

**Bringing Measurement Back In:  
Methodological Foundations of the Electoral Democracy  
Index**

**Gerardo L. Munck and Jay Verkuilen**

[munck@usc.edu](mailto:munck@usc.edu)  
School of International Relations  
University of Southern California

[jayv@uiuc.edu](mailto:jayv@uiuc.edu)  
Department of Psychology, Quantitative Division  
University of Illinois, Urbana-Champaign

Paper prepared for delivery at the 2003 Annual Meeting of the  
American Political Science Association,  
Philadelphia, August 28 - 31, 2003.

Copyright by the American Political Science Association.

Measurement of the core concepts used in the study of politics has a great impact on our knowledge about politics. It affects the way we describe the world—and hence the questions that seem important to study—and the causal propositions we consider to be valid. Even more importantly, it affects the kind of advice political scientists offer on the key political and public policy issues of the day. Yet it is remarkable how little attention is given to the generation of data and the methodology of measurement. Indeed, it is fair to say that the mainstream view in political science is that measurement is a necessary task that should be quickly transcended or, if at all possible altogether skipped, so that researchers' energies can focus on a task seen as much more important: the testing of causal hypotheses.

The failure to acknowledge the importance of data generation and the methodology of measurement, and the propensity to take shortcuts, has severe costs. In a nutshell, it is associated with illusionary gains in knowledge, which sooner or later are questioned. Thus, it is time that political scientists become more suspicious of the tendency toward premature knowledge claims that is so pervasive in the discipline and put more emphasis on the measurement of key concepts as a foundation of knowledge, that is, as a task that affects the possibility of providing sound descriptive and/or causal analysis and, ultimately, of offering responsible advice.

Some hopeful signs of an appreciation of the important of measurement can be gleaned in various quarters. Various important, ambitious research projects have tackled the issue of data generation as an integral part of the research process.<sup>1</sup> In

---

<sup>1</sup> Examples include Lijphart's (1999) work on consensus democracy, which dedicates great attention, chapter after chapter, to the measurement of the key concepts it uses; Przeworski et al.'s (2000) work on democracy and development, which called for the construction of a new index on democracy and the compilation of a data set of a range of political and economic concepts; Fearon and Laitin's (2003) ongoing work on ethnic conflict, a project that has gone hand in hand with efforts to measure

addition, some leading journals in political science have published articles on the methodology of measurement.<sup>2</sup> But these are exceptions. The standard practice in most quantitative research is to draw upon readily available data sets that have been used to the point that little new insight can be drawn from them and that, to begin with, are fairly poor measures of the concepts that are used in theorizing. The problem with the qualitative literature, though of a different nature, tends to be as serious. In this type of research, the link between theory and observation is usually given more consideration. But the richness of narrative accounts is routinely delinked from an effort to systematically code the cases. In turn, efforts to develop the methodology of measurement in political science lag considerably behind other disciplines. Indeed, the sort of attention to measurement that is routine in disciplines such as psychology and sociology, and fields such as education, is quite foreign to political science. Thus, it is no surprise that the most significant advances in the methodology of measurement do not come from political science.

This paper seeks to respond to the need to bring measurement back in to political science. It does so through an in-depth discussion of the methodological issues involved in the measurement of democracy, a key concept in political science and, more broadly, in the social sciences. More concretely, it does so by offering a detailed exposition of the methodological choices followed in the construction of a new index of democracy, the electoral democracy index (EDI).<sup>3</sup> As the discussion shows, the generation of data is an extremely complex aspect of the research process

---

ethnic fractionalization (Fearon 2002); and Schmitter's work with Schneider on democracy (Schmitter and Schneider 2003).

<sup>2</sup> See, among others, Kritzer (1996), Jacoby (1999), and Adcock and Collier (2001).

<sup>3</sup> This index has been prepared for a new *Report on Democratic Development in Latin America*, an initiative of the Regional Bureau for Latin America and the Caribbean of the United Nations Development Programme (UNDP). In constructing this index we followed both the methodology of measurement and the specific recommendations made to improve upon existing democracy indices outlined in Munck and Verkuilen (2002). Therefore, this paper is a companion piece to our previous work on the conceptualization and measurement of democracy.

that involves an iterative process of theorizing and testing, along with a range of consequential choices, which are seldom appreciated and seriously discussed. Thus, by explaining and illustrating the issues involved in the generation of indices, that is, aggregate data, we seek to encourage further discussion of the methodology of measurement, and to promote a more self-conscious approach to the generation of data, within political science.

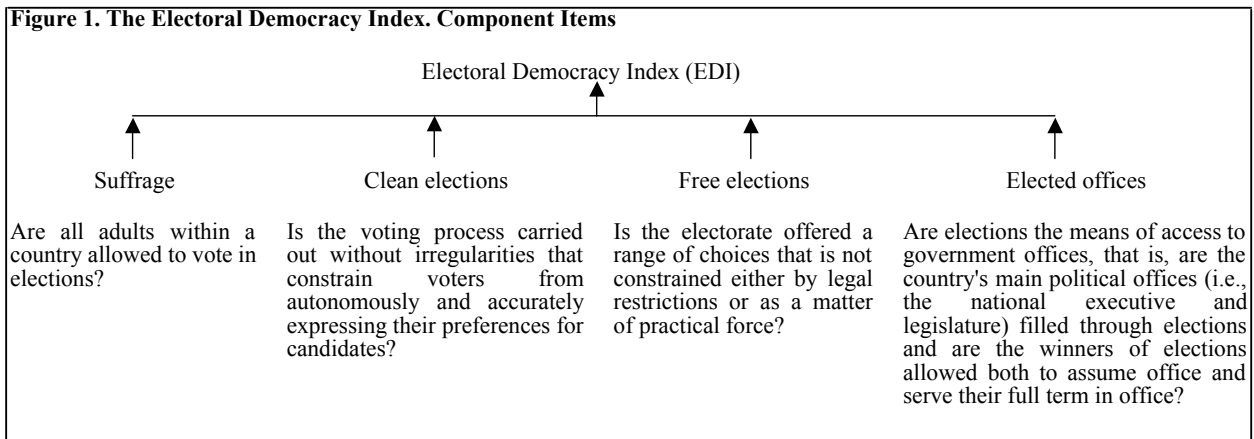
The paper is organized in terms of the three broad challenges that must be tackled in the generation of indices: conceptualization, measurement and aggregation (Munck and Verkuilen 2002).<sup>4</sup> In each section we discuss the theoretical considerations and empirical tests that informed the multiple choices involved in data generation. We conclude with a few points about the recommended interpretation and use of the EDI. (The data themselves will be presented at the APSA panel.)

## **I. Conceptualization: The Selection of Component Items**

The first and probably most important step in the construction of the EDI—the choice of component items to be included in the index—led to the identification of four items—suffrage, clean elections, free elections and elected offices—which are briefly defined in Figure 1.

---

<sup>4</sup> Further discussion of the mathematics used here—based on fuzzy set theory—can be found in Smithson (1987), Verkuilen (2002), and Smithson and Verkuilen (forthcoming).



These items were selected for a number of reasons. First and foremost, they correspond to the core elements that theorists of democracy have routinely invoked in defining a democratic regime, and tap into a series of questions that are generally considered central, indeed necessary, to an assessment of the democratic character of a political regime (Dahl 1971, 1989, 1998; Rokkan 1970: Ch. 4; O'Donnell 2001).

Second, all four items refer to citizenship rights that are the state's responsibility to uphold and can be clearly interpreted in terms existing democratic theory. Thus, these items avoid the problems associated with items such as voter turnout or electoral disproportionality that reflect both state choices and citizens' actions. This helps ensure that the index can be clearly interpreted as a measure of the extent to which the state guarantees citizenship rights regarding the political regime, as opposed to aspects of citizen action. Likewise, they avoid the problems associated with measures that have an unclear implication for the democraticness of the regime, such as the difference between proportional and majoritarian electoral rules or between presidential and parliamentary systems. These other aspects are

certainly not unimportant but they are not as clearly connected to the democraticness of the regime as the four chosen.

Third, these items were selected because they could be measured on the basis of observables on the basis of a standardized methodology. This criterion meant that some components that might have been included but that would require a reliance on surveys of perceptions were dropped. Moreover, this criterion meant that certain components that might have been measured on the basis of observables were not included, due to the practical constraints imposed by the decision to not relying on readily available data based on observables that nonetheless was based on a different methodology.

## **II. Measurement: The Scoring of Component Items**

The second and most labor intensive step in the construction of the EDI—the measurement of the EDI's four component items—involved a range of choices, starting with the construction of scales and the definition of the coding rules, moving on to the coding process itself, to the generation of a rectangular data set with normalized scores, the ultimate product of this stage in the measurement process. In addition to explaining and justifying how each choice was made, the results of a sensitivity analysis, aimed at assessing the precision of measurement of the membership function, are presented.

*Scales and Coding Rules.* The first set of choices pertained to the coding rules governing the coding process. The scales—three five-point ordinal scales and one three-point ordinal scale—were constructed by initially determining theoretically meaningful endpoints and then identifying distinct scale values as far apart conceptually as possible, starting with the midpoint. That is, the scale values were

chosen to reflect distinctions identified in the literatures as being important, avoiding verifiable but nonetheless small variations between cases (see Table 1). For cases not seen as fitting precisely any of the points on the various ordinal scales, the possibility of using pluses and minuses—analogueous to the assignment of academic grades—was introduced as a means to record intermediate values. This extra “fine tuning” step was taken as quite secondary to the basic assignment of scale values, however.

**Table 1. Scales of the EDI's Component Items**

*Suffrage:* This component responds to the question whether all adults within a country are allowed to vote in elections. Beyond the holding of elections, the measure does not include procedures that might hamper the effective use of the right to vote, such as access to the polls.

0 = no election is held to install the government;  
1 = only some men have the right to vote (there are restrictions related to property, gender and literacy);  
2 = most men enjoy the right to vote (there are restrictions related to gender and literacy);  
3 = most men and most women enjoy the right to vote (there are restrictions related to literacy);  
4 = the right to vote is universally recognized (however, even in countries with universal suffrage some restrictions may exist, affecting groups such as the military, the police, the clergy, foreign residents and citizens living abroad)

*Clean Elections:* This component responds to the question whether the voting process is carried out without irregularities that constrain voters from autonomously and accurately expressing their preferences for candidates. It does not include issues related to the competitiveness of the elections nor even if the winner of an election is allowed to assume office and if all the offices are elected.

0 = major irregularities in the voting process that have a determinative effect on the result of elections (e.g. alteration in the election for the national executive and/or the balance of power in parliament);  
1 = significant irregularities in the voting process (e.g. intimidation of voters, violence against voters,  
2 = lack of significant irregularities in the voting process (e.g. elections that might include "technical" irregularities but not any systematic bias of considerable weight).

*Free Elections:* This component responds to the question whether the electorate is offered a range of choices that is not constrained either by legal restrictions or as a matter of practical force. The measure does not include factors that affect the ability of parties and candidates to compete in equality of conditions, such as public financing, access to the mass media and the use of public resources.

0 = single party system;  
1 = ban on major party;  
2 = ban on minor party;  
3 = restrictions of a legal or practical nature that significantly affect the ability of potential candidates to run for office and/or the formation of political parties (for example, systematic assassinations and intimidation of candidates, ban on popular candidates, legal or practical restrictions that prevent the formation of parties or that lead parties to boycott the elections);  
4 = essentially unrestricted conditions for the presentation of candidates and the formation of parties.

*Elected Offices:* This component responds to the question whether elections are the means of access to government offices, that is, whether the country's main political offices (i.e., the national executive and legislature) are filled through elections and the winners of elections allowed both to assume office and serve their full term in office.

0 = none of the main political offices are filled through elections or all of the main political office holders are forcefully displaced from office and replaced by unconstitutional rulers;  
1 = a few political offices are filled by winners of elections or most of the main political office holders are forcefully displaced from office and replaced by unconstitutional rulers;  
2 = the president or the parliament are not elected or are forcefully displaced from office and replaced by unconstitutional rulers;  
3 = the president or the parliament are elected but the president is displaced from office and/or replaced by semiconstitutional means; or a significant number of parliamentarians are not elected or are forcefully displaced from office;  
4 = All of the main political offices are filled by elections and none of the main political office holders are displaced from office unless their removal from power and their replacement is based on strictly constitutional grounds.

The scales were constructed in such a way that each point on the scales corresponded to relatively concrete situations and events, and that coding decisions could be conducted based on clearly delineated standards and, as indicated above, strictly on the basis of observables. Data based on surveys of perceptions were not included in the index. Furthermore, as a way to further ensure the replicability of the coding exercise, as well as to guard against arbitrariness, the importance of

documenting the basis for coding decisions through reference to publicly available sources of information was stressed.

Scores for each case in each year were not required. Rather, scores for three of the components—suffrage, clean elections and free elections—were only required for those years when elections were held. The conditions for elections are affected by events and decisions made between elections and the coding drew upon information between election periods. But the significance of these events and decisions for the process whereby actors gain access to government offices, the core concern in the measurement exercise, becomes crystallized in the electoral event itself. Thus, though scores were assigned to some components only during elections years, these scores were seen as synthesizing a broader process.

*The Coding Process.* The second set of choices concerned the coding process itself. In this regard, two supplementary processes were used to code the cases. An initial coding was conducted by a single coder, on the basis of extensive research and consultations with a large number of experts over a period of many months. In addition, the assigned scores were presented and discussed in depth in various meetings, included one with a group of invited participants who work in diverse settings (politics, academia, international organizations) and come from different countries in the Americas (Argentina, Brazil, Canada, Colombia, Ecuador, Mexico, the United States and Uruguay). These discussions led to the identification of disagreements, which led to further research, and yet further group discussions. In the end, through this iterative process, a large degree of consensus emerged concerning the coding of the EDI's four dimensions.

*The Generation of a Rectangular Data Set with Normalized Scales.* The third set of choices pertained to the transformation of the scores on the component scales into a

rectangular data set, that is, a data set that includes numerical scores for all cases on all variables and all years, with normalized scales.

There were several steps taken to provide numbers in all years. First, a number of fairly mechanical issues were tackled. Pluses and minuses were turned into numbers by adding or subtracting 0.33 from base score (for e.g., a 3 plus was turned into a 3.33). Dashes (-) used to indicate that the assignation of a score was not applicable, because the government not elected, were turned into zeros (0). In addition, the scores for two of the components that were assigned scores only for the year in which an election was held—suffrage and free elections—were extended to the intervening years by simply carrying over the score for a given year to subsequent years, until a new score has been assigned (either because an election was held after a period in which there was a non-elected government, or a new election was held, or the election process was interrupted). The justification for this procedure is that the way in which a government originates continues to be a characteristic that affects its nature beyond the moment of its installation. In the case of the clean elections item a slightly more complex procedure was followed. A three-point scale was used in coding this item for reasons of interpretability. Nonetheless, the 1 in this scale does not actually represent a midpoint; it is a lot closer to the 2. Thus, each 1 was turned into a 3 and each 2 into a 4. In addition, because this item distinguishes the values assigned for elections for president and parliament, scores are not simply carried from election to election. Rather, the scores are an average of the scores for the elections for president and parliament.

A second issue dealt with was the problem entailed by the assignation of only one score per country per year. This practice, which is standard, obeys reasons of parsimony and is well justified. After all, the point of generating an index is to

offer a synthesis of the state of a country. But it does run into a variety of problems because the situation in countries changes over the course of a year and only one score is used to characterize the entire year. In some cases the solution is relatively easy. Thus, when a key event such an election was held toward the end of the year, the change in status due to that event was registered in the following year.<sup>5</sup> When events were held in the first half of the year, they were computed in that year.<sup>6</sup> In other cases, however, the solution was harder. In some cases, even though an event happened in the second half of the year it was registered in that year.<sup>7</sup> Also problematic were cases when more than one critical event happens in one year.<sup>8</sup>

Third, the component scales were normalized, that is, translated into a common metric, through a simple linear norming to the unit interval:

$$\text{normed value} = \text{raw value} / \text{maximum raw value}$$

Unfortunately nearly any choice made would be somewhat arbitrary because there are no widely acceptable units of electoral freedom comparable to units such as kilograms or dollars. But the choice of normalization procedure, as applied to the five-point ordinal scales—with the modification introduced in the clean elections item when transforming the scores on the component scales into a rectangular data

---

<sup>5</sup> For example, even though elections ending a period of military dominated governments were held in Guatemala in 1985, these elections were held in late 1985 and only led to the change in government in January 1986. Thus, even when scores were assigned for the components items for 1985, in computing the EDI these scores were entered for 1986.

<sup>6</sup> For example, El Salvador's 1994 election was held in March and led to a change in government in June. Thus the change was registered in 1994.

<sup>7</sup> This is the case, for example, of the fraudulent elections in the Dominican Republic held in May election, leading to the inauguration of the new president in August.

<sup>8</sup> For example, in 2000 Peru held two heavily questioned elections, in April and May, leading to the inauguration of Fujimori as president in July and then his resignation in November. In this case, the problematic elections were registered in 2000 and the rectification of the situation was registered as occurring in 2001.

set, the scales used to measure all four component items were five-point ordinal scales—is transparent and justifiable.

A major objection is likely to be that the scales are not ratio scales per se but rather five-point ordinal scales with an interval established by investigator fiat. Thus a cogent objection to norming into the unit interval is that it uses inadmissible “meaningless” mathematics. To address this concern, we quote Warren Torgerson (1958: 23-24) at length:

There is certainly nothing *wrong* or *logically incorrect* with the [use of investigator fiat].... In all these cases, one or more observable properties are selected which on a priori grounds are judged to be related to the concept of interest. A measure of the observable property itself or of a simple or weighted sum of several such observable properties is taken as the measure of the concept of interest. Ordinarily, the values so obtained are treated as though they were measures on an interval or ratio scale. This is clear, for example, when curves are fitted by the process of least squares or when product-moment correlations, means, or standard deviations are computed. All these presuppose that distance has meaning. Hence, either explicitly or implicitly, the experimenter is measuring the attribute on an interval scale whose order and distance characteristics have obtained meaning initially through definition alone.

The discovery of stable relationships among variables so measured can be as important as among variables measured in other ways. Indeed, it really makes little difference whether the present scale of length, for example, had been obtained originally through arbitrary definition, through a relation with other, established variables, or through a fundamental process. The concept is a good one. It has entered into an immense number of simple relations with other variables. And this is, after all, the major criterion of the value of a concept. If the result is the same, the way in which a concept is originally introduced is of limited importance.... [But] the major difficulty with measurement by fiat is the tremendous number of ways in which such defined scales can be constructed.

The question, then, is how our choice can be justified. And to this question the following response can be offered. First, all the scales have theoretically meaningful endpoints and can be assumed to fall in the unit interval, with 0 indicating a total lack of the property and 1 indicating full possession of the property. That is, the bottom value of the ordinal scale corresponds to the negation of the property in question while the top value corresponds to the complete

possession. A case with suffrage 0 has no suffrage at all while a case with suffrage 1 after norming has complete adult suffrage, the theoretically established standard. Thus, the problem of distance relates only to the points between the endpoint. Second, most of the scales were constructed in such a manner that each point on the scale is theoretically interpretable and in such a way that the distinct scale values are as far apart conceptually as possible. That is, the scale values were chosen to reflect distinctions identified in the literatures as being important, avoiding verifiable but nonetheless small variations between cases. Therefore, the likelihood of introducing any large error is relatively small. Though other, more sophisticated psychometric methods could be used, they are more complicated and less accessible, heavily data dependant and frequently do not perform all that much better than this simple procedure. In this sense it bears emphasizing that even though statistical optimality is important, it is not the only relevant criterion.

*Intercoder Reliability and Error Estimates.* Due to reasons of time, a formal intercoder reliability test was not conducted. Nonetheless, to gain a sense of whether other coders might have provided somewhat different assignments of the EDI's component items, the precision of measurement of the membership function was assessed through a sensitivity analysis (Saltelli et al. 2000). This analysis relies on perturbations of the assigned codings according to an experimental design and an examination of the resulting overall "replicated" indices. Thus, it uses mathematics to create "virtual" coders that are biased in various ways, for instance downwardly biased in rating one or more components. The results of this test showed that the EDI is quite stable—rank correlations with all the other "replications" were .99 or greater—and the shifts in mean and spread were quite predictable, showing a predictable negative or positive bias depending on the cell in the experimental

design. This test also provided some basic error bars for the EDI based on the “replications.” Generous bands for index values from around .25 to .75 are about  $\pm.07$  and reasonably conservative bands are  $\pm.1$ . By the most conservative possible standard, the EDI values are within  $\pm.2$ . The width is reasonably constant throughout the cited interval, but the precise boundaries depend on the value of the index and are in general narrower near the endpoints. A check on the mathematics was conducted using the inversion of the well-known and highly conservative Kolmogorov-Smirnov test for the distribution function—based on entirely different mathematics—and similar results were produced.

### **III. Aggregation: The Reduction of the Data**

The third step in the construction of the EDI—the aggregation of the EDI’s four component items—involved a choice with regards to the aggregation rule. In addition, various tests were conducted to assess the robustness of the proposed aggregation rule, the dimensionality of the component items, and influence of the component items on the EDI. Thus, after explaining and offering an initial justification for the choice of aggregation rule, these tests are presented and their results are interpreted.

*The Choice of Aggregation Rules.* The choice of aggregation rule that formalizes the relationship among the index’s component items was resolved through the use of a simple aggregation rule. The core insight used in selecting an aggregation rule is the well-established view that the EDI’s four component items are parts that constitute a system by virtue of the way in which they combine together and, moreover, that these four component items are so fundamental to the overall characterization of the regime that the absence of any one would simply render the regime non-democratic

(Sartori 1987: 184-85; Przeworski, Alvarez, Cheibub and Limongi 2000: 28-29, 57; O'Donnell 2001: 10, 12). For example, as theorists of democracy have long argued, the fact that Soviet-type systems had elections with full suffrage was meaningless, from the perspective of democracy, because the electorate did not have a choice among alternative candidates and because these elections did not lead to the occupation of offices that effectively exercised state power. Therefore, the four component items of the EDI are posited as individually necessary conditions, which are unsubstitutable and have equal weight.<sup>9</sup> This conception is formalized by calculating the product of the value of each of the component items. In formal terms, then, the EDI is calculated according to the following equation:

$$\text{Electoral Democracy Index} = \text{Suffrage} \times \text{Clean Elections} \times \text{Free Elections} \times \text{Elected Offices}$$

This equation captures a key insight of democratic theory: that when any one component item is totally lacking, the regime should simply be considered as non-democratic. In effect, this operation ensures that a value of zero on any of the four component items leads to classifying a case as a non-democracy. This is a “tough” standard, which can be seen as less “forgiving” than other aggregation rules. Thus, it bears emphasizing that precisely because this conception of the component items of the EDI as individually necessary conditions is highly demanding, it is used in conjunction with a conservative criterion in assigning zeros, indicating the total lack of a certain property, to the component items. This is so both in the sense that the scales were constructed in such a way that a zero would be called for only in the extreme cases where a property that is widely seen as vital to the existing of democracy is totally lacking, and in the sense that the evidence required for

---

<sup>9</sup> Whether these items might be considered jointly sufficient is a different matter.

assigning a zero had to be compelling. Thus, only when democratic norms are blatantly and undisputedly disregarded would the EDI code a country as being non-democratic.

*The Robustness of Aggregation Rules.* A test was conducted comparing four possible aggregation rules for combining the EDI's component items: the product of the four components actually used for the EDI, the minimum value of the four scaled components, the geometric mean of the four components, and the arithmetic mean of the four components. The results showed that no matter which rule is used, the rank correlations are all very high, indicating that the general ordering of cases is preserved. But there are differences between the indices, with the arithmetic mean and the geometric mean being similar to each other, and the minimum and the EDI being similar to each other. The most important difference is between the means and standard deviations (SDs). The arithmetic and geometric means have means of .92 and .91, respectively, and SDs of .20 and .21, respectively. In contrast, the minimum and the EDI have means of .84 and .82, respectively, and SDs of .26 and .28, respectively. This suggests that the latter rules do a better job at spreading the cases out and avoiding a bunching up of cases that makes it hard to interpret their differences with much clarity. In any event, many cases are assigned a value of 1, which suggests that the EDI scale is "topping out" in the current time period.

*The Dimensionality of the Component Items.* A test of the scalability of the EDI's four component items gave a Cronbach's alpha of .92, suggesting that the EDI is a measure of a unidimensional phenomenon. However, when a test was performed on two periods (1960-1985 and 1990-2002), the resulting Cronbach alphas were of .95 and .23 respectively. This indicates that while in the earlier period the components were unidimensional, this is no longer true in the post-1990 period. This finding is

consistent with the theory used in selecting the aggregation rules for the EDI. It also makes sense with the case histories, as the earlier period was one in which authoritarian regimes were generally complete in their suppression of democracy. In effect, it is important to note that standard additive measurement models rely on the assumption that aggregation operates on multiple parallel measurements. In contrast, because the component items of the EDI are theoretically considered unsubstitutable, the decision to aggregate up to one single score is not invalidated by any potential deviations from unidimensionality. The EDI in this sense is much more similar to the production function in economics than the ability test from psychometrics for which the Cronbach's procedure was devised. The factors of production for democracy may or may not move in tandem but lack of any of them renders their product unavailable. Thus, this scalability test offers further validation for the choice of the proposed aggregation rules as opposed to the quite common additive rule.

*Influence of the Component Items.* An indirect test of the influence of the component items further supports the view that the components of the EDI do not vary in the same way across the entire time period that is covered. The results of this test are presented in Table 2, which shows the means (standard deviations in parentheses) for the normalized components, and the EDI, unconditionally and conditional on time period (< 1990 or not). And they can be summarized as follows. Prior to 1990 all the EDI's components vary. This is particularly true because of the presence of cases of pure authoritarianism. The free elections item has the lowest mean, indicating that it is the most frequently violated component, but all components are down. In contrast, after 1990, the suffrage item has no variance and drops out entirely. And the clean elections and free elections items vary the most, hence

entailing the most variation in the EDI. This confirms the impression one would get looking at the component data tables. But a key point is also that most cases are at the top of the rating.

**Table 2. Component Means and Standard Deviations**

<b>Sample</b>	<b>Suffrage</b>	<b>Clean elections</b>	<b>Free elections</b>	<b>Elected offices</b>	<b>EDI</b>
Unconditional (n = 288)	.96 (.19)	.90 (.22)	.89 (.25)	.94 (.20)	.82 (.27)
< 1990 (n = 54)	.76 (.40)	.73 (.39)	.60 (.43)	.78 (.40)	.52 (.40)
≥ 1990 (n = 234)	1.00 (0.00)	.94 (.13)	.96 (.09)	.98 (.08)	.89 (.17)

#### **IV. Conclusion: Interpreting and Using the EDI**

The EDI is a 0.00-1.00 scale, with 0.00 indicating non-democracy and any number above 0.00 indicating a degree of electoral democracy, with higher scores referring to greater degrees of electoral democracy. To avoid confusion, it is important to note that the index should not be interpreted as an evaluation of a government's actions. Rather, it is a measure of a state of a system that is affected by the action and inaction of a government, as well as other state agents and societal actors. Moreover, it bears stressing that the concept that is measured is *electoral* democracy. This concept is not as narrow as some see it. Thus, though it is focused squarely on the holding of inclusive, clean and free elections, it encompasses more than "mere elections." Not only are the conditions for holding such elections affected by developments between elections. In addition, it calls for a consideration of what happens to the governments themselves between elections. But the EDI is certainly not a broad measure of democracy. Rather, it is a measure of a conception of a democratic political regime based on the most widely shared understanding about fundamental political rights. And this is very significant. On the one hand, this

means that any shortcomings detected by the EDI must be considered as important restrictions of citizen's political rights. On the other hand, the fact that a country received a perfect score of 1.00 should not be interpreted as having no room to improvement either with regards to dimensions not included in the index or relative more demanding standards of the component items of the EDI.

The EDI can be used for comparative purposes, whether to compare a country with itself or to other countries. Of these two uses, a comparison of a country with itself at a different point in time is usually easier to interpret. After all, a country may well have made notable improvements yet fall behind other countries if these other countries have made even greater strides. However, in either case, it is crucial to stress that any such comparisons should always be based on substantial and never on small differences. The reason for this is that the EDI, as any index, has a certain degree of measurement error and within the error bounds, it is inadvisable to make strong statements about differences. Indeed, as was estimated through the sensitivity analysis, generous error bars for EDI values between about .25 and .75 are roughly  $\pm .07$ . Thus, any pair of cases that differ by less than this value—for example a country with a EDI of .85 and one of .95—are simply too close to validly distinguish. Hence, it is methodologically unjustifiable to offer an overly precise ranking of countries, as is commonly done in the context of other indices, which simply transforms the scores of the EDI into a ranking without taking into consideration the degree of uncertainty associated with the EDI scores.

The identification of benchmark cases that are prototypical representations of the features that are associated with a range of scores can help to give concreteness to the meaning of each number. Moreover, the EDI can be used as a flag, in the sense that the specific scores of each country invite the reader to go back to the

tables on the component items to identify precisely what feature or features account for a country's score. In this way the EDI can be used as a valuable analytical tool, in that it offers a summary score that allows its users to identify the distinctiveness of each country's political regime in terms of its various component items but also in terms of the relationship among the constituent parts of the regime and their contribution to the whole.

### References

- Adcock, Robert N. and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review* Vol. 95, N° 3 (September): 529-46.
- Dahl, Robert A. 1971. *Polyarchy* (New Haven, CT.: Yale University Press).
- Dahl, Robert A. 1989. *Democracy and its Critics* (New Haven, CT.: Yale University Press).
- Dahl, Robert A. 1998. *On Democracy* (New Haven, CT.: Yale University Press).
- Fearon, James D. 2002. "Ethnic Structure and Cultural Diversity around the World: A Cross-National Data Set on Ethnic Groups," Department of Political Science, Stanford University, August 11, 2002.
- Fearon, James D. and David D. Laitin. 1996. Explaining Interethnic Cooperation," *American Political Science Review* Vol. 90, N° 4 (December): 715-35.
- Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War," *American Political Science Review* Vol. 97, N° 1 (February): 75-90.
- Jacoby, William G. 1999. "Levels of Measurement and Political Research: An Optimistic View," *American Journal of Political Science* Vol. 43, N° 1 (January): 271-301.
- Kritzer, Herbert M. 1996. "The Data Puzzle: The Nature of Interpretation in Quantitative Research," *American Journal of Political Science* Vol. 40, N° 1 (February): 1-32.
- Lijphart, Arend. 1999. *Patterns of Democracy. Government Forms and Performance in Thirty-Six Countries* (New Haven, CT.: Yale University Press).
- Munck, Gerardo L. and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices," *Comparative Political Studies* Vol. 35, N° 1 (February): 5-34.
- O'Donnell, Guillermo. 2001. "Democracy, Law, and Comparative Politics," *Studies in Comparative International Development* Vol. 36, N° 1 (Spring): 7-36.
- Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and*

- Well-Being in the World, 1950-1990* (New York: Cambridge University Press).
- Rokkan, Stein. 1970. *Citizens, Elections, and Parties: Approaches to the Comparative Study of the Processes of Development* (New York: David McKay).
- Saltelli, Andrea, Karen Chan, and Evelyn Marian Scott (eds.). 2000. *Sensitivity Analysis* (New York: Wiley).
- Sartori, Giovanni. 1987. *The Theory of Democracy Revisited Part 1: The Contemporary Debate* (Chatham, N.J.: Chatham House Publishers).
- Schmitter, Philippe C. and Carsten Q. Schneider. 2003. "The Liberalization of Autocracy and the Consolidation of Democracy. Exploring a Cross-Regional Time-Series Data Set," Paper prepared for delivery at the 2003 Annual Meeting of the American Political Science Association, August 28 - August 31, 2003.
- Smithson, Michael J. 1987. *Fuzzy Set Analysis for Behavioral and Social Sciences* (New York: Springer-Verlag).
- Smithson, Michael J. and Jay Verkuilen. Forthcoming. *Fuzzy Set Theory* (Thousand Oaks, Cal: Sage Publications).
- Torgerson, Warren S. 1958. *Theory and Methods of Scaling* (New York: Wiley).
- Verkuilen, Jay. 2002. *Problems in Comparative and Cross-National Analysis: A Fuzzy Set Approach*, unpublished Ph.D. dissertation, University of Illinois, Urbana, IL.