

**Gerardo L. Munck\* e Jay Verkuilen\*\***

## *Trazendo de volta a mensuração: fundamentos metodológicos do índice de democracia eleitoral<sup>1</sup>*

*Bringing measurement back in: methodological foundations of the electoral democracy index*

A mensuração dos principais conceitos usados no estudo da política tem um grande impacto sobre o nosso conhecimento a respeito do tema. Ela afeta a forma como descrevemos o mundo – e, portanto, as questões que parecem importantes para o estudo – e as proposições causais que consideramos válidas. Ainda mais importante, afeta o tipo de aconselhamento que os cientistas políticos oferecem sobre as principais questões referentes à política e às políticas públicas do momento. No entanto, impressiona a pouca atenção dada à geração de dados e à metodologia de mensuração. Na verdade, pode-se dizer que o ponto de vista dominante na ciência política é o de que a mensuração é uma tarefa necessária, que deve ser rapidamente superada ou, se possível, ignorada completamente, para que as energias dos pesquisadores possam se concentrar em uma tarefa considerada muito mais importante: o teste de hipóteses causais.

---

Do original "Producing and Validating Political Data: The Case of the Electoral Democracy Index". Direitos autorais concedidos pelo autor Gerardo Munck. Tradução de Roberto Cataldo Costa. Revisão técnica da tradução por Luis Felipe Miguel.

\* É professor da Faculdade de Relações Internacionais da Universidade do Sul da Califórnia, em Los Angeles (CA), Estados Unidos. E-mail: [munck@usc.edu](mailto:munck@usc.edu).

\*\* É professor do Departamento de Psicologia da Universidade de Illinois, em Urbana-Champaign (IL), Estados Unidos. E-mail: [jayv@uiuc.edu](mailto:jayv@uiuc.edu).

<sup>1</sup> Artigo preparado para apresentação à Reunião Anual da Associação Norte-Americana de Ciência Política na Filadélfia, de 28 a 31 de agosto de 2003. Direitos reservados à American Political Science Association. Tradução de Roberto Cataldo Costa.

O não reconhecimento da importância da geração de dados e da metodologia de mensuração, bem como da propensão a tomar atalhos, tem custos elevados. Em poucas palavras, essas características estão associadas a ganhos de conhecimento ilusórios, que serão questionados mais cedo ou mais tarde. Assim, é hora de os cientistas políticos desconfiarem mais da tendência a afirmações de conhecimento prematuras, tão comuns na disciplina, e darem mais ênfase à mensuração de conceitos fundamentais como base do conhecimento, isto é, como tarefa que afeta a possibilidade de apresentar uma análise descritiva e/ou causal sólida e, por fim, de oferecer aconselhamento responsável.

Alguns sinais auspiciosos de uma valorização da importância da mensuração podem ser identificados em diversas áreas. Vários projetos de pesquisa importantes e ambiciosos têm abordado a questão da geração de dados como parte integrante do processo de pesquisa<sup>2</sup>. Além disso, algumas publicações importantes em ciência política têm publicado artigos sobre a metodologia da mensuração (Kritzer, 1996; Jacoby, 1999; Adcock e Collier, 2001).

Porém, essas são exceções. A prática padrão na maioria das pesquisas quantitativas é recorrer a conjuntos de dados já disponíveis, os quais já foram tão usados que poucas ideias novas podem ser obtidas a partir deles, além de serem, para começo de conversa, medidas bastante pobres dos conceitos usados na teorização. O problema da literatura qualitativa, embora de natureza diferente, tende a ser igualmente grave. Nesse tipo de pesquisa, a relação entre teoria e observação costuma receber mais consideração. Mas a riqueza das descrições narrativas normalmente é desvinculada de um esforço para codificar sistematicamente os casos. Por sua vez, os esforços para desenvolver a metodologia de mensuração em ciência política ficam consideravelmente aquém dos que acontecem em outras disciplinas. Na verdade, o tipo de atenção à mensuração que é comum em disciplinas como a psicologia e a sociologia e em campos como a educação é bastante estranho à ciência política. Portanto, não

---

2 Entre os exemplos, estão o trabalho de Lijphart (1999) sobre democracia do consenso, que dedica grande atenção, capítulo após capítulo, à mensuração dos conceitos fundamentais que usa; o trabalho de Przeworski *et al.* (2000) sobre democracia e desenvolvimento, que chama à construção de um novo índice sobre a democracia e à compilação de um conjunto de dados de uma série de conceitos políticos e econômicos; o trabalho permanente de Fearon e Laitin (2003) sobre o conflito étnico, um projeto que tem andado de mãos dadas com os esforços para medir o fracionamento étnico (Fearon, 2002); e o trabalho de Schmitter e Schneider (2003) sobre a democracia.

surpreende que os avanços mais significativos na metodologia da mensuração não venham deste campo.

Este artigo procura responder à necessidade de trazer a mensuração de volta à ciência política, e faz isso por meio de uma discussão aprofundada das questões metodológicas envolvidas na mensuração da democracia – um conceito fundamental no campo e, mais amplamente, nas ciências sociais. Em termos mais concretos, o artigo apresenta uma exposição detalhada das escolhas metodológicas feitas na construção de um novo índice de democracia, o índice de democracia eleitoral (IDE)<sup>3</sup>. Como mostra a discussão, a geração de dados é um aspecto extremamente complexo do processo de investigação, que envolve um processo iterativo de teorização e teste, com uma gama de escolhas resultantes, as quais raramente são apreciadas e discutidas com seriedade. Assim, ao explicar e ilustrar as questões envolvidas na geração de índices, ou seja, dados agregados, buscamos incentivar uma discussão mais aprofundada sobre a metodologia de mensuração e promover uma abordagem mais consciente à geração de dados dentro da ciência política.

O trabalho está organizado em termos dos três desafios gerais que devem ser abordados na geração de índices: conceituação, mensuração e agregação (Munck e Verkuilen, 2002)<sup>4</sup>. Em cada seção, discutimos as considerações teóricas e os testes empíricos que informaram as múltiplas escolhas envolvidas na geração de dados. Concluimos com alguns pontos sobre a interpretação e o uso recomendados do IDE.

### **Conceituação: a seleção de itens componentes**

O primeiro passo, e provavelmente o mais importante, para a construção do IDE, que é a escolha dos itens a ser incluídos no índice, levou à identificação de quatro deles, definidos de forma breve no Quadro 1: sufrágio, eleições limpas, eleições livres e cargos eletivos.

3 Esse índice foi preparado para um novo Relatório sobre o Desenvolvimento Democrático da América Latina, uma iniciativa do Escritório Regional para a América Latina e o Caribe do Programa de Desenvolvimento das Nações Unidas (PNUD). Na construção do índice, seguimos a metodologia da mensuração e as recomendações específicas feitas para melhorar os índices de democracia existente descritos em Munck e Verkuilen (2002). Portanto, este artigo é um complemento ao nosso trabalho anterior sobre conceituação e mensuração da democracia.

4 Uma discussão mais aprofundada da matemática usada aqui – com base na teoria dos conjuntos difusos – pode ser encontrada em Smithson (1987), Verkuilen (2002), e Smithson e Verkuilen (2006).

Quadro 1 – Índice de democracia eleitoral (IDE): itens componentes

Índice de democracia eleitoral (IDE)↑			
↑ Sufrágio	↑ Eleições limpas	↑ Eleições livres	↑ Cargos eletivos
Todos os adultos dentro de um determinado país podem votar nas eleições?	O processo de votação é realizado sem irregularidades que restrinjam a possibilidade de os eleitores expressarem suas preferências por candidatos de forma autônoma e precisa?	O eleitorado tem uma gama de opções não restringida, seja por restrições jurídicas, seja por uma questão de força prática?	As eleições são o meio de acesso aos cargos de governo, ou seja, os principais cargos do país (isto é, Executivo e Legislativo nacionais) são preenchidos por meio de eleições, e os vencedores das eleições podem assumir o cargo e cumprir integralmente seus mandatos?

Esses itens foram selecionados por uma série de razões. Em primeiro lugar, correspondem aos elementos centrais que os teóricos da democracia costumam invocar ao definir um regime democrático, além de estarem relacionados com uma série de questões que geralmente são consideradas centrais, na verdade, necessárias, para uma avaliação do caráter democrático de um regime político (Dahl, 1971, 1989, 1998; Rokkan, 1970, cap. 4; O'Donnell, 2001).

Em segundo lugar, todos os quatro itens dizem respeito a direitos de cidadania que o Estado tem responsabilidade de preservar e podem ser claramente interpretados em termos da teoria democrática existente. Assim, evitam os problemas associados a itens como participação em eleições ou a desproporcionalidade eleitoral, que refletem opções do Estado e ações dos cidadãos. Isso ajuda a garantir que o índice possa ser interpretado claramente como uma medida de até onde o Estado garante os direitos de cidadania em relação ao regime político, ao contrário de aspectos da ação cidadã. Da mesma forma, evita os problemas associados a medidas que têm uma implicação incerta para o caráter democrático do regime, como a diferença entre regras eleitorais proporcionais e majoritárias ou entre sistemas presidencialistas e parlamentaristas. Esses outros aspectos certamente não são insignificantes, mas não sua ligação com o caráter democrático do regime não é tão clara quanto a dos quatro escolhidos.

Em terceiro lugar, esses itens foram selecionados porque poderiam ser medidos com base em observáveis a partir de uma metodologia padronizada. Esse critério fez com que alguns componentes que poderiam ter sido incluídos, mas que exigiriam trabalhar com pesquisas de percepções, fossem excluídos. Além disso, o critério significa que certos componentes que poderiam ter sido medidos com base em observáveis não fossem incluídos, devido às restrições práticas impostas pela decisão de não trabalhar com dados já disponíveis baseados em observáveis que, todavia, basearam-se em uma metodologia diferente.

### **Mensuração: a pontuação dos itens**

O segundo passo na construção do IDE, e que demanda mais trabalho – a mensuração dos quatro itens que compõem o índice –, envolveu uma gama de escolhas, começando com a construção de escalas e a definição das regras de codificação, avançando ao próprio processo de codificação, à geração de um conjunto de dados retangular com pontuações normalizadas, o produto final dessa fase do processo de mensuração. Além de explicar e justificar a forma como cada escolha foi feita, são apresentados os resultados de uma análise de sensibilidade voltada a avaliar o quanto a mensuração da função de pertinência foi precisa.

*Escalas e regras de codificação* – O primeiro conjunto de escolhas diz respeito às regras que regem o processo de codificação. As escalas – três escalas ordinais de cinco pontos e uma escala ordinal de três pontos – foram construídas determinando-se, inicialmente, pontos finais teoricamente significativos e depois identificando-se valores de escala distintos, com a maior distância conceitual possível, a partir do ponto central. Ou seja, os valores da escala foram escolhidos de modo a refletir as distinções identificadas na literatura como importantes, evitando as variações verificáveis, mas, ainda assim, pequenas entre casos (Quadro 2). Para os casos que se considerou não se encaixarem precisamente em qualquer um dos pontos nas várias escalas ordinais, introduziu-se a possibilidade de utilização dos sinais de positivo e negativo – análoga à atribuição de notas acadêmicas – como um meio para registrar valores intermediários. No entanto, essa etapa extra de “sintonia fina” foi considerada bastante secundária para a atribuição básica de valores de escala.

## Quadro 2 – Escalas dos itens que compõem o IDE

---

*Sufrágio* – Este componente responde à seguinte pergunta: todos os adultos em um país podem votar nas eleições? Além da realização de eleições, a medida não inclui procedimentos que possam dificultar o uso efetivo do direito de voto, como acesso às urnas.

0 = nenhuma eleição é realizada para instalar o governo.

1 = apenas alguns homens têm direito de votar (há restrições relacionadas com propriedade, gênero e alfabetização).

2 = a maioria dos homens tem direito de votar (há restrições relacionadas com gênero e alfabetização).

3 = a maioria dos homens e a maioria das mulheres têm direito de votar (há restrições relacionadas com a alfabetização).

4 = o direito ao voto é reconhecido universalmente (no entanto, mesmo em países com sufrágio universal, pode haver algumas restrições, afetando grupos como militares, polícia, clero, residentes estrangeiros e cidadãos que moram no exterior).

*Eleições limpas* – Este componente responde à seguinte pergunta: o processo de votação é realizado sem irregularidades que restrinjam a possibilidade de os eleitores expressarem suas preferências por candidatos de forma autônoma e precisa? Não inclui questões relacionadas com a competitividade das eleições, ou mesmo se o vencedor das eleições pode assumir o cargo e cumprir integralmente seu mandato.

0 = graves irregularidades no processo de votação que tenham efeito determinante no resultado das eleições (por exemplo, alteração na eleição para o executivo nacional e/ou o equilíbrio de poder no parlamento).

1 = irregularidades significativas no processo de votação (por exemplo, intimidação e/ou violência contra eleitores).

2 = ausência de irregularidades significativas no processo de votação (por exemplo, eleições que possam incluir irregularidades “técnicas”, mas sem qualquer viés sistemático de importância considerável).

*Eleições livres* – Este componente responde à seguinte pergunta: o eleitorado tem uma gama de opções não restringida, seja por restrições jurídicas, seja por uma questão de força prática? A medida não inclui fatores que afetem a capacidade de partidos e candidatos de competir em igualdade de condições, tais como financiamento público, acesso aos meios de comunicação de massas e o uso dos recursos públicos.

0 = sistema de partido único.

1 = proibição de partido importante.

2 = proibição de partido menos importante.

3 = restrições de natureza legal ou prática que afetem significativamente a capacidade dos potenciais candidatos de concorrer a um cargo e/ou a formação de partidos políticos (por exemplo, assassinatos e intimidação sistemáticos de candidatos, proibição de candidatos populares, restrições legais ou práticas que impeçam a formação de partidos ou que levem partidos a boicotarem as eleições).

4 = condições essencialmente irrestritas à apresentação de candidatos e à formação de partidos.

---

## Quadro 2 – Escalas dos itens que compõem o IDE (continuação)

---

*Cargos eletivos* – Este componente responde à seguinte pergunta: as eleições são o meio de acesso a cargos de governo, ou seja, os principais cargos políticos do país (isto é, Executivo e Legislativo nacionais) são preenchidos por meio de eleições e os vencedores das eleições podem assumir o cargo e cumprir integralmente seus mandatos?

0 = nenhum dos principais cargos políticos é preenchido por meio de eleições ou todos os principais ocupantes de cargos políticos são destituídos à força do cargo e substituídos por governantes inconstitucionais.

1 = alguns cargos políticos são preenchidos por vencedores das eleições ou a maior parte dos principais ocupantes de cargos políticos é destituída à força do cargo e substituída por governantes inconstitucionais.

2 = o presidente ou o parlamento não são eleitos ou são destituídos à força do cargo e substituídos por governantes inconstitucionais.

3 = o presidente ou o parlamento são eleitos, mas o presidente é destituído do cargo e/ou substituído por meios semiconstitucionais, ou um número significativo de parlamentares não é eleito ou é destituído à força do cargo.

4 = todos os principais cargos políticos são preenchidos por meio de eleições e nenhum dos principais ocupantes de cargos políticos é destituído do cargo, a menos que sua remoção do poder e sua substituição tenham bases estritamente constitucionais.

---

As escalas foram construídas de maneira que cada ponto correspondesse a situações e eventos relativamente concretos e que as decisões de codificação pudessem ser tomadas a partir de padrões bem definidos e estritamente com base em observáveis. Dados baseados em pesquisas sobre percepções não foram incluídos no índice. Além disso, como forma de garantir ainda mais a replicabilidade do exercício de codificação, bem como protegê-la da arbitrariedade, foi salientada a importância de documentar a base das decisões de codificação por meio de referência a fontes públicas de informação.

Não foram necessárias pontuações para cada caso anualmente. Em vez disso, a pontuação para três dos componentes – sufrágio, eleições limpas e eleições livres – foi forçosa apenas para os anos em que foram realizadas eleições. As condições para as eleições são afetadas por eventos e decisões tomadas entre eleições, e a codificação se baseou em informações entre períodos eleitorais. Mas o significado desses eventos e decisões para o processo pelo qual os atores têm acesso a cargos governamentais – preocupação central do exercício de mensuração – cristaliza-se no próprio evento eleitoral. Assim, embora só se tenham atribuído pontuações a alguns componentes durante anos de eleição, essas pontuações foram consideradas síntese de um processo mais amplo.

*O processo de codificação* – O segundo conjunto de escolhas estava relacionado com o próprio processo de codificação. Nesse sentido, foram usados dois processos complementares para codificar os casos. Uma codificação inicial foi realizada por um único codificador, com base em uma ampla pesquisa e em consultas a um grande número de especialistas, durante um período de vários meses. Além disso, as pontuações atribuídas foram apresentadas e discutidas profundamente em vários encontros, um deles com um grupo de participantes convidados que trabalham em diversos contextos (política, academia, organizações internacionais) e vêm de diferentes países das Américas (Argentina, Brasil, Canadá, Colômbia, Equador, México, Estados Unidos e Uruguai). Essas discussões levaram à identificação de divergências, o que gerou novas pesquisas e mais discussões em grupo. No final, com esse processo iterativo, houve um elevado grau de consenso com relação à codificação das quatro dimensões do IDE.

*A geração de um conjunto de dados retangular com escalas normalizadas* – O terceiro conjunto de escolhas diz respeito à transformação das pontuações das escalas de componentes em um conjunto de dados retangular, ou seja, um conjunto de dados que inclui pontuações numéricas para todos os casos em todas as variáveis e em todos os anos, com escalas normalizadas.

Foram tomadas várias medidas para fornecer números em todos os anos. Em primeiro lugar, abordou-se uma série de questões bastante mecânicas. Sinais de positivo e negativo foram transformados em números, adicionando-se ou se subtraindo 0,33 da pontuação-base (por exemplo, um 3+ foi transformado em 3,33). Traços (–) usados para indicar que a atribuição de uma pontuação não se aplicava porque o governo não era eleito foram transformados em zeros (0). Além disso, as pontuações de dois dos componentes a que foram atribuídas pontuações apenas para o ano em que houve eleição – sufrágio e eleições livres – foram estendidas aos anos intermediários simplesmente transportando-se a pontuação de um determinado ano aos anos seguintes, até que uma nova pontuação fosse atribuída (seja porque houve eleição depois de um período em que havia um governo não eleito, seja porque houve uma nova eleição ou, ainda, porque o processo eleitoral foi interrompido). Esse procedimento justifica-se pela maneira como um governo se origina continua a afetar sua natureza para além do momento de sua instalação. No caso do item “eleições limpas”, seguiu-se um procedimento um pouco mais complexo, com o uso de uma escala de três pontos em sua

codificação, por razões de interpretabilidade. No entanto, o 1 na escala não representa realmente um ponto médio, sendo muito mais próximo do 2. Assim, cada 1 foi transformado em um 3, e cada 2, em um 4. Além disso, como esse item distingue os valores atribuídos a eleições para presidente e parlamento, as pontuações não são simplesmente transportadas de eleição para eleição. Em vez disso, são uma média das pontuações relativas a eleições para presidência e parlamento.

Uma segunda questão foi o problema acarretado pela atribuição de apenas uma pontuação por país a cada ano. Essa prática, que é padrão, obedece a razões de parcimônia e é bem justificada. Afinal de contas, a razão de se gerar um índice é apresentar uma síntese do estado de um país. Mas ela enfrenta uma série de problemas porque a situação dos países muda ao longo de um ano e se usa apenas uma pontuação para caracterizar o ano todo. Em alguns casos, a solução é relativamente fácil. Assim, quando houve um evento fundamental no final do ano, como uma eleição, a mudança de *status* em consequência desse evento foi registrada no ano seguinte<sup>5</sup>. Quando houve eventos no primeiro semestre do ano, eles foram computados no mesmo ano<sup>6</sup>. Em outros casos, no entanto, a solução foi mais difícil. Em algumas situações, mesmo que um evento tivesse acontecido na segunda metade do ano, ele foi registrado neste ano<sup>7</sup>. Também foram problemáticos os casos em que mais de um evento importante acontece em um ano<sup>8</sup>.

Em terceiro lugar, as escalas componentes foram normalizadas, isto é, traduzidas para uma métrica comum, através de uma normalização linear simples em relação ao intervalo unitário:

$$\text{valor normalizado} = \text{valor bruto} / \text{valor bruto máximo}$$

5 Por exemplo, apesar de terem sido realizadas as eleições que puseram fim a um período de governos dominados por militares na Guatemala, em 1985, essas eleições aconteceram no final de 1985 e só levaram à mudança de governo em janeiro de 1986. Assim, mesmo quando se atribuíram pontuações aos itens componentes para 1985, ao calcular o IDE, essas pontuações foram inseridas com relação a 1986.

6 Por exemplo, a eleição de 1994 em El Salvador foi realizada em março e levou a uma mudança de governo em junho. Assim, a mudança foi registrada em 1994.

7 É o caso das eleições fraudulentas na República Dominicana, realizadas em maio, levando à posse do novo presidente em agosto.

8 Por exemplo, em 2000, o Peru realizou duas eleições muito questionadas, em abril e maio, levando à posse de Fujimori como presidente, em julho, e, em seguida, à sua demissão, em novembro. Nesse caso, as eleições problemáticas foram registradas em 2000 e a retificação da situação foi registrada em 2001.

Infelizmente, quase qualquer escolha seria um tanto arbitrária, porque não existem unidades amplamente aceitáveis de liberdade eleitoral comparáveis a unidades como quilogramas ou dólares. Mas a escolha do procedimento de normalização, como se aplica à escala ordinal de cinco pontos (com a modificação introduzida no item “eleições limpas” quando se transformaram as pontuações nas escalas componentes em um conjunto retangular de dados, as escalas usadas para medir todos os quatro itens componentes eram escalas ordinais de cinco pontos) é transparente e justificável.

É provável que uma grande objeção seja a de que as escalas não são escalas de razão em si, mas escalas ordinais de cinco pontos, com um intervalo estabelecido por critério do investigador. Assim, uma objeção convincente à normalização para o intervalo unitário é a de que ele usa uma matemática “insignificante”, inadmissível. Ao abordar essa questão, Warren Torgerson afirma:

Certamente não há nada de errado nem de logicamente incorreto com o [uso do critério do investigador] [...]. Em todos esses casos, seleciona-se uma ou mais propriedades observáveis que se consideram, aprioristicamente, relacionadas ao conceito de interesse. Uma medida da propriedade observável em si, ou de uma soma simples ou ponderada de várias dessas propriedades observáveis, é tomada como a medida do conceito de interesse. Normalmente, os valores assim obtidos são tratados como se fossem medidas em um intervalo ou escala de razão. Isso fica claro, por exemplo, quando as curvas são ajustadas pelo processo dos quadrados mínimos ou quando são calculadas correlações produto-momento, médias ou desvios-padrão. Tudo isso pressupõe que a distância seja significativa. Assim, explícita ou implicitamente, o experimentador está medindo o atributo em uma escala de intervalo cujas características de ordem e distância se tornaram significativas inicialmente apenas através de definição.

A descoberta de relações estáveis entre variáveis medidas dessa forma pode ser tão importante como entre as variáveis medidas de outras maneiras. Na verdade, realmente faz pouca diferença se a escala atual de comprimento, por exemplo, foi obtida originalmente através de definição arbitrária, através de uma relação com outras variáveis estabelecidas ou através de um processo fundamental. O conceito é bom, e estabeleceu um número imenso de relações simples com outras variáveis. E esse é, apesar de tudo, o principal critério do valor de um conceito. Se o resultado é o mesmo,

o modo como um conceito é originalmente introduzido tem importância limitada [...] [Mas] a grande dificuldade com a mensuração por critério do investigador é o imenso número de maneiras em que essas escalas definidas podem ser construídas (Torgerson, 1958, p. 23-4).

Assim, a pergunta é como a nossa escolha pode ser justificada. E, para essa pergunta, pode-se oferecer a resposta a seguir. Em primeiro lugar, todas as escalas têm pontos finais teoricamente significativos e se pode supor que caíam no intervalo unitário, com 0 indicando uma ausência total de propriedade e 1 indicando posse total da propriedade. Ou seja, o valor mais baixo da escala ordinal corresponde à negação da propriedade em questão, enquanto o valor máximo corresponde à posse completa. Um caso com sufrágio 0 não tem qualquer sufrágio, enquanto um caso com sufrágio 1 após normalização tem sufrágio adulto total – o padrão estabelecido teoricamente. Dessa forma, o problema da distância se refere apenas aos pontos entre os pontos finais. Em segundo lugar, a maior parte das escalas foi construída de tal modo que cada ponto na escala seja teoricamente interpretável e que os distintos valores da escala tenham a maior distância conceitual possível entre si. Isto é, os valores da escala foram escolhidos de modo a refletir distinções identificadas na literatura como sendo importantes, evitando variações verificáveis, mas pequenas, entre os casos. Portanto, a probabilidade de se introduzir qualquer erro grande é relativamente pequena. Embora se pudessem usar outros métodos psicométricos mais sofisticados, eles são mais complicados e menos acessíveis, muito dependentes de dados e, muitas vezes, não têm um desempenho muito melhor do que este procedimento simples. Nesse sentido, cabe enfatizar que, apesar de importante, a otimização estatística não é o único critério relevante.

*Confiabilidade intercodificadores e estimativas de erro* – Devido a razões de tempo, não foi realizado teste formal de confiabilidade intercodificadores. No entanto, para se ter uma ideia de se outros codificadores poderiam ter dado atribuições um pouco diferentes dos itens componentes do IDE, a acuidade da mensuração da função de pertinência foi avaliada através de uma análise de sensibilidade (Saltelli *et al.*, 2000). Essa análise se baseia em perturbações dos códigos atribuídos de acordo com um desenho experimental e no exame dos índices globais “replicados” que resultam daí. Portanto, ela usa a matemática para criar codificadores “virtuais” que têm

vários vieses – por exemplo, um viés para baixo na classificação de um ou mais componentes. Os resultados desse teste mostraram que o IDE é bastante estável – as correlações de ordem com todas as outras “repetições” foram de 0,99 ou maiores, e as mudanças na média e na distribuição foram bastante previsíveis, mostrando um viés negativo ou positivo previsível, dependendo da célula no desenho experimental. Esse teste também forneceu algumas barras de erro básicas para o IDE com base nas “replicações”. Faixas generosas para valores de índice em torno de 0,25 a 0,75 são de cerca de  $\pm 0,07$  e faixas razoavelmente conservadoras são de  $\pm 0,1$ . Pelo padrão mais conservador possível, os valores de IDE estão dentro de  $\pm 0,2$ . A amplitude é razoavelmente constante ao longo do intervalo citado, mas os limites exatos dependem do valor do índice e, próximo aos pontos finais, costumam ser mais estreitos. Realizou-se uma verificação da matemática usando a inversão do conhecido e altamente conservador teste de Kolmogorov-Smirnov para a função de distribuição – com base em matemática totalmente diferente – e foram obtidos resultados semelhantes.

### **Agregação: a redução dos dados**

O terceiro passo na construção do IDE – a agregação dos quatro itens componentes – envolveu uma escolha no que diz respeito à regra de agregação. Além disso, vários testes foram realizados para avaliar a robustez da regra de agregação proposta, a dimensionalidade dos itens componentes, bem como a influência dos itens componentes do IDE. Dessa forma, depois de se explicar e oferecer uma justificativa inicial para a escolha da regra de agregação, esses testes são apresentados e seus resultados, interpretados.

*A escolha de regras de agregação* – A escolha da regra de agregação, que formaliza a relação entre os itens componentes do índice, foi resolvida com o uso de uma regra simples. A ideia central usada na escolha de uma regra de agregação é a visão bem estabelecida de que os quatro itens componentes do IDE são partes que constituem um sistema em virtude da maneira com que se combinam e, além disso, que esses quatro itens componentes são tão fundamentais para a caracterização geral do regime que a ausência de qualquer um deles simplesmente tornaria o regime não democrático (Sartori, 1987, p. 184-5; Przeworski, Alvarez, Cheibub e Limongi, 2000, p. 28-9 e 57; O'Donnell, 2001, p. 10 e 12). Por exemplo, como há muito argumentam os teóricos da democracia, o fato de que os sistemas de tipo soviético tinham

eleições com sufrágio completo era insignificante do ponto de vista da democracia, porque o eleitorado não tinha uma escolha entre diferentes candidatos e porque aquelas eleições não levavam à ocupação de cargos que efetivamente exercessem o poder de Estado. Por conseguinte, os quatro itens componentes do IDE são postulados como condições individualmente necessárias, que são insubstituíveis e têm peso igual<sup>9</sup>. Esta concepção é formalizada calculando-se o produto do valor de cada um dos itens componentes. Em termos formais, portanto, o IDE é calculado de acordo com a seguinte equação:

$$\text{IDE} = \text{sufrágio} \times \text{eleições limpas} \times \text{eleições livres} \times \text{cargos eletivos}$$

Essa equação captura uma visão fundamental da teoria democrática: a de que quando algum item componente inexistente totalmente, o regime deve simplesmente ser considerado como não democrático. Com efeito, essa operação garante que um valor de zero em qualquer um dos quatro itens componentes faz com que se classifique um caso como “não democracia”. Este é um padrão “duro”, o que pode ser visto como menos “indulgente” do que outras regras de agregação. Assim, cabe enfatizar que, precisamente por ser muito exigente, essa concepção dos itens componentes do IDE como condições individualmente necessárias é usada em conjunto com um critério conservador na atribuição de zeros ao item, indicando a ausência total de uma determinada propriedade. Isso é válido tanto porque as escalas foram construídas de tal forma que só se atribuirá zero nos casos extremos, em que uma propriedade considerada vital para a existência da democracia estiver totalmente ausente, quanto porque a evidência necessária para se atribuir zero teve de ser convincente. Assim, apenas quando as normas democráticas são flagrante e indiscutivelmente desconsideradas é que o IDE codifica um país como não democrático.

*A robustez de regras de agregação* – Foi realizado um teste comparando quatro possíveis regras de agregação para combinar os itens componentes do IDE: o produto dos quatro componentes efetivamente usados no IDE, o valor mínimo dos quatro componentes na escala, a média geométrica dos quatro componentes, bem como sua média aritmética. Os resultados mostraram que, não importa qual regra se use, as correlações de ordem são todas muito altas, indicando que o ordenamento geral dos casos é preservado. Mas há

9 Se esses itens podem ser considerados suficientes em seu conjunto ou não é outra questão.

diferenças entre os índices, com a média aritmética e a média geométrica sendo semelhantes entre si, e o mínimo e o IDE sendo semelhantes entre si. A diferença mais importante se dá entre as médias e os desvios-padrão (DPs). As médias aritmética e geométrica têm médias de 0,92 e 0,91 e DPs de 0,20 e 0,21, respectivamente. Em contraste, o mínimo e o IDE têm médias de 0,84 e 0,82 e DPs de 0,26 e 0,28, respectivamente. Isso sugere que essas últimas regras funcionam melhor para distribuir os casos e evitar uma acumulação que torne difícil interpretar suas diferenças com muita clareza. De qualquer forma, muitos casos têm atribuído a si um valor de um, o que sugere que a escala de IDE está “chegando ao máximo” no período de tempo atual.

*A dimensionalidade dos itens componentes* – Um teste de escalabilidade dos quatro itens componentes do IDE resultou em um alfa de Cronbach de 0,92, sugerindo que o IDE é a medida de um fenômeno unidimensional. No entanto, ao se realizar um teste em dois períodos (1960-1985 e 1990-2002), os alfas de Cronbach resultantes foram de 0,95 e 0,23, respectivamente, indicando que, embora no período anterior os componentes fossem unidimensionais, isso já não se aplica ao período pós-1990. Essa conclusão é coerente com a teoria usada na seleção das regras de agregação para o IDE. Também tem sentido em relação aos históricos de caso, já que, no período anterior, os regimes autoritários geralmente eram completos em sua supressão da democracia. Com efeito, é importante notar que os modelos de mensuração aditivos padrão trabalham com a premissa de que a agregação opera em múltiplas medições paralelas. Por outro lado, como os itens componentes do IDE são teoricamente considerados insubstituíveis, a decisão de agregar em uma única pontuação não é invalidada por quaisquer desvios potenciais em relação à unidimensionalidade. O IDE, nesse sentido, é muito mais parecido com a função de produção em economia do que o teste de capacidade de psicometria, para o qual o procedimento de Cronbach foi elaborado. Os fatores de produção para a democracia podem avançar em conjunto ou não, mas a falta de qualquer um deles torna seu produto indisponível. Assim, esse teste de escalabilidade oferece mais validação para a escolha das regras de agregação propostas em oposição à regra aditiva, bastante comum.

*Influência dos itens componentes* – Um teste indireto da influência dos itens componentes dá mais sustentação à visão de que os componentes do IDE não variam da mesma forma durante todo o período coberto. Os resultados desse teste são apresentados na Tabela 1, que mostra as médias

(desvios-padrão entre parêntesis) para os componentes normalizados, e o IDE, condicionalmente ou não ao período de tempo (<1990 ou não). Antes de 1990, todos os componentes do IDE variam. Isso é particularmente verdadeiro por causa da presença de casos de autoritarismo puro. O item “eleições livres” tem a menor média, indicando que é o componente violado com mais frequência, mas todos os componentes estão em baixa. Por outro lado, depois de 1990, o item “sufrágio” não tem variância e desaparece completamente. E os itens “eleições limpas” e “eleições livres” variam muito, acarretando, assim, a maior parte da variação no IDE. Isso confirma a impressão que se teria ao examinar as tabelas de dados sobre os componentes. Mas um ponto importante também é que a maioria dos casos está no topo da classificação.

Tabela 1 – Meios de componentes e desvios-padrão

Amostra	Sufrágio	Eleições limpas	Eleições livres	Cargos eletivos	IDE
Incondicional (n = 288)	0,96 (0,19)	0,90 (0,22)	0,89 (0,25)	0,94 (0,20)	0,82 (0,27)
<1990 (n = 54)	0,76 (0,40)	0,73 (0,39)	0,60 (0,43)	0,78 (0,40)	0,52 (0,40)
≥ 1990 (n = 234)	1,00 (0,00)	0,94 (0,13)	0,96 (0,09)	0,98 (0,08)	0,89 (0,17)

### Conclusão: interpretando e usando o IDE

O IDE é uma escala de 0,00 a 1,00, com 0,00 indicando não democracia, qualquer número acima de 0,00 indicando um grau de democracia eleitoral e pontuações maiores se referindo a graus mais elevados de democracia eleitoral. Para evitar confusão, é importante observar que o índice não deve ser interpretado como uma avaliação das ações de um governo. Em vez disso, é uma medida do estado de um sistema que é afetado pela ação e a inação de um governo, bem como de outros agentes do Estado e atores da sociedade. Além disso, cabe salientar que o conceito medido é a democracia eleitoral. Esse conceito não é tão estreito quanto alguns o consideram. Portanto, embora seja direcionado diretamente à realização de eleições incluídas, limpas e livres, ele abrange mais do que “meras eleições”. Não apenas as condições para a realização dessas eleições são afetadas pelos acontecimentos entre eleições; também é necessária uma reflexão sobre o que acontece com os

próprios governos entre eleições. Mas o IDE certamente não é uma medida ampla da democracia, e sim a medida de uma concepção de regime político democrático, baseada no entendimento mais amplamente compartilhado sobre direitos políticos fundamentais. E isso é muito significativo. Por um lado, significa que quaisquer deficiências detectadas pelo IDE devem ser consideradas importantes restrições aos direitos políticos do cidadão. Por outro lado, o fato de um país ter recebido uma pontuação perfeita, de 1,00, não deve ser interpretado como se não houvesse ali espaço para aprimoramento, no que diz respeito a dimensões não incluídas no índice ou a padrões relativamente mais exigentes dos itens que compõem o IDE.

O IDE pode ser usado para fins de comparação, seja de um país consigo mesmo ou com outros países. Desses dois usos, uma comparação de um país consigo mesmo em um ponto diferente no tempo geralmente é mais fácil de interpretar. Afinal, o país pode muito bem ter feito melhorias importantes, mas, ainda assim, estar atrás de outros se estes fizeram avanços ainda maiores. No entanto, em ambos os casos é fundamental ressaltar que qualquer comparação desse tipo deve ser sempre baseada em diferenças substanciais, e nunca em diferenças pequenas. A razão para isso é que o IDE, como qualquer índice, tem um certo grau de erro de mensuração e, dentro dos limites de erro, não é aconselhável fazer declarações contundentes sobre diferenças. De fato, como foi estimado por meio da análise de sensibilidade, barras de erro generosas para valores de IDE entre cerca de 0,25 e 0,75 são de cerca de  $\pm 0,07$ . Assim, quaisquer dois casos que difiram por menos do que esse valor – por exemplo, um país com um IDE de 0,85 e um de 0,95 – simplesmente estão próximos demais para serem diferenciados de forma válida. Por isso, é metodologicamente injustificável oferecer uma classificação excessivamente precisa de países, como geralmente se faz no contexto de outros índices que simplesmente transformam as pontuações do IDE em uma classificação, sem levar em consideração o grau de incerteza associado a essas pontuações.

A identificação dos casos de referência que sejam representações prototípicas das características associadas a uma série de pontuações pode ajudar a dar concretude ao significado de cada número. Além disso, o IDE pode ser utilizado como um indicador, no sentido de que pontuações específicas de cada país convidam o leitor a voltar aos quadros sobre os itens componentes para identificar precisamente qual característica ou características explicam a pontuação do país. Dessa forma, o IDE pode ser usado como uma valiosa

ferramenta analítica, na medida em que oferece uma pontuação concisa que permite a seus usuários identificarem a peculiaridade do regime político de cada país com base não só em seus diversos itens componentes, mas também na relação entre as partes constituintes do regime e sua contribuição para o todo.

## Referências

- ADCOCK, Robert N. & COLLIER, David (2001). "Measurement validity: a shared standard for qualitative and quantitative research". *American Political Science Review*, v. 95, n. 3, p. 529-46.
- DAHL, Robert A. (1971). *Polyarchy*. New Haven (CT): Yale University Press.
- \_\_\_\_\_. (1989). *Democracy and its critics*. New Haven (CT): Yale University Press.
- \_\_\_\_\_. (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". *Annual Meeting of the American Political Science Association*. Boston, 29 ago.-1 set.
- FEARON, James D. & LAITIN, David D. (1996). "Explaining interethnic cooperation". *American Political Science Review*, v. 90, n. 4, p. 715-35.
- \_\_\_\_\_. (2003). "Ethnicity, insurgency, and civil war". *American Political Science Review* v. 97, n. 1, 75-90.
- JACOBY, William G. (1999). "Levels of measurement and political research: an optimistic view". *American Journal of Political Science*, v. 43, n. 1, p. 271-301.
- KRITZER, Herbert M. (1996). "The data puzzle: the nature of interpretation in quantitative research". *American Journal of Political Science* v. 40, n. 1, p. 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. & VERKUILEN, Jay (2002). "Conceptualizing and measuring democracy: evaluating alternative indices". *Comparative Political Studies*, v. 35, n. 1, p. 5-34.
- O'DONNELL, Guillermo (2001). "Democracy, law, and comparative politics". *Studies in Comparative International Development*, v. 36, n. 1, p. 7-36.
- PRZEWORSKI, Adam; ALVAREZ, Michael E.; CHEIBUB José Antonio & LIMONGI, Fernando (2000). *Democracy and development: political in-*

- stitutions 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; CHAN, Karen & SCOTT, Evelyn Marian (orgs.) (2000). *Sensitivity analysis*. New York: Wiley.
- SARTORI, Giovanni (1987). *The theory of democracy revisited. Part 1: The contemporary debate*. Chatham (NJ): Chatham House Publishers.
- SCHMITTER, Philippe C. & SCHNEIDER, Carsten Q. (2003). "The liberalization of autocracy and the consolidation of democracy. exploring a cross-regional time-series data set". *Annual Meeting of the American Political Science Association*. Philadelphia, 28-31 ago.
- SMITHSON, Michael J. (1987). *Fuzzy set analysis for behavioral and social sciences*. New York: Springer.
- SMITHSON, Michael J. & VERKUILEN, Jay (2006). *Fuzzy set theory*. Thousand Oaks (CA): Sage.
- 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*. Tese (doutorado) em ciência política. Urbana (IL): University of Illinois.

## Resumo

A medição dos principais conceitos utilizados no estudo da política tem um grande impacto no nosso conhecimento sobre ela. Afeta a nossa forma de descrever o mundo (e, portanto, as questões que parecem importantes para o estudo) e as proposições causais que consideramos válidas. Ainda mais importante, afeta o tipo de conselho que cientistas políticos podem oferecer sobre os principais temas de política pública e as questões políticas do dia. No entanto, é notável a pouca atenção é dada para a geração de dados e a metodologia de medição. Na verdade, é justo dizer que a visão dominante na ciência política é que a medição é uma tarefa necessária que deve ser rapidamente transcendida ou, se possível completamente ignorada, de modo que as energias dos pesquisadores se concentrem em uma tarefa vista como muito mais importante: o teste de hipóteses causais. A falha em reconhecer a importância da geração de dados e a metodologia da medição, e a propensão a tomar atalhos, tem custos graves. Em poucas palavras, ela leva a ganhos ilusórios em conhecimento, que mais cedo ou mais tarde são questionados. Assim, é hora de que os cientistas políticos desconfiem da tendência de reivindicações

prematuros de conhecimento, tão difundida na disciplina, e coloquem mais ênfase na medição de conceitos-chave como uma base de conhecimento, isto é, como uma tarefa que afeta a possibilidade de fornecer análise descritiva e/ou causal sólida e, em última análise, de oferecer conselhos responsáveis.

**Palavras-chave:** metodologia, índice de democracia

## **Abstract**

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.

**Keywords:** methodology, democracy index