point estimate, although only a couple of these analyses report them (Pemstein, Meserve & Melton 2010; Treier & Jackman 2008).

In sum, the discriminatory power of even the most refined democracy indices is generally too low to justify confidence that a country that scores a few points higher than another is actually more democratic (Armstrong 2011; Pemstein, Meserve & Melton 2010). According to one of the most rigorous analyses, Polity scores are so imprecise that one cannot be confident that the United States in 2000 was any more democratic than the top 70 of 153 countries (Treier & Jackman 2008).

¹² Questions can also be raised about whether these indices are properly regarded as interval scales. This is a difficult problem, although Pemstein, Meserve, and Melton (2010) offer a solution that is incorporated into many V-Dem measures.

Aggregation

Since democracy is a multi-faceted concept all composite indices must wrestle with the aggregation problem— which indicators to combine into a single index, whether to add or multiply them, and how much to weigh them. Different solutions to the aggregation problem lead to quite different results (Munck & Verkuilen 2002; Munck 2009; Goertz 2006).

In order for any aggregation scheme to be successful, rules must be clear, they must be operational, and they must reflect an accepted definition of democracy. Otherwise, the resulting measure is not valid. Although most extant indices have fairly explicit aggregation rules, they are only rarely justified explicitly with reference to a particular definition of democracy, including the relationship among the attributes and between the attributes and the overarching concept.

The aggregation rules used by most democracy indices are additive, with an (implicit or explicit) weighting scheme; they are sums, averages, or weighted averages of various components. It is far from obvious that this is the most appropriate aggregation rule. Others recommend that one should consider the various components of democracy as necessary (non-substitutable), mutually constitutive (interactive), or both (Goertz 2006: 95–127; Munck 2009; Schneider 2010).

A more inductive approach may also be taken to the aggregation problem. Coppedge, Alvarez & Maldonado (2008) apply an exploratory principal components analysis of a large set of democracy indicators, identifying two dimensions that they label Contestation and Inclusiveness. Other writers analyze extant indices as reflections of a (unidimensional) latent variable. This is the approach taken by the UDS (Pemstein, Meserve & Melton 2010; see also Bollen & Jackman 1989; Bollen & Paxton 2000; Treier & Jackman 2008). However, problems of definition are implicit in any factor-analytic or latent-variable index, for an author must decide which indicators to include in the sample—requiring a judgment about which extant indices are measuring “democracy” and which are not—and how to interpret commonalities among the chosen indicators. This is not solvable simply by referring to the labels assigned to the indicators in question. Consider that many of the most well-known and widely regarded democracy indices are packaged as “rights,” “liberties,” or “freedom” rather than democracy per se, and do not necessarily measure exactly the concepts they purport to measure. Moreover, while factor-analytic and other latent variable approaches allow for the incorporation of multiple sources of data, thereby reducing some sources of error, they remain biased by any systematic error common to the chosen data sources.

Assessing Validity and Reliability

All empirical measures aim to achieve validity and reliability. Validity refers to whether the proposed index measures what it purports to measure (the concept of interest) in an unbiased fashion. Reliability refers to an estimate of how precise that index might be, i.e., whether replications of the measurement procedure would achieve the same result.¹³

We have already discussed potential problems of validity and reliability arising from choices in definition, sources, and aggregation. Evidently, there is cause for concern. In this section, we discuss methods of assessment.

One approach is to ask multiple experts to code each question in a survey. If the coding is conducted independently (a problem addressed above), we may regard the degree of inter-coder agreement as evidence of relative consensus, at least at the indicator level. If, however, the original coding is not preserved, or not made public, it is not possible to use this information to judge validity or reliability. Several projects, such as Freedom House and BTI, consult more than one expert and describe a process for reconciling disagreements. However, it is extremely rare for a project to fully report the extent of the original disagreements, inter-coder reliability, or confidence bounds around the reconciled estimates. Inter-coder reliability tests are not common practice among democracy indices, as noted in Table 1. Freedom House does not conduct such tests in any formal sense. Polity has done so, but only for a single year (1999), and it required a good deal of hands-on training before coders reached an acceptable level of coding accuracy. This suggests that other coders might not reach the same decisions simply by reading Polity's coding manual. And this, in turn, can contribute to the problem of conceptual validity, in which key concepts are not well matched to the empirical data.

Another approach is to reexamine scores produced by key indices after the process is complete. This *ex post* analysis usually focuses on specific countries well known to the specialists conducting the review. A recent study by scholars of Central America alleges major flaws in coding for Costa Rica, El Salvador, Honduras, Guatemala, and Nicaragua in Polity and the Vanhanen indices — errors that, the authors suspect, also characterize other indices and other countries (Bowman, Lehoucq & Mahoney 2005). Of course, it is possible that regional specialists would also disagree with each other, which returns us to the need for transparency and estimates of uncertainty.

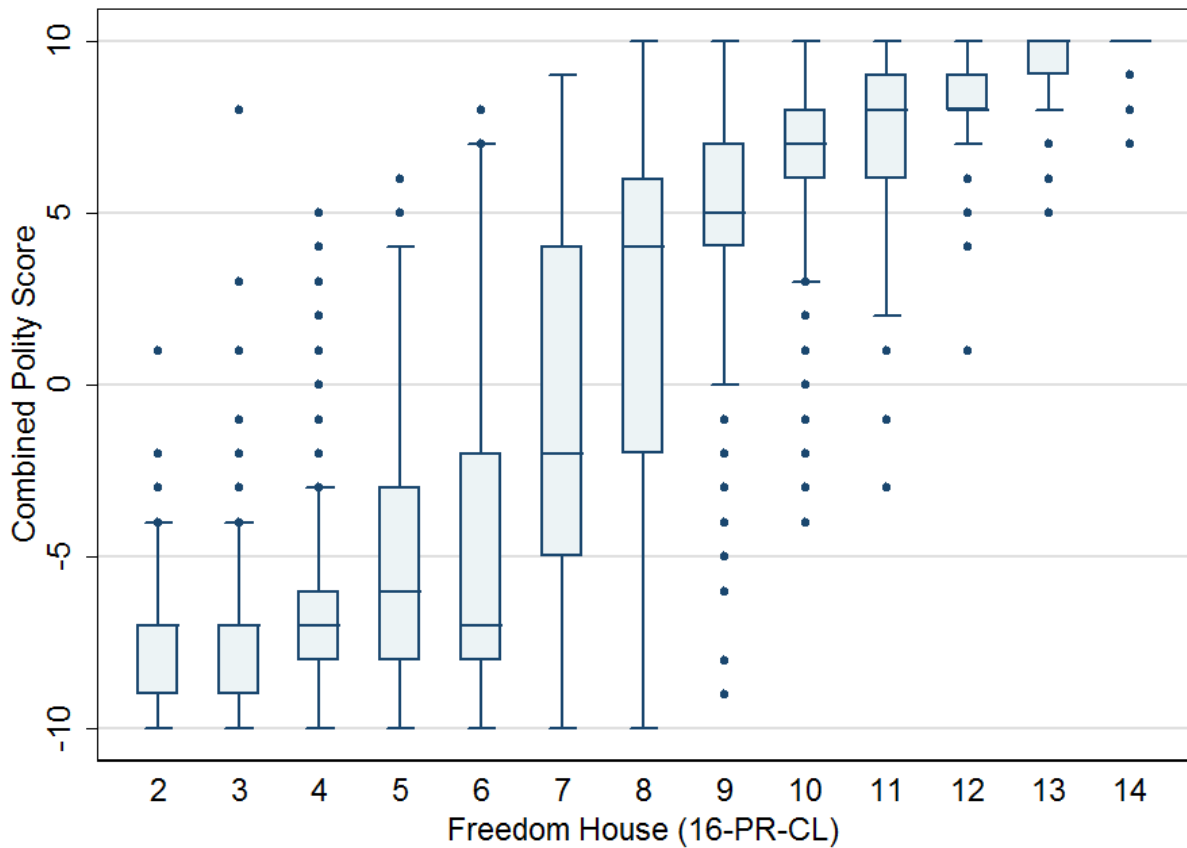
¹³ In practice, these concepts are often enmeshed. For example, convergent validity tests attempt to measure validity by examining reliability. That is why we deal with validity and reliability together, in the same section.

Freedom House – alone among the projects surveyed in Table 1 – issues narrative country reports each year that accompany its annual coding of countries. This helps to bridge the gap between assigned scores and complex realities on the ground, explaining why a country may have achieved a higher or lower score relative to the previous year. Of course, this does not resolve potential disagreements over those scores.

A third approach is to examine correlations across democracy indices for evidence of agreement. Encouragingly, the correlation between Polity2 and Political Rights – the dominant indices, by most accounts – is a respectable 0.88 (Pearson’s r). Yet, closer examination reveals that the consensus is largely the product of countries lying at the democratic extreme—Canada, Sweden, the United States, *et al.* When countries with the top two scores on the Freedom House Political Rights scale are eliminated, Pearson’s r drops to 0.63. Figure 1 confirms that although Polity and Freedom House data are highly correlated, the agreement is mostly at the extremes. Intermediate values often diverge, as shown by the much higher confidence intervals. This is problematic, especially when one considers that scholars and policymakers are usually interested in precisely those countries lying in the middle of the distribution—countries that are neither highly autocratic nor highly democratic.¹⁴

¹⁴ For extensive cross-country tests see Hadenius and Teorell (2005a).

Figure 1: Intercorrelations between Polity and Freedom House



Boxplot of inter-correlations between Polity2 and Freedom House (Political Rights and Civil Liberties combined, transformed so that larger numbers indicate higher scores). Boxes indicate the inter-quartile range (IQR: 25th to 75th percentile), so each box contains half of the observations. The line within the box is the median. Vertical lines sticking out of the top and bottom of each box ("adjacent lines") are the most extreme values within 1.5 interquartile range of the nearer quartile. Dots represent values beyond the adjacent lines, i.e., outliers.

Not surprisingly, these measurement differences translate into different results when democracy is employed in analyses. Przeworski & Limongi (1997), for example, find that per capita income was not associated with transitions to democracy, but Zachary Elkins (2000) shows that their result depended on using a binary measure of democracy; when graded indices such as Freedom House or Polity were substituted, the correlation between income and transitions returned to significance. More generally, Casper & Tufis (2003) show that few explanatory variables (beyond per capita income) have a consistently significant correlation with levels of democracy when different democracy indices are employed. In fact, the *only* predictor that both remained significant regardless of how democracy was measured and survived all their

robustness checks was Rae's index of party-system fractionalization (an aspect of party competition, which is arguably a component of democracy itself).¹⁵ Thus, we have good reasons to suspect that extant indices suffer problems of validity and that these problems are consequential. They impact what we think is going on in the world.

II. Varieties of Democracy

In this section we outline the distinctive features of the V-Dem project. While most other projects are focused on developing one or two very high-level indices, V-Dem is focused on the construction of a wide-ranging database consisting of a series of measures of varying ideas of what democracy is or ought to be, a wide variety of meso-level indices of different components of such ideals of democracy, and about 350 specific indicators (as laid out in our Methodology document). As such, its goal is orthogonal to Polity, Freedom House, et al. We hope that V-Dem indicators and indices will complement, not replace, the more highly aggregated indices produced by other groups.

In addition to disaggregation, several features of the V-Dem project deserve emphasis:

- Historical data extending back to 1900 (eventually to 1789)
- Multiple, independent coders for each (non-factual) question
- Inter-coder reliability tests, incorporated into a Bayesian measurement model.
- Confidence bounds for all point estimates associated with non-factual questions
- Multiple indices reflecting varying theories of democracy
- Transparent aggregation procedures
- All data freely available, including original coder-level judgments (exclusive of any personal identifying information)

In the following sections we discuss each of these features, beginning with the multidimensional qualities of democracy, the aggregation procedures used for these indices, the benefits of disaggregation, and finally, a residual category of “additional payoffs.”

¹⁵ See also Hadenius and Teorell (2005a) and Högström (2013).

Principles

Many problems of conceptualization and measurement stem from the decision to represent democracy as a single point score (based on a binary, ordinal, or interval index). Summary measures have their uses. Sometimes one wants to know whether a country is democratic or non-democratic or how democratic it is overall (Bayer & Bernhard 2010). Even so, the goal of summarizing a country's regime type is elusive, as the proliferation of democracy indices suggests. The highly abstract and contested nature of democracy impedes an authoritative measurement of the concept that is viable across countries and time-periods.

Naturally, one can always impose a particular definition, insist that *this* is democracy, and then go forward with the task of measurement. But this is unlikely to convince anyone not already predisposed to the author's point of view. Moreover, even if one could gain agreement over the definition and measurement of democracy, a great deal of useful information about the world would be lost while aggregating up to such a general concept. V-Dem helps with this task insofar as it offers the base-level indicators that any index requires. This mitigates the need for new indices to start from scratch, collecting data for all countries and all time-periods.

While there is no consensus on what democracy at-large means (beyond the diffuse notion of "rule by the people"), one may discern from the voluminous literature seven traditions with rather distinct sets of core values: electoral, liberal, majoritarian, consensual, participatory, deliberative, and egalitarian. We refer to these principles, summarized in Table 2, as "Varieties of Democracy." We hope that these seven principles, taken together, offer a fairly comprehensive accounting of the concept of democracy.

- The *electoral* principle of democracy embodies the core value of making rulers responsive to citizens through periodic elections. In the V-Dem conceptual scheme, electoral democracy is captured by Robert Dahl's (1971, 1989) conceptualization of "polyarchy." It is the idea that democracy is achieved through competition among leadership groups, which vie for the electorate's approval during periodic elections before a broad electorate. Parties and elections are the critical instruments in this largely procedural account of the democratic process. Following Dahl, we also count the existence of freedom of association that goes beyond political parties, a free media and freedom of expression ensuring possibilities for enlightened understanding in selecting leaders, and alternative sources of information on the (in)actions of political elites. Although many additional factors might be regarded as important for ensuring and enhancing electoral contestation (e.g., additional civil liberties and an independent judiciary), these factors are

often viewed as secondary to electoral institutions (Dahl 1956; Przeworski et al. 2000; Schumpeter 1950) and in the V-Dem scheme are classified as aspects of other principles of democracy.

In the V-Dem conceptual scheme, the electoral principle is important for all other conceptions of democracy. We also consider it fundamental: we would not want to call a regime without elections “democratic” in any sense. At the same time, countries can have “democratic qualities” without being complete polyarchies. We see electoral dimension as a continuum.

We also recognize that the electoral principle in itself does not capture various understandings of democracy that emphasize non-electoral properties and that critique electoral democracy as being insufficient. These critiques have given rise to additional principles, each of which is designed to correct one or more limitations of electoral democracy.

- The *liberal* principle of democracy stresses the intrinsic value of protecting individual rights against potential “tyranny of the majority” and state repression. This is achieved through constitutionally protected civil liberties, strong rule of law, and effective checks and balances that limit the use of executive power. These are seen as defining features of the liberal aspect of democracy, not simply as aids to political competition. The liberal model takes a negative view of political power insofar as it judges the quality of democracy by the limits placed on government.¹⁶
- The *majoritarian* principle of democracy reflects the idea that the will of the majority should be sovereign. Accordingly, democracy is improved if it ensures that the many prevail over the few in terms of making decisions and act on policy issues thus boosting what is often referred to as governing capacity. This also reflects the ideal that one party should clearly be accountable to the electorate in order to make responsiveness possible. To facilitate this, political institutions should centralize and concentrate, rather than disperse, power (within the context of competitive elections). This may be facilitated by a unitary constitution, unicameralism, plurality electoral laws (or majoritarian two-round

¹⁶ See Dahl (1956) on “Madisonian Democracy”; see also Gordon (1999), Hamilton, Madison & Jay (1992), Hayek (1960), Held (2006, ch. 3), Hirst (1989), Mill (1958), Vile (1998).

systems) leading to two-party systems, governing party domination of legislative committees, no constitutional provisions for supermajorities, no or weak judicial review, and so forth—in other words, few veto players.¹⁷

- The *consensual* principle of democracy is in several ways the opposite to the majoritarian vision and emphasizes that the political institutions should encourage, in the extreme, mandate the inclusion of as many political perspectives as possible. Accordingly, democracy is improved in the consensual sense if it makes it easier for small groups to be represented in the political system and make their voices heard, and that require the national head of government to share power with other political actors and bodies. This also reflects the ideal that responsiveness is accomplished when each interest can have its own party represented. Consensual democracy therefore emphasizes proportional electoral laws making large party systems possible, having two (or more) legislative chambers, forming oversized multiparty cabinets, separating national and subnational political units (federalism), constitutional provisions of supermajorities, strong judicial review, among other attributes.¹⁸
- The *participatory* principle of democracy embodies the values of direct rule and active participation by citizens in all political processes. It is usually viewed as a lineal descendant of the “direct” (i.e., non-representative) model of democracy. The motivation for participatory democracy is uneasiness about a bedrock practice of electoral democracy: delegating authority to representatives. Direct rule and involvement by citizens is preferred, wherever practicable. And within the context of representative government, the participatory element is regarded as the most democratic element of the polity. This model of democracy thus takes suffrage for granted, emphasizing turnout (actual voting) as well as non-electoral forms of participation such as citizen assemblies, party primaries, referenda, juries, social movements, public hearings, town hall meetings, and other forums of citizen engagement.¹⁹
- The *deliberative* principle of democracy enshrines the core value that political decisions in

¹⁷ See Bagehot (1963), Ford (1967), Goodnow (1900), Lijphart (1999), Lowell (1889), Ranney (1962), Schattschneider (1942), Tsebelis (2002), Wilson (1956).

¹⁸ Our definition of consensus democracy is almost identical to Lijphart’s formal definition of consensus democracy (Lijphart 1999, 3-4). However, Lijphart’s book, and his prior work on consociationalism, imply that his version of consensus democracy is designed to foster the inclusion of different religious, linguistic, or ethnic communities. We prefer to include these attributes in our egalitarian principle. See also Mansbridge (1983) and Powell (2000).

¹⁹ See Barber (1988), Benelo & Roussopoulos (1971), Dewey (1916), Fung & Wright (2003), Macpherson (1977), Mansbridge (1983), Pateman (1976), Rousseau (1984), Young (2000).

pursuit of the public good should be informed by respectful and reason-based dialogue at all levels rather than by emotional appeals, solidary attachments, parochial interests, or coercion. According to this principle, democracy requires more than an aggregation of existing preferences. It therefore focuses on the process by which decisions are reached in a polity. A deliberative process is one in which public reasoning focused on the common good motivates political decisions—as contrasted with emotional appeals, solidary attachments, parochial interests, or coercion. There should also be respectful dialogue at all levels—from preference formation to final decision—among informed and competent participants who are open to persuasion (Dryzek 2010: 1). “The key objective,” writes David Held (2006: 237), “is the transformation of private preferences via a process of deliberation into positions that can withstand public scrutiny and test.” Some political institutions serve a specifically deliberative function, such as consultative bodies (hearings, panels, assemblies, courts); polities with these sorts of institutions might be judged more deliberative than those without them. However, the more important issue is the degree of deliberativeness that can be discerned across all powerful institutions in a polity (not just those explicitly designed to serve a deliberative function) and among the citizenry.²⁰

- The *egalitarian* principle of democracy holds that material and immaterial inequalities inhibit the actual use formal political (electoral) rights and liberties. It therefore addresses the goal of political equality across social groups – as defined by income, wealth, education, ethnicity, religion, caste, race, language, region, gender, sexual identity, or other ascriptive characteristics. Ideally, all groups should enjoy equal *de jure* and *de facto* capabilities to participate; to serve in positions of political power; to put issues on the agenda; and to influence policymaking. (This does not entail equality of power between leaders and citizens, as leaders in all polities are by definition more powerful.) Following the literature in this tradition, gross inequalities of health, education, or income are understood to inhibit the exercise of political power and the *de facto* enjoyment of political rights. Hence, a more equal distribution of these resources across social groups may be needed in order to achieve political equality.²¹

²⁰ See Bohman (1998), Elster (1998), Fishkin (1991), Gutmann & Thompson (1996), Habermas (1984, 1996), Held (2006, ch. 9), Offe (1997). A number of recent studies have attempted to grapple with this concept empirically; see Bächtiger (2005), Dryzek (2009, 2010), Mutz (2008), Ryfe (2005), Steiner et al. (2004), Thompson (2008).

²¹ See Ake (1999), Bernstein (1961, 1996), Dahl (1982, 1989), Dewey (1916, 1930), Dworkin (1987, 2000), Gould (1988), Lindblom (1977), Meyer (2007), Offe (1983), Sen (1999), Walzer & Miller (1995). Many of the writings

Thus, while most indices focus only on democracy's electoral and liberal attributes, V-Dem seeks to measure a broader range of attributes associated with the concept of democracy, as summarized in Table 2. It is important to recognize that the core values enshrined in the varying principles of democracy sometimes conflict with one another. Such contradictions are implicit in democracy's multidimensional character. Having separate indices that represent these different facets of democracy will make it possible for policymakers and academics to examine potential tradeoffs empirically. At present, we provide all indices except for majoritarian and consensual democracy. We plan to produce these last two indices in the near future.

The formulas we use to construct high-level indices (HLIs) such as the Electoral Democracy Index reflect both a family resemblance logic and the classical or Sartorian logic of necessary conditions (Collier and Mahon 1993; Goertz 2006; Sartori 1970). In practice this means that while an improvement in the score of every element increases the overall score of the democracy index in question, its marginal contribution is conditional on the score of all other elements.²² Collectively, these thick versions of the five concepts capture significant “varieties of democracy,” and thus offer a fairly comprehensive account of the concept of democracy.

cited previously under participatory democracy might also be cited here. Taking a somewhat different stand on this issue, Beetham (1999) and Saward (1998: 94-101) do not request an equal distribution of resources. Rather, they consider access to basic necessities in the form of health care, education, and social security to be democratic rights as they make participation in the political process possible and meaningful.

²² For a detailed description and justification of the aggregation procedures, see Coppedge et al. (forthcoming) and the *V-Dem Methodology* document.

Table 2:
Properties of Democracy

I. Electoral

Core Values: Contestation, competition.

Question: Are important government offices filled by free and fair multiparty elections before a broad electorate?

Institutions: Elections, political parties, competitiveness, suffrage, turnover.

II. Liberal

Core Values: Individual liberty, Protection against tyranny of majority and state repression.

Question: Is power constrained and are individual rights guaranteed?

Institutions: Civil liberties, independent bodies (media, interest groups); separation of powers, constitutional constraints on the executive, strong judiciary with political role.

III. Majoritarian

Core Values: Majority rule, Governing capacity, Accountability.

Question: Does the majority rule via one party, and can implement its policies?

Institutions: Consolidated and centralized, with special focus on the role of political parties, single-member districts, first-past-the post electoral rules.

IV. Consensual

Core Values: Voice and representation of all groups, possibly sharing power.

Question: How numerous, independent, and diverse are the groups and institutions that participate in policymaking?

Institutions: Federalism, PR, supermajorities, oversized cabinets, multiple parties.

V. Participatory

Core Values: Direct, active participation in decision-making by the people.

Question: Do citizens participate in political decision-making?

Institutions: voting, civil society, strong local government, direct democracy instruments.

VI. Deliberative

Core Values: Reasoned debate and rational arguments.

Question: Are political decisions the product of public deliberation based on reasoned and rational justification?

Institutions: Media, hearings, panels, other deliberative and consultative bodies.

VII. Egalitarian

Core Values: Equal political empowerment.

Question: Are all citizens equally empowered to use their political rights?

Institutions: Formal and informal practices that safeguard or promote equal distribution of resources and equal treatment.

Disaggregation

Within each principle, it is helpful to disaggregate further. At lower levels of abstraction these concepts become more tractable. While the world may never agree on whether the overall level of democracy in India can be summarized as a “4” or “5” (on some scale), we may agree on scores for this and other countries on more specific properties such as judicial independence, freedom of association, or legislative constraints on the executive.

Another advantage is the degree of discrimination that a disaggregated set of indicators offers relative to extant composite indices. While holistic measures of democracy float hazily over the surface, the indicators and indices of a disaggregated dataset are comparatively specific and precise. Contrasts and comparisons become correspondingly acute.

The data that we are collecting as part of the V-Dem project is disaggregated down to the level of country-date-variable-expert observations. (Of course, most end-users will be interested in indicators, not individual five expert ratings of an indicator. Even so, the latter may be useful for those who wish to construct their own measurement model.) As such, V-Dem offers one of the largest country-level databases in existence. There are approximately 350 indicators, of which about 175 (C-type indicators) are coded by 5 experts per item. Each of the indicators will eventually cover 206 countries, most of which will be coded over the course of a century (and eventually, two centuries). At present, we have data for 173 countries. This results in a dataset of more than 20 million data points, which will grow as it is updated (annually or semi-annually). For anyone interested in issues relating to democracy, this disaggregated format will be a vast resource for exploration and testing. It will allow policymakers and researchers to clarify how, specifically, one country’s democratic features differ from others in the same region, or across regions, and how a country’s features have changed over time.

Additional Payoffs

Beyond the high level of disaggregation, several additional features distinguish V-Dem from other measurement projects. First, we make all data (aggregated and disaggregated) freely available (once data collection and cleaning is complete). Second, we employ multiple coders for each (non-factual) question, respecting coder independence and allowing for inter-coder reliability tests. Third, we extend indicators of democracy back through modern history to 1900 (and eventually, in another wave of data collection currently in process, to 1789). Finally, we offer not only point scores measuring various properties of democracy, but also confidence

bounds for all point estimates associated with non-factual questions.

While other projects focused on the measurement of democracy may contain one or several of these features, none combines them all. These features of V-Dem are described in detail in the *Methodology* document. Here, we touch upon some of the anticipated payoffs from this approach to conceptualization and measurement. (For thoughts on the possible role of V-Dem indicators in the impact evaluation of democracy assistance programs see Appendix A.)

A detailed knowledge of cases is especially helpful in the context of country assessments. How can policymakers determine which aspects of a polity are most in need of assistance? It seems clear that for assessing the potential impact of programs focused on different elements of a polity it is helpful to have measures at hand that offer a differentiated view of the subject. Intuitively, the greatest effectiveness is achieved when program interventions are targeted on the weakest element of democracy in a country. A large set of fully differentiated indicators should make it possible to both identify those elements and test the assumption behind such choices.

Relatedly, the V-Dem dataset will allow policymakers and researchers to track a single country's progress or regress through time. One will be able to specify which facets of a polity have improved, and which have remained stagnant or declined. This means that the longstanding question of regime transitions will be amenable to empirical tests. When a country transitions from autocracy to democracy (or vice versa), which elements come first? Are there common patterns, a finite set of sequences, prerequisites? Or, is every transition unique? Do transition patterns effect the consolidation of democracy? With a large set of indicators measured over many years, it would become possible for the first time to explore transition sequences.²³ Does a newly vibrant civil society lead to more competitive elections, or to an authoritarian backlash? Do accountable elected officials create an independent judiciary, or does an independent judiciary make officials accountable? Similar questions could be asked about the relationships among citizenship, voting, parties, civil society, and other components of democracy, perhaps with the assistance of sequence-based econometrics (Abbott 1995; Abbott & Tsay 2000; Wu 2000).

Note that insofar as one wishes to judge trends, trend lines are necessary. A single snapshot of the contemporary world reveals nothing about the direction or speed at which countries are moving toward, or away from, democracy. Even trends in a short span of recent years can be very misleading, as many democratization paths contain many years of stasis

²³ Sequencing is explored by Schneider and Schmitter (2004b) with a smaller set of indicators and a shorter stretch of time. See also McFaul (2005) and Møller and Skaaning (2010).

punctuated by sudden movements toward or away from democracy. Assessments of global trends require even more data, as some countries move in opposite directions in any given year; “waves” of democratization exist only on average, with many exceptions (Huntington 1991).

The V-Dem database also enables one to test democracy’s causal effect as an independent variable (Carbone 2009; Gerring, Thacker, Alfaro 2012; Ross 2006; for an example, see Gerring et al. 2015). Does democracy hinder economic growth, contain inflation, promote public order, or ensure international peace? Answering such questions requires a lengthy time-series because the effects of these factors play out over many years. They also require a great deal of disaggregation because we need to know, as specifically as possible, which elements of democracy are related to which results. This is helpful from the policy perspective as well as from the analytic perspective, so that we can gain insight into causal mechanisms. Whether democracy is looked upon as an independent (causal) variable, or a dependent (outcome) variable, we need to know which aspect of this complex construct is at play.

Another advantage of the V-Dem approach derives from its data-collection strategy, which is open to public scrutiny and commentary. Although coders are anonymous, procedures of collection and aggregation are replicable. Experts with deep knowledge of each country’s history are enlisted. Several experts are asked to code each indicator wherever the latter requires a degree of judgment. Extensive tests of reliability and validity are provided, as described in the Methodology document. Periodic revisions are planned rather than ad hoc, and where re-coding is required it is applied to the whole dataset so that consistency is preserved through time and across polities.

With democracy indicators, as with most things, the devil is in the details. Getting these details right, and providing full transparency at each stage, should enhance the precision and validity of the measurement instrument, as well as its legitimacy in the eyes of potential end-users.

The importance of creating consensus on these matters can hardly be over-emphasized. The purpose of a set of democracy indicators is not simply to guide rich-world policymaking bodies such as the European Commission, USAID, the World Bank, and the UNDP. As soon as a set of indicators becomes established and begins to influence international policymakers, it also becomes fodder for dispute in countries around the world. A useful set of indicators is one that claims the widest legitimacy. A poor set of indicators is one that is perceived as a tool of western influence or a mask for the forces of globalization. The hope is that by disaggregating the elements of democracy down to levels that are more coherent and operational it may be possible to generate a broader consensus around this vexed subject. Likewise, when V-Dem indicators are

used to construct aggregate indices at least everyone will know what lies behind a country's score in a particular year since the indicators used to compose an index, along with the underlying ratings by country experts, are available for inspection.

III. A Plurality of Approaches

By way of concluding, let us return to the main theme of this paper. Democracy is a complex topic and there will always be different ways to conceptualize and measure it, each of which may be appropriate for a particular goal.

As an example, let us consider the matter of *scales*. In many settings, it is vital to distinguish between democratic and non-democratic polities; for this purpose, a binary scale is required. In some settings, it may be important to distinguish a number of categorical differences that are ordered with respect to each other; for this purpose, an ordinal scale is appropriate. In many settings, it may be more useful to have an overall measure of democracy that incorporates many facets of this multidimensional concept and is sensitive to fine differences of degree; for this purpose, an aggregate interval measure is required.

The type of scale provided by an index is an obvious instance of how diverse end-uses drive the construction of diverse indices. Naturally, there are many other factors a researcher must take into account when constructing a scale, and this is why we possess so many democracy indices (of which Table 1 is only a partial list).

As a second example, let us consider the approach to *ground-level measurement*. Relying on multiple independent coders, V-Dem is able to enlist the expertise of thousands of country experts, each of whom is intimately acquainted with the details of at least one country and at least one area related to the ambient concept of democracy. This also allows for a consistent and transparent scoring procedure in which codings are aggregated across coders with a formal model in order to produce a point estimate and an estimate of uncertainty (a confidence interval) that can be easily replicated.

By contrast, other projects either rely on one or several in-house experts to code the whole world (e.g., Polity and DD) or apply an informal system of deliberation across a team of country experts and in-house experts (e.g., Freedom House and BTI). These informal systems are, by their very nature, opaque and hence difficult to replicate. And they do not allow for an estimate of the error associated with each coding – error that we must assume affects all judgments pertaining to unobservable concepts such as democracy.

That said, the traditional approach to measurement also has an important benefit that we should not lose sight of – assuring a high degree of cross-country comparability. The meaning of a “4” should be the same for India as for Afghanistan if they are both coded by the same person or if the coders are encouraged to communicate with each other, reconciling their differences in the final product or submitting to a higher authority, who reconciles those differences. The V-Dem project, by contrast, must rely on a complex measurement model, assisted by “lateral” coding (coding of neighboring countries), to adjust each coder’s thresholds so that they have a uniform meaning. This is a fraught process, needless to say – though it will be enhanced by the addition of “vignette” codings in the next wave of data collection.

Many other tradeoffs might be identified. But the point is by now clear. We can and should encourage a diversity of approaches to measuring democracy. This is an explicit goal of the current project, as signaled by its title. Just as we imagine varieties of democracy we may acknowledge varieties of democracy projects, of which V-Dem is – we hope – an honorable addition.

References

- Abbott, Andrew. 1995. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21: 93–113.
- Abbott, Andrew, Angela Tsay. 2000. "Sequence Analysis and Optimal Matching Methods in Sociology." *Sociological Methods and Research* 29(1): 3–33.
- Altman, David, Aníbal Pérez-Liñán. 2002. "Assessing the Quality of Democracy: Freedom, Competitiveness and Participation in Eighteen Latin American Countries." *Democratization* 9(2): 85–100.
- Alvarez, Michael, Jose A. Cheibub, Fernando Limongi, Adam Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31(2): 3–36.
- Armstrong II, David A. 2011. "Stability and Change in the Freedom House Political Rights and Civil Liberties Measures." *Journal of Peace Research* 48(5): 653–662.
- Arndt, Christiane, Charles Oman. 2006. *Uses and Abuses of Governance Indicators*. Paris: OECD Development Centre.
- Bächtiger, André, ed. 2005. Special issue: "Empirical Approaches to Deliberative Democracy." *Acta Politica* 40(1–2).
- Bagehot, Walter. 1963 [1867]. *The English Constitution*. Ithaca: Cornell University Press
- Barber, Benjamin. 1988. *The Conquest of Politics: Liberal Philosophy in Democratic Times*. Princeton: Princeton University Press.
- Bayer, Resat, Michael Bernhard. 2010. "A Test of the Relative Utility of Scalar and Dichotomous Measures: The Operationalization of Democracy and the Strength of the Democratic Peace." *Conflict Management and Peace Science* 27(1): 85–101.
- Beetham, David, ed. 1994. *Defining and Measuring Democracy*. London: Sage.
- Beetham, David. 1999. *Democracy and Human Rights*. Cambridge: Polity.
- Beetham, David, Sarah Bracking, Iain Kearton, Stuart Weir, eds. 2001. *International IDEA Handbook on Democracy Assessment*. The Hague: Kluge Academic Publishers.
- Beetham, David. 2004. "Towards a Universal Framework for Democracy Assessment." *Democratization* 11(2):1–17.
- Benelo, George C., and Dimitrios Roussopoulos, eds. 1971. *The Case for Participatory Democracy: Some Prospects for a Radical Society*. New York: Viking Press.
- Berg-Schlosser, Dirk. 2004a. "Indicators of Democracy and Good Governance as Measures of the Quality of Democracy in Africa: A Critical Appraisal." *Acta Politica* 39(3): 248–278.
- Berg-Schlosser, Dirk. 2004b. "The Quality of Democracies in Europe as Measured by Current Indicators of Democratization and Good Governance." *Journal of Communist Studies and Transition Politics* 20(1): 28–55.
- Bernhard, Michael, Timothy Nordstrom, Christopher Reenock. 2001. "Economic Performance, Institutional Intermediation, and Democratic Survival." *Journal of Politics* 63(3): 775–803.
- Bernstein, Eduard. 1961. *Evolutionary Socialism: A Criticism and Affirmation*, Random House.
- Bernstein, Eduard. 1996. *Selected Writings of Eduard Bernstein, 1900-1921*. Amherst: Prometheus Books.
- Bertelsmann Stiftung. [various years]. *Bertelsmann Transformation Index [2003–]: Towards Democracy and a Market Economy*. Washington, DC: Brookings Institution Press.
- Besancon, Marie. 2003. *Good Governance Rankings: The Art of Measurement*. World Peace Foundation Reports 36. Cambridge, MA: World Peace Foundation.

- Bohman, James. 1998. "Survey Article: The Coming of Age of Deliberative Democracy." *Journal of Political Philosophy* 6(4): 400–25.
- Boix, Carles, Michael Miller, Sebastian Rosato. 2013. "A Complete Dataset of Political Regimes, 1800-2007." *Comparative Political Studies* 46(12), 1523-1554.
- Bollen, Kenneth A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45(3): 370–390.
- Bollen, Kenneth A. 1993. "Liberal Democracy: Validity and Method Factors in Cross-National Measures." *American Journal of Political Science* 37(4): 1207–1230.
- Bollen, Kenneth A. 2001. *Cross-National Indicators of Liberal Democracy, 1950–1990*. 2nd ICPSR version. Chapel Hill, NC: University of North Carolina and Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Bollen, Kenneth A., Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(1): 58–86.
- Bollen, Kenneth A., Robert W. Jackman. 1989. "Democracy, Stability, and Dichotomies." *American Sociological Review* 54(4): 612–621.
- Bollen, Kenneth, Pamela Paxton, Rumi Morishima. 2005. "Assessing International Evaluations: An Example from USAID's Democracy and Governance Program." *American Journal of Evaluation* 26(2): 189-203.
- Bowman, Kirk, Fabrice Lehoucq, James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38(8): 939–970.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Bernhard Weßels. 2012. "The Democracy Barometer: A New Instrument to Measure the Quality of Democracy and Its Potential for Comparative Research." *European Political Science* 11(4): 519-536.
- Burnell, Peter. 2008. "From Evaluating Democracy Assistance to Appraising Democracy Promotion." *Political Studies* 56(2): 414-434.
- Carbone, Giovanni. 2009. "The Consequences of Democratization." *Journal of Democracy* 20(2): 123-137.
- Carothers, Thomas. 1999. *Aiding Democracy Abroad: The Learning Curve*. Washington, DC: Carnegie Endowment.
- Carothers, Thomas. 2006. "The Backlash against Democracy Promotion." *Foreign Affairs* 85 (Mar/Apr).
- Casper, Gretchen, Claudiu Tufis. 2003. "Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets." *Political Analysis* 11(2): 196–203.
- Cheibub, Jose Antonio, Jennifer Gandhi, James Raymond Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143(1–2): 67–101.
- Chua, A.L. 2003. *The World on Fire: How Exporting Free Market Democracy Breeds Ethnic Hatred and Global Instability*. New York: Anchor Books.
- Collier, David, Steven Levitsky. 1997. "Democracy with Adjectives: Conceptual Innovation in Comparative Research." *World Politics* 49(3): 430–451.
- Collier, David and James Mahon (1993). "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87(4): 845-855.
- Coppedge, Michael, John Gerring with David Altman, Michael Bernhard, Steven Fish, Allen Hicken, Matthew Kroenig, Staffan I. Lindberg, Kelly McMann, Pamela Paxton, Holli A. Semetko, Svend-Erik Skaaning, Jeffrey Staton, Jan Teorell. 2011. "Conceptualizing and Measuring Democracy: A New Approach." *Perspectives on Politics* 9(1): 247–267.
- Coppedge, Michael, Wolfgang H. Reinicke. 1990. "Measuring Polyarchy." *Studies in Comparative*

- International Development* 25(1): 51–72.
- Coppedge, Michael, Angel Alvarez, Claudia Maldonado. 2008. “Two Persistent Dimensions of Democracy: Contestation and Inclusiveness.” *Journal of Politics* 70(3): 335–350.
- Coppedge, Michael, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell. Forthcoming. “Measuring High Level Democratic Principles using V-Dem Data.” *International Political Science Review*.
- Cox, M., John Ikenberry, T. Inoguchi (eds). 2000. *American Democracy Promotion: Impulses, Strategies and Impacts*. Oxford: Oxford University Press.
- Crawford, Gordon. 2001. *Foreign Aid and Political Reform: A Comparative Analysis of Democracy Assistance and Political Conditionality*. Palgrave Macmillan.
- Cunningham, Frank. 2002. *Theories of Democracy*. London: Routledge.
- Dahl, Robert A. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.
- Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Dahl, Robert A. 1982. *Dilemmas of Pluralist Democracy*. New Haven: Yale University Press.
- Dahl, Robert A. 1989. *Democracy and its Critics*. New Haven: Yale University Press.
- Dewey, John. 1916. *Democracy and Education*. New York: The Free Press.
- Dewey, John. 1930. *Individualism*. New York: Minton, Balch & Company.
- Diamond, Larry, Leonardo Morlino, eds. 2005. *Assessing the Quality of Democracy*. Baltimore: Johns Hopkins University Press.
- Dryzek, John S. 2009. “Democratization as Deliberative Capacity Building.” *Comparative Political Studies* 42(11): 1379–1402.
- Dryzek, John S. 2010. *Foundations and Frontiers of Deliberative Governance*. Oxford: Oxford University Press.
- Dunn, John. 2005. *Setting the People Free: The Story of Democracy*. London: Atlantic Books.
- Economist Intelligence Unit (EIU). 2010. *Democracy Index 2010: Democracy in Retreat*. Economist Intelligence Unit. (http://www.eiu.com/public/democracy_index.aspx), accessed January 25, 2011.
- Dworkin, Ronald. 1987. *Law’s Empire*. Cambridge: Harvard University Press.
- Dworkin, Ronald. 2000. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge: Harvard University Press.
- Elkins, Zachary. 2000. “Gradations of Democracy? Empirical Tests of Alternative Conceptualizations.” *American Journal of Political Science* 44(2): 287–294.
- Elster, Jon, ed. 1998. *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Epstein, S., N. Serafino, F. Miko. 2007. *Democracy Promotion: Cornerstone of U.S. Foreign Policy?* Washington, DC: Congressional Research Service.
- Finkel, Steve, Anibal Pérez-Liñán, Mitchell A. Seligson. 2007. “The Effects of U.S. Foreign Assistance on Democracy Building, 1990–2003.” *World Politics* 59(3): 404–439.
- Ford, Henry Jones. 1898/1967. *The Rise and Growth of American Politics*. New York: Da Capo.
- Foweraker, Joe, Roman Krznaric. 2000. “Measuring Liberal Democratic Performance: An Empirical and Conceptual Critique.” *Political Studies* 48(4):759–787.
- Freedom House. 2015. *Methodology*. Freedom in the World 2015. New York. (https://freedomhouse.org/sites/default/files/Methodology_FIW_2015.pdf), accessed December 2, 2015.
- Fung, Archon, and Erik Olin Wright, eds. 2003. *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. London: Verso.

- Gasiorowski, Mark J. 1996. "An Overview of the Political Regime Change Dataset." *Comparative Political Studies* 29(4): 469–83.
- Gates, Scott, Håvard Hegre, Mark P. Jones, Håvard Strand. 2006. "Institutional Inconsistency and Political Instability: Polity Duration, 1800–2000." *American Journal of Political Science* 50(4):893–908.
- Gerring, John, Carl Henrik Knutsen, Svend-Erik Skaaning, Jan Teorell, Michael Coppedge, Staffan Lindberg, Matthew Maguire. 2015. *Electoral Democracy and Human Development*. Varieties of Democracy Institute Working Paper No. 9.
- Gerring, John, Strom C. Thacker, Rodrigo Alfaro. 2012. "Democracy and Human Development." *Journal of Politics* 74(1): 1-17.
- Gleditsch, Kristian S., Michael D. Ward. 1997. "Double Take: A Re-examination of Democracy and Autocracy in Modern Politics." *Journal of Conflict Resolution* 41(3): 361–383.
- Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.
- Goldsmith, Arthur A. 2001. "Donors, Dictators and Democrats in Africa." *Journal of Modern African Studies* 39(3): 411-436.
- Goodnow, Frank J. 1900. *Politics and Administration*. New York: Macmillan.
- Gould, Carol C. 1988. *Rethinking Democracy: Freedom and Social Cooperation in Politics, Economy, and Society*. Cambridge: Cambridge University Press.
- Green, Andrew T., R.D. Kohl. 2007. "Challenges of Evaluating Democracy Assistance: Perspectives from the Donor Side." *Democratization* 14(1): 151-165.
- Guilhot, Nicolas. 2005. *The Democracy Makers: Human Rights and International Order*. New York: Columbia University Press.
- Gutmann, Amy, and Dennis Thompson. 1996. *Democracy and Disagreement*. Cambridge: Harvard University Press.
- Habermas, Jürgen. 1984. *The Theory of Communicative Action. Volume One: Reason and the Rationalization of Society*. Boston: Beacon Press.
- Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: MIT Press.
- Hadenius, Axel. 1992. *Democracy and Development*. Cambridge: Cambridge University Press.
- Hadenius, Axel, Jan Teorell. 2005. *Assessing Alternative Indices of Democracy*. Committee on Concepts and Methods Working Paper Series, August.
- Hamilton, Alexander, James Madison, and John Jay. 1992 [1787–88]. *The Federalist*. Ed. William R. Brock. London: Everyman/J.M. Dent.
- Hayek, Friedrich A. 1960. *The Constitution of Liberty*. London: Routledge & Kegan Paul.
- Hearn, Julie 2000. "Aiding Democracy? Donors and Civil Society in South Africa." *Third World Quarterly* 21(5): 815-30.
- Held, David. 2006. *Models of Democracy*, 3d ed. Cambridge: Polity Press.
- Hirst, Paul Q., ed. 1989. *The Pluralist Theory of the State: Selected Writings of G.D.H. Cole, J.N. Figgis, and H.J. Laski*. New York: Routledge.
- Högström, John. 2013. "Does the Choice of Democracy Measure Matter? Comparisons between the Two Leading Democracy Indices, Freedom House and Polity IV." *Government and Opposition* 48(2): 201–221.
- Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- Hyde, Susan, Nikolay. 2012. "Which Elections Can Be Lost?" *Political Analysis* 20(2): 191-210.

- Jackman, Simon. 2008. "Measurement." Pp. 111-151 in Janet Box-Steffensmeier, Henry Brady, and David Collier (eds), *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- Kaufmann, Daniel, Aart Kraay, Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators: Methodology and Analytical Issues*. World Bank Policy Research Working Paper No. 5430.
- Keyssar, Alexander. 2000. *The Right to Vote: The Contested History of Democracy in the United States*. New York: Basic Books.
- Kurtz, Marcus J., Andrew Schrank. 2007. "Growth and Governance: Models, Measures, and Mechanisms." *Journal of Politics* 69(2): 538–554.
- Landman, Todd. 2003. *Map-Making and Analysis of the Main International Initiatives on Developing Indicators on Democracy and Good Governance*. Unpublished manuscript, University of Essex.
- Landman, Todd, Edzia Carvalho. 2008. *Issues and Methods in Comparative Politics: An Introduction*, 3rd Edition. London: Routledge.
- Lijphart, Arend. 1997. "Unequal Participation: Democracy's Unresolved Dilemma." *American Political Science Review* 91(1): 1-14.
- Lijphart, Arend. 1999. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.
- Lindberg, Staffan I. 2006. *Democracy and Elections in Africa*. Baltimore: Johns Hopkins University Press.
- Lindberg, Staffan I., Michael Coppedge, John Gerring, Jan Teorell, et al. 2014. "V-Dem: A New Way to Measure Democracy." *Journal of Democracy* 25(3): 159-169.
- Lindblom, Charles E. 1977. *Politics and Markets: The World's Political-Economic Systems*. New York: Basic Books.
- Lively, Jack. 1975. *Democracy*. Oxford: Basil Blackwell.
- Lowell, A. Lawrence. 1889. *Essays on Government*. Boston: Houghton Mifflin.
- Macpherson, C.B. 1977. *The Life and Times of Liberal Democracy*. Oxford: Oxford University Press.
- Mainwaring, Scott, Daniel Brinks, Aníbal Pérez-Liñán. 2001. "Classifying Political Regimes in Latin America, 1945-1999." *Studies in Comparative International Development* 36(1): 37-65.
- Mansbridge, Jane. 1983. *Beyond Adversarial Democracy*. Chicago: University of Chicago Press.
- Marshall, Monty G., Ted Gurr, Keith Jagers. 2014. *Polity IV Project: Political Regime Characteristics and Transitions, 1800–2006*. (http://home.bi.no/a01110709/PolityIV_manual.pdf), accessed April 9, 2015.
- McFaul, Michael. 2005. "Transitions from Postcommunism." *Journal of Democracy* 16(3): 5–19.
- McHenry, Dean E., Jr. 2000. "Quantitative Measures of Democracy in Africa: An Assessment." *Democratization* 7(2): 168–185.
- McMahon, Edward R. 2001. "Assessing USAID's Assistance for Democratic Development: Is it Quantity versus Quality?" *Evaluation* 7(4): 453–467.
- Merkel, Wolfgang. 2004. "Embedded and Defective Democracies." *Democratization* 11(5): 33-58.
- Meyer, Thomas. 2007. *The Theory of Social Democracy*. Cambridge: Polity Press
- Mill, John Stuart. 1958 [1865]. *Considerations on Representative Government*, 3d ed., ed. Curren V. Shields. Indianapolis: Bobbs-Merrill.
- Møller, Jørgen, Svend-Erik Skaaning. 2013. *Democracy and Democratization*. London: Routledge.
- Møller, Jørgen, Svend-Erik Skaaning. 2010. "Marshall Revisited: The Sequence of Citizenship Rights in the Twenty-First Century." *Government and Opposition* 45(4): 457-483.
- Moon, Bruce E., Jennifer Harvey Birdsall, Sylvia Ciesluk, Lauren M. Garlett, Joshua H. Hermias, Elizabeth Mendenhall, Patrick D. Schmid, Wai Hong Wong. 2006. "Voting Counts:

- Participation in the Measurement of Democracy.” *Studies in Comparative International Development* 41(2): 3–32.
- Munck, Gerardo L. 2009. *Measuring Democracy: A Bridge between Scholarship and Politics*. Baltimore: John Hopkins University Press.
- Munck, Gerardo L. 2016. “What is Democracy? A Reconceptualization of the Quality of Democracy.” *Democratization* 23(1): 1-26.
- Munck, Gerardo L., Jay Verkuilen. 2002. “Conceptualizing and Measuring Democracy: Alternative Indices.” *Comparative Political Studies* 35(1): 5–34.
- Mutz, Diana C. 2008. “Is Deliberative Democracy a Falsifiable Theory?” *Annual Review of Political Science* 11: 521–38.
- Naess, Arne, Jens Christophersen, Kjell Kvalo. 1956. *Democracy, Ideology and Objectivity: Studies in the Semantics and Cognitive Analysis of an Ideological Controversy*. Oslo: Universitetsforlaget.
- Offe, Claus. 1983. "Political Legitimation through Majority Rule?" *Social Research* 50(4): 709-756.
- Offe, Claus. 1997. "Micro-aspects of Democratic Theory: What Makes for the Deliberative Competence of Citizens?" Pp. 81-104 in Axel Hadenius (ed.), *Democracy's Victory and Crisis*. Cambridge: Cambridge University Press.
- Pateman, Carole. 1976. *Participation and Democratic Theory*. Cambridge: Cambridge University Press.
- Paxton, Pamela. 2000. “Women’s Suffrage in the Measurement of Democracy: Problems of Operationalization.” *Studies in Comparative International Development* 35(3): 92–110.
- Pemstein, Dan, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, and Farhad Miri. 2015. “The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data.” University of Gothenburg, Varieties of Democracy Institute: V-Dem Working Papers Series No. 20
- Pemstein, Daniel, Stephen Meserve, James Melton. 2010. “Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type.” *Political Analysis* 18(4): 426–449.
- Pew Research Center. 2014. *Religion in Latin America: Widespread Change in a Historically Catholic Region*. <http://www.pewforum.org/files/2014/11/Religion-in-Latin-America-11-12-PM-full-PDF.pdf> (accessed March 13, 2015).
- Pickel, Susanne, Toralf Stark, and Wiebke Breustedt. 2015. “Assessing the Quality of Quality Measures of Democracy: A Theoretical Framework and its Empirical Application.” *European Political Science* 14(4): 496-520.
- Powell, G. Bingham. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*. New Haven: Yale University Press.
- Proyecto Estado de la Nación. 2001. *Informe de la auditoría ciudadana sobre la calidad de la democracia en Costa Rica*. San José, Costa Rica: Proyecto Estado de la Nación.
- Przeworski, Adam, Fernando Limongi. 1997. “Modernization: Theories and Facts.” *World Politics* 49(1) 155-183.
- Ranney, Austin. 1962. *The Doctrine of Responsible Party Government: Its Origins and Present State*. Urbana: University of Illinois.
- Reich, Gary. 2002. “Categorizing Political Regimes: New Data for Old Problems.” *Democratization* 9(4): 1-24.
- Ross, Michael. 2006. “Is Democracy Good for the Poor?” *American Journal of Political Science* 50(4): 860–74.
- Rousseau, Jean-Jacques. 1984. *A Discourse on Inequality*. Trans. by Maurice Cranston. Middlesex, England: Penguin.
- Ryfe, David M. 2005. “Does Deliberative Democracy Work?” *Annual Review of Political Science* 8:

- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64(4): 1033-1053.
- Saward, Michael. 1998. *The Terms of Democracy*. Oxford: Polity.
- Saward, Michael. 2003. *Democracy*. Cambridge: Polity.
- Schattschneider, E.E. 1942. *Party Government*. New York: Rinehart.
- Schedler, Andreas. 2012. "Judgment and Measurement in Political Science." *Perspectives on Politics* 10(1): 21-36.
- Schneider, Carsten Q. 2010. *Issues in Measuring Political Regimes*. Working Paper 2010/12, Center for the Study of Imperfections in Democracy, Central European University. (<http://disc.ceu.hu/working-papers>).
- Schneider, Carsten Q., Philippe C. Schmitter. 2004a. *The Democratization Dataset, 1974–2000*. Unpublished manuscript, Department of Political Science, Central European University.
- Schneider, Carsten Q., Philippe C. Schmitter. 2004b. "Liberalization, Transition and Consolidation: Measuring the Components of Democratization." *Democratization* 11(5): 59–90.
- Schumpeter, Joseph A. 1950 [1942]. *Capitalism, Socialism and Democracy*. New York: Harper & Bros.
- Seawright, Jason, David Collier. 2014. "Rival Strategies of Validation: Tools for Evaluating Measures of Democracy." *Comparative Political Studies* 47(1): 111–138.
- Sen, Amartya. 1999. "Democracy as a Universal Value." *Journal of Democracy* 10(1): 3-17.
- Skaaning, Svend-Erik, John Gerring, Henrikas Bartusevičius. 2015. "A Lexical Index of Electoral Democracy." *Comparative Political Studies* 48(12): 1491-1525.
- Steiner, Jurg, Andre Bächtiger, Markus Spornli, and Marco R. Steenbergen. 2004. *Deliberative Politics in Action: Analysing Parliamentary Discourse*. Cambridge: Cambridge University Press.
- Sudders, Matthew, Joachim Nahem. 2004. *Governance Indicators: A Users' Guide*. Oslo: UNDP.
- Thomas, Melissa A. 2010. "What Do the Worldwide Governance Indicators Measure?" *European Journal of Development Research* 22(1): 31–54.
- Thompson, Dennis F. 2008. "Deliberative Democratic Theory and Empirical Political Science." *Annual Review of Political Science* 11: 497–520.
- Treier, Shawn, Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1): 201–17.
- Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton: Princeton University Press and the Russell Sage Foundation.
- USAID. 1998. *Handbook of Democracy and Governance Program Indicators*. Technical Publication Series PNACC-390. Washington: USAID Center for Democracy and Governance.
- Vanhanen, Tatu. 2000. "A New Dataset for Measuring Democracy, 1810–1998." *Journal of Peace Research* 37(2): 251–265.
- Vermillion, Jim. 2006. "Problems in the Measurement of Democracy." *Democracy at Large* 3(1): 26–30.
- Vile, M.J.C. 1967/1998. *Constitutionalism and the Separation of Powers*. Indianapolis: Liberty Fund.
- Walzer, Michael and David Miller. 1995. *Pluralism, Justice, and Equality*. Oxford: Oxford University Press.
- Weale, Albert. 1999. *Democracy*. New York: St. Martin's Press.
- Wilson, Woodrow. 1956 [1885]. *Congressional Government*. Baltimore: Johns Hopkins University Press.
- Wu, Lawrence L. 2000. "Some Comments on 'Sequence Analysis and Optimal Matching

- Methods in Sociology: Review and Prospect.' ” *Sociological Methods and Research* 29(1): 41–64.
- Young, Iris M. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.
- Youngs, Richard (ed.). 2010. *The European Union and Democracy Promotion: A Critical Global Assessment*. Baltimore: The Johns Hopkins University Press.

Appendix A: Impact Evaluation

Policymakers are increasingly concerned with evaluating the impact of the programs that they fund and organizations and practitioners are facing increased demands for systematic and objective evaluation of their projects – both individually and at an aggregate level. Does democracy promotion enhance prospects of democracy? Could it lead to other undesirable outcomes such as greater instability?²⁴ Is the impact of democracy assistance uneven – having a positive effect in some contexts and a negative effect in others? How should democracy promotion programs be constructed and which sectors in/parts of a society should be targeted?

These are difficult, but nonetheless fundamental, issues. Thus, we want to consider carefully the extent to which V-Dem data might contribute to knowledge about program and project evaluation. We shall make the case that V-Dem data is valuable in two aspects: (1) identifying which areas and aspects of democracy to target, and (2) evaluating whether these efforts have produced desired outcomes.

Let us begin at the program level. Imagine a program focused on improving capacity among civil society groups in a country. Regardless of research design, the *proximal* impact of this program is best measured with very specific indicators – presumably focused on the groups that are being offered assistance (or might be offered assistance) – collected for this purpose. V-Dem has little to offer at this micro-level. V-Dem data, however, may be used to gauge whether an intervention has had societal impact, i.e., across civil society at-large. Here, V-Dem indicators focused on civil society may prove useful – more useful, at any rate, than highly aggregated indices such as those produced by Polity and Freedom House, which do not afford the opportunity to examine how particular aspects of democracy may have changed in response to an intervention.

Of course, the funder could also decide to collect their own (custom-designed) societal-level indicators for this purpose. However, there are several reasons to suppose that V-Dem might offer an advantage over custom macro-level indicators. First, custom data collection would significantly increase the cost of program evaluation. Second, one faces the problem of bias since those who construct and measure the outcomes of concern typically have a stake in the outcome. Third, in order to discern subtle changes in a trend-line that might be a product of

²⁴ For various sides of this issue see Bollen, Paxton, Morishima (2005), Burnell (2008), Carothers (1999, 2006), CEUDAP (2008), Chua (2003), Cox, Ikenberry, Inoguchi (2000), Crawford (2001), Epstein, Serafino, Miko (2007), Finkel, Perez-Linan, Seligson (2007), Goldsmith (2001), Green, Kohl (2007), Guilhot (2005), Hearn (2000), McMahan (2001), Young (2010).

a specific intervention it is usually necessary to construct a long time-series, extending back well before the initiation of the program. This may be difficult to do, or at least costly. Fourth, a custom set of indicators is likely to be restricted to a single country, prohibiting comparisons with neighboring countries. These comparisons often help in detecting spurious causal relationships. So, all things considered, we can identify several advantages of using V-Dem to measure country-level outcomes.

Now let us consider macro-level interventions. By this, we mean the entire set of democracy promotion efforts undertaken by large agencies such as the World Bank, the UNDP, USAID, or the DAC countries at-large. Our interest might be country-specific (how have DAC democracy promotion efforts affected the quality of democracy in Haiti?) or cross-national (how have DAC democracy promotion efforts affected the quality of democracy across the developing world?). At this level, V-Dem offers a significant advantage over extant democracy indices. First, V-Dem allows one to investigate the impact of policies on specific properties of democracy (rather than looking only at democracy at-large). It could be, for example, that democracy promotion is more successful in bolstering the quality of elections than in enhancing the quality of civil society. If so, this would provide important information for funders to consider in terms of what areas to focus on. Likewise, if funding categories are disaggregated – so that programs focused on elections can be distinguished from programs focused on civil society – then these specific connections can be probed. This is the approach taken by a recent and highly influential study of USAID democracy promotion (Finkel, Pérez-Liñán, & Seligson 2007). A shortcoming of this study is that the outcomes are often not well-calibrated to the inputs, a product of the paucity of disaggregated indicators of democracy when Finkel et al. conducted their analysis.

One should be especially cautious when attempting to infer causal relationships between (non-randomized) inputs and outputs measured at the country level. However, we must also bear in mind that randomized control trials (RCTs) are rarely viable at a societal level or with respect to distal outcomes. Here, V-Dem offers a significant improvement over extant indicators, as demonstrated above.

One further point may be added. One of the problems of evaluating democracy promotion efforts is determining whether a program was actually responsible for observed improvements when improvements may have happened anyway, due to other conditions. The advantage of V-Dem's highly disaggregated measures is that provides a set of placebo tests for any causal proposition. If foreign aid in a country focuses on civil society and not on other facets of democracy we might expect that improvements in the strength and independence of civil society would outstrip improvements in the quality of elections or the independence of the

judiciary (to mention just two out of many alternative outcomes). V-Dem indicators are designed to capture these discrepant tendencies.

Appendix B: Key Terms

V-Dem documents employ several terms which, in ordinary usage, have similar or overlapping meanings. Because our rather specialized project deals with conceptualization and measurement at multiple levels of specificity, we have found it necessary to assign more precise meanings to some of these terms to promote clear communication within the team. Some of these terms—*attribute*, *property*, *dimension*, *principle*, *conception*, and *variety*—refer to concepts. Some—*variable*, *indicator*, *index*, *scale*, and of course *measure*—refer only to measures. Yet others—*component* and *subcomponent*—can refer to either a concept or a measure, depending on the context.

Attributes (or sometimes *elements*) are the most specific conceptual building blocks we use to discuss democracy and related concepts. Many of our survey questions attempt to ask about a single attribute, for example, “What percentage of the lower (or unicameral) chamber of the legislature is directly elected in popular elections?” Although any of these questions could also be seen as a compendium of multiple attributes (What does it mean to be a legislature? What is a “popular” election?), in a project covering all countries for more than a century, there are degrees of specificity that it is not practical to approach, so attributes are the most specific concepts that we consider feasible to measure.

In V-Dem usage, *properties* are more general concepts than attributes. We speak of the participatory property of participatory democracy, for example, to call attention to the participatory aspects of participatory democracy, as distinct from the other features that it may share with egalitarian, liberal, electoral, or deliberative democracy. Because they are at a relatively high level of generality, properties tend to contain many attributes.

A *dimension* is a property with an added empirical characteristic: it describes a straight line connecting two poles of a concept. It is practically synonymous with *scale*. Often we reserve the term dimension for properties whose attributes also can be arrayed between the same two poles. For example, if “civil liberty” is a dimension, many specific civil liberties are correlated: if a case has a high degree of freedom of discussion, it tends to have high degrees of freedom of movement, freedom to organize, freedom from political murder, and so on. There are exceptions, however, when there are accepted ways of reducing multidimensional attributes to a single dimension. For example, male suffrage and female suffrage vary somewhat independently but they can be combining into a dimension of adult suffrage.

When we wish to make reference to the various intellectual traditions that have fostered debate about what democracy should be, we prefer the term *principles*. Principles are properties with normative connotations. For example, when describing theories of deliberative democracy,

it is necessary to refer to philosophers such as Habermas who argue for the principle that governments must earn their authority to rule by respectfully providing citizens with rationales for their decisions—a normative claim. However, by referring to various principles we are not endorsing them, only saying that others do. Principles are not necessarily dimensions, as realizing a principle can require achieving a high standard on more than one dimension; and dimensions are not necessarily principles; but both are special types of properties.

At the most general level, we speak of *conceptions* of democracy. These are more complex notions that allow for a version of democracy to embrace multiple properties and dimensions. They are attempts to define more holistic, thick concepts that approach natural-language understandings of democracy. In doing so, they tend to overlap with other general concepts of democracy. For example, our conceptions of liberal, participatory, deliberative, and egalitarian democracy all include electoral democracy and therefore overlap quite a bit. *Varieties of democracy* is a looser term that could refer both to these conceptions and to other notions of democracy, such as direct democracy (which in our conceptual scheme is a property of participatory democracy).

Very often we refer to *components* and *subcomponents*. These have relative meanings, which are useful when describing the structure of either a concept or an index. For example, egalitarianism is a component of egalitarian democracy, but egalitarianism in turn has its own components, including health and educational equality, which are therefore subcomponents of egalitarian democracy. The V-Dem conceptual scheme sometimes distinguishes five or more levels of specificity. Because these terms are relative, knowing whether a concept is a component or a subcomponent does not reveal how general or specific it is in an absolute sense.

Variables or *indicators* are interchangeable terms for measures of a small number of attributes. Properties, principles, dimensions, and conceptions contain many attributes. Measuring them therefore requires an *index* (plural: *indices*), which we define as a measure constructed from multiple variables or indicators. Components and subcomponents can be measured by either variables or indices, depending on how general they are. For example, we define several fairly general “component indices.” However, at times we also refer to variables and indices as components or subcomponents when describing the hierarchical relations among them. *Measures* is the umbrella term for variables, indicators, and indices.

Appendix C: Search Terms

Index	Google Hits	Google Scholar Cites
Bernhard...: <i>BNR</i>	“Bernhard” “Nordstrom” and “Reenock”	Economic Performance, Institutional Intermediation, and Democratic Survival
Bertelsmann: <i>BTI</i>	“Bertelsmann Transformation Index”	Bertelsmann Transformation Index – all editions
Boix...: <i>BMR</i>	“Boix Miller Rosato”	Boix, Miller and Rosato
Coppedge...: <i>Inclusiveness</i>	“Coppedge, Alvarez, Maldonado” “Inclusiveness”	Two Persistent Dimensions of Democracy: Contestation and Inclusiveness
Coppedge...: <i>Contestation</i>	“Coppedge, Alvarez, Maldonado” “Contestation”	Two Persistent Dimensions of Democracy: Contestation and Inclusiveness
EIU: <i>EIU index</i>	“EIU index of democracy”	EIU index of democracy
FH: <i>Civil liberty</i>	“Freedom House” “Civil Liberties”	Freedom House Civil liberties
FH: <i>Countries at Crossroads</i>	“Freedom House” “Countries at the Crossroads”	Freedom House Countries at the Crossroads
FH: <i>Nations in Transit</i>	“Freedom House” “Nations in Transit”	Freedom House Nations in Transit
FH: <i>Political rights</i>	“Freedom House” “Political Rights”	Freedom House Political Rights
Merkel...: <i>Democracy Barometer</i>	“Democracy Barometer”	The Democracy Barometer: A New Instrument to Measure the Quality of Democracy and Its Potential for Comparative Research – all editions
Pemstein...: <i>UDS</i>	“Unified Democracy Scores”	Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type
Polity IV: <i>Polity2</i>	“Polity IV”	Polity IV
Przeworski...: <i>DD</i>	“Democracy and Dictatorship” Cheibub	Democracy and Dictatorship, plus the Revisited
Skaaning...: <i>Lexical</i>	“Lexical Index of Electoral Democracy”	Lexical Index of Electoral Democracy
Vanhanen: <i>Competition</i>	“Index Competition” Vanhanen	New Dataset for Measuring Democracy, 1810–1998
Vanhanen: <i>Participation</i>	“Index Participation” Vanhanen	New Dataset for Measuring Democracy, 1810–1998
WGI: <i>Voice & Accountability</i>	“Voice & Accountability” WGI	Worldwide Governance Indicators: Methodology and Analytical Issues – all versions
V-Dem: <i>[various]</i>	“Varieties Of Democracy project”	Varieties of Democracy Codebook

Note: Search terms for results posted in the final columns of Table 1.