

Estratégias rivais de validação: ferramentas para avaliar medidas de democracia

Rival strategies of validation: tools for evaluating measures of democracy

Jason Seawright
David Collier

Resumo

O desafio de encontrar as ferramentas adequadas para validar o processo de mensurar, tem sido objeto de preocupação permanente, na Ciência Política. Este artigo procura analisar quatro diferentes modelos de validação, e trás exemplos da pesquisa transnacional sobre democracia: a abordagem os níveis-de-medição, o modelo de equação estrutural com variáveis ocultas (latentes), a tradição pragmática e o método baseado em estudo de casos. Os metodologistas tem disputado de forma bastante contundente os méritos de modelos alternativos. Nós incentivamos os pesquisadores- e também os analistas da democracia- a prestar mais atenção nessas disputas e com isso reavaliar os pontos fortes e fracos das ferramentas de validação que eles tem usado. Um apêndice online trás um resumo da avaliação de seis grupos de dados sobre democracia sob a perspectiva de abordagens alternativas de validação. O objetivo geral é abrir um novo espaço de discussão sobre estratégias alternativas de validação.

Palavras-chave

Democracia; Multi Método; Pesquisa Transnacional; Modelo de Equação Estrutural.

Abstract

The challenge of finding appropriate tools for measurement validation is an abiding concern in political science. This article considers four traditions of validation, using examples from cross-national research on democracy: the levels-of-measurement approach, structural-equation modeling with latent variables, the pragmatic tradition, and the case-based method. Methodologists have sharply disputed the merits of alternative traditions. We encourage scholars – and certainly analysts of democracy – to pay more attention to these disputes and to consider strengths and weaknesses in the validation tools they adopt. An online appendix summarizes the evaluation of six democracy data sets from the perspective of alternative approaches to validation. The overall goal is to open a new discussion of alternative validation strategies.

Keywords

Democracy; Multimethod; Cross-national Research; Structural-equation Modeling.

Introdução

Pesquisadores, normalmente, enfrentam dificuldades quando têm que escolher uma, entre as ferramentas alternativas de mensuração de validade em pesquisa comparativa. Alguns autores usam a proximidade de correlação com outras medidas transnacionais, para defender seus indicadores, baseando seus argumentos na já bem estabelecida ideia de validação convergente. Outros aprofundam esta abordagem utilizando complexos modelos estatísticos para a construção de indicadores e avaliar erros. Em alguns casos, a principal preocupação está relacionada com os níveis de medição, e também tem aqueles que procuram avaliar e melhorar a qualidade dos seus indicadores focando a sua atenção em evidências presentes em um, ou alguns casos analisados.

Este artigo procura explorar quatro diferentes tipos de abordagem alternativa de medição, usando exemplos tirados da pesquisa transnacional sobre democracia. Esta, certamente, se constitui uma tarefa bastante desafiadora, porque assim como a literatura sobre inferência causal, os trabalhos sobre o processo de validação de medição têm causado muita controvérsia. Certamente, pesquisadores inseridos em qualquer um dos campos, em certos casos, são extremamente hostis a abordagens diferentes. Antigas e mais recentes críticas têm rejeitado certas abordagens classificando-as como “má análise de dados e má ciência” e como “enganosa” (VELLEMAN e WILKINSON, 1993, p. 70 e 72), como “excessivamente senso comum” (FREEDMAN, 1987, p. 102), como um “desastre” (CLIFF, 1983, p. 116), como, “ciência patológica” (MICHELL, 2008, p. 10), como “ofuscante” (DUNCAN, 1984, p. 135), como “obstáculos” para o progresso das Ciências Sociais refletindo “preguiça conceitual” (BLALOCK, 1982, p. 109 e 110), e “impeditivos” ao progresso científico (YOUNG, 1981, p. 357). Tukey (1986) defende uma abordagem com mais “sensatez” e menos “puritanismo”. Com esse espírito, a sua lista de maus conselhos metodológicos inclui o sarcástico “não pense, use estatística” (TUKEY, 1986, p. 202, 243, 244 e 247)¹.

Rotineiramente, pesquisadores tem usado ferramentas de avaliação de medidas desses quatro modelos, sem o devido escrutínio e reflexão sobre a sua acuracidade. Com o objetivo de revisar os pontos fortes e fracos de cada abordagem,

¹ Velleman e Wilkinson (1993) comentam sobre o modelo níveis de medicao, e Freedman (1987). Cliff (1983), e Michell (2008) criticam o modelo de equações estruturais. Tukey (1986) defende o modelo pragmático e tem como preocupação central o que ele vê como desnecessárias e indesejáveis restrições impostas pelo modelo níveis de medição. Blalock (1982), Duncan (1984) e Young (1981) rejeita legitimidade de trabalhos com dados nominais, os quais certamente incluiriam uma boa parte do que é feito no modelo baseado em estudo de casos.

nós estamos encorajando escolhas mais criteriosas e informadas sobre a validação de medidas. A tabela 1 apresenta uma visão geral dos quatro modelos, assim como, importantes exemplos trazidos da literatura sobre democracia. Como em qualquer classificação, algumas sobreposições vão ser encontradas, porém mesmo assim, a classificação continua sendo útil para a distinção, dos métodos alternativos.

O modelo *Níveis-de-medição* (LoM, sigla em Inglês) concentra a sua atenção na clássica distinção entre tipos de escala -tais como nominal, ordinal, intervalo e razão- e tenta fortalecer os seus indicadores através da transformação, sem distorção, das informações contidas em cada tipo de escala.

O modelo *Equação estrutural com variáveis ocultas* (latentes) concentra-se na elaboração de modelos estatísticos que agregam indicadores de forma aditiva, mas geralmente com pesos diferenciados, com o objetivo de medir um “valor” do conceito oculto (latente) para cada caso – neste estágio de democracia. Desta forma procura melhorar a validade da mensuração, através da eliminação de erros na criação das novas variáveis.

Seawright e Collier

Tabela 1 – Quatro Modelos de Validação de Medidas

	Níveis de Medição	Equação Estrutural com Variáveis Ocultas (latentes)	Pragmático	Estudo de Casos
Principal contribuição	Trata os níveis de medição como uma visão empírica básica sobre os indicadores e como um guia para formas mais adequadas de análise de dados.	Usa múltiplos indicadores e suposições sobre as relações descritivas e causais na tentativa de reduzir os erros de medição.	Focaliza sua atenção na aplicação dos indicadores; secundariamente preocupa-se com os níveis de medição e outras características dos dados analisados.	Avalia as medidas com base na análise profunda de um ou alguns casos. Tenta descobrir se os dados são plausíveis a luz do detalhamento do caso.
Ferramentas usadas	Escala Guttman, modelo Rasch, teoria item-response.	Equação estrutural com variáveis ocultas, análise fatorial, teoria item-response.	Correlação entre indicadores sem nenhum modelo específico explícito, validação monológica, regressão ALSOS, Em alguns casos usa a escala	Testes de correspondência entre a classificação de um caso usando um indicador, e o conhecimento

			multidimensional, testes de confiabilidade entre múltiplos.	histórico contextual.
Exemplos	Coppedge e Reinicke (1990) e P. J. Baker e Koesel (2001) usam a escala nominal para criar um ranking de ordenamento de regimes	Bollen (1993) e Bollen e Paxton (2000) fazem uma estimativa dos erros e do grau de viés político.	Elkins (2000) avalia indicadores alternativos testando-os em comparação com hipóteses já estabelecidas.	Bowman, Lehoucq, e Mahoney (2005) usam o conhecimento de casos na América Central para reavaliar indicadores transnacionais.
Crítica de outros Metodologistas	Níveis de medição, as vezes, podem não ser importantes para estatística e inferência causal.	Complexidade e a não-testabilidade de pressupostos no modelo de medição.	Desatenção a links entre indicadores e os conceitos que se quer medir.	O foco em características específicas dos casos estudados pode obscurecer uma verdadeira comparação entre este grupo de casos e uma gama maior deles.

Fonte: Elaboração própria.

No modelo *pragmático* a mensuração se torna valida e apropriada quando ela se mostra útil para um proposito específico, ou um dado contexto. Esta abordagem rejeita especialmente os pressupostos condicionantes que sustentam SEM-L e LoM, considerando que as distinções entre LoM podem ser de limitada relevância em alguns contextos. Testes de validade que usam correlações simples de duas variáveis, sem postular um modelo estatístico subjacente, também servem aqui.

Finalmente, no modelo baseado no estudo de casos, a atenção é centrada em um conjunto muito bem selecionado de detalhes empíricos para cada caso. Mesmo que um indicador seja visto como útil e valido sob o ponto de vista de um outro modelo de validação, para a perspectiva baseada no estudo de casos, este indicador pode ser revisado e rejeitado se ele não for plausível para os casos que estiverem sendo estudados naquele momento.

Pesquisa avaliando medidas de democracia tem o mérito de incluir exemplos importantes trazidos de todos os quatro modelos de validação. Assim, esta literatura

oferece um produtivo foco por ver como os diferentes modelos são aplicados a um domínio específico.

Termos e Distinções-chave

Validade, confiabilidade e erros de medição. A *Validade* diz respeito a preocupação se um indicador, de forma plausível, está medindo a ideia conceitual que ele se propõe a medir. Validação de mensuração envolve uma série de ferramentas para avaliar esta plausibilidade. A *confiabilidade* é atingida se o pesquisador pode, de maneira relativamente segura, acreditar que repetindo a aplicação de uma medida a um estável grupo de casos poderia obter resultados consistentes entre si. A visão padrão, no meio científico, diz que muito embora validade e confiabilidade sejam conceitos distintos, uma medida não deveria ser considerada válida se ela não for confiável. *Erros de medição* podem ser vistos como oriundos de falhas nos procedimentos de medição e de erros daqueles responsáveis pela sua aplicação².

Escalas nominais. As escalas nominais desempenham um papel importante nas pesquisas sobre democracia. Portanto – é essencial observar que, ao contrário do que a desatualizada avaliação negativa, feita pela geração anterior de pesquisadores³ - a visão padrão atual diz que as escalas nominais tem um lugar central no esforço de se alcançar medidas validas, assim como, validas inferências causais⁴. As distinções categóricas nominais, também tem um lugar de destaque nas discussões conceituais sobre democracia⁵.

Definição Procedimental de Democracia. Nós focamos aqui na definição “procedimental”, “institucional” identificadas com o trabalho de Schumpeter e Dahl⁶. Esta é a definição padrão utilizada em pesquisa comparativa sobre democracia analisada neste artigo.

Conceito de Fundo, Conceito Sistematizado e Validade de Medida. O conceito de democracia possui diversos sentidos, e a definição procedimental se constitui apenas um deles. Gallie (1956) descreveu democracia como um “conceito

² Esta discussão se baseia, em parte, em Adcock e Collier (2001).

³ Por exemplo, Young (1981, p. 357), Duncan (1984, p. 126), e Blalock (1982, p. 109 e 110).

⁴ A clássica formulação é Stevens (1946, 1951 e 1975); Collier, LaPorte e Seawright (2012) oferecem uma atualizada discussão. O uso extensivo de logit, probit e variáveis binárias tem desempenhado um papel fundamental em legitimar escalas nominais entre pesquisadores que usam métodos quantitativos. Sobre estas técnicas, ver Aldrich e Nelson (1984), Pampel (2000), e Hardy (1993).

⁵ Sartori (1970 e 2009), Collier e Adcock (1999).

⁶ Schumpeter (1950), cap. 22), Dahl (1971, cap. 1). Ver também O'Donnell e Schmitter (1986, p. 8) e Przeworski et al. (2000, p. 18-22).

contestado por excelência” (p. 148). É produtivo tratar este amplo conjunto de sentidos como “conceito de fundo”. Este pode ser contrastado com o “conceito sistematizado” que os analistas extraíram do conceito de fundo. A escolha de um específico conceito sistematizado normalmente deriva de um espectro de preocupações teóricas e às vezes normativas que instigam uma determinada linha de investigação. Nesta presente discussão, o conceito sistematizado corresponde a definição procedimental de democracia.

Para discutir validade de medida em uma determinada literatura, a análise deve ter seu foco principal no relacionamento entre a medida especificada e o conceito sistematizado. Esta recomendação ressalta um ponto crucial sobre validação de medidas. Como acabamos de observar, o relacionamento entre o conceito de fundo e o conceito sistematizado depende de questões teóricas e normativas, e é contra produtivo se pensar em termos de validade conceitual. Em vez disso, para reiterar, a ideia de validade é melhor para a relação entre indicadores e o conceito sistematizado.

Modelos Estatísticos. Estes são centrais para o SEM-L e não desempenham um papel relevante para os outros modelos de mensuração. Um modelo estatístico é um conjunto de equações as quais relacionam dados observáveis a parâmetros subjacentes, com base em suposições, por exemplo, sobre quais variáveis, forma funcional, sequência temporal, a heterogeneidade causal, e como as possibilidades são representadas. Um modelo estatístico com essas características não deve ser confundido com “modelos formais” da teoria dos jogos.

Contribuição de Ferramentas de Validação mais específicas. Os quatro grande modelos de validação usam, em combinações diferentes, ferramentas mais específicas para a validade de medidas⁷.

Validação de Conteúdo. Esta ferramenta é compartilhada por todas as abordagens. O foco é se o indicador cobre significativamente o grupo de elementos - conceitualizados como universo de conteúdo- que correspondem ao conceito sistematizado a ser medido. É quando, às vezes, o indicador tem sua validade aparente, ele simplesmente faz sentido.

Validação Convergente-Discriminante. Esta ferramenta é importante para o modelo pragmático e SEM-L. A ideia aqui é que indicadores medindo um mesmo conceito vão estar muito mais correlacionados entre si do que indicadores medindo conceitos diferentes, e indicadores com uma correlação fraca entre si podem medir de forma mais acurada conceitos diferentes. Uma versão padrão desta abordagem na

⁷ Ver novamente Adcock e Collier (2001) para uma discussão desses critérios específicos de validação.

literatura sobre democracia – a qual nós vemos como parte do modelo pragmático – é examinar correlações simples entre os indicadores, sem sistematicamente utilizar um modelo estatístico. A informalidade desta abordagem faz dela parte integrante do modelo pragmático. SEM-L, em contraste, usa uma versão altamente sofisticada da validação convergente-discriminante, baseado em elaborados modelos estatísticos, e no reconhecimento de que as correlações fortes ou fracas podem produzir não apenas relações descritivas entre as variáveis, mas também relações causais, e estas precisam ser resolvidas.

Validação Monológica. Esta ferramenta está definitivamente identificada com o modelo pragmático, a qual se baseia naquilo que pode ser visto como abordagem não ortodoxa – mas de fato, as vezes, ainda produtiva. Ela toma como ponto de partida relações causais previamente estabelecidas e examina se essas relações podem ser repetidas com um novo indicador. Com efeito, ela adota a hipótese causal e usa-a para testar a medida.

Modelo Equação Estrutural com Variáveis Ocultas (SEM-L, Sigla em Inglês)

Visão Geral

Este método se baseia na ferramenta chamada validação convergente-discriminante. Ele induz o pesquisador a distinguir dois tipos de relações. Assim, as associações entre os indicadores são hipotetizadas para refletir um misto de (a) descrições alternativas do mesmo conceito subjacente e/ou (b) relações causais entre diferentes conceitos. Essas hipóteses são utilizadas para construir um modelo estatístico, o qual será usado para estimar parâmetros descritivos e explanatórios (BOLLEN, 1989; BOLLEN, RABE-HESKETH e SKRONDAL, 2008)⁸.

SEM-L, seguidamente identificado com o LISREL software package (BOLLEN, 1989, *passim*), é fundamental para literatura sobre democracia comparada, devido as grandes contribuições de Kenneth Bollen. SEM-L também envolve vários tipos de análise fatorial (FABRIGAR e VEGENER, 2011; KIM e MUELLER, 1978a e 1978b) e teoria Item-response (RECKASE, 2009). Discussões econométricas do modelo erros-em-variáveis (GREENE, 2000, p. 375-383) também se encaixam muito bem nesta abordagem.

SEM-L, na realidade, avalia os indicadores para medir de forma precisa os conceitos sob investigação, e para lidar da forma mais efetiva possível com os erros de medição. Ao criar um modelo, os pesquisadores fazem suposições de que os dados

⁸ Ver também Hayduk (1987), Mueller (1996), e Kaplan (2008).

observados são gerados devido a influência de variáveis ocultas não observáveis⁹ – as quais presumivelmente refletiriam os conceitos de interesse. Desta forma, na literatura sobre democracia, os pesquisadores supõem que determinados indicadores refletem, de forma imperfeita, um “verdadeiro” nível de democracia subjacente, assim como eles conceitualizaram. Os analistas devem fazer suposições sobre (a) a verdadeira dimensão da democracia, (b) qual a dimensão que cada variável oculta tem em uma relação medida com cada indicador observado, (c) os erros existentes em cada indicador, e por vezes, (d) relações causais entre as variáveis ocultas.

Alguns elementos dessas quatro suposições podem ser testados por análise estatística. No entanto, não obstante quaisquer testes, eles permanecem como suposições. Além disso – e isso se constitui um dilema deste modelo – resultados empíricos são difíceis de interpretar, a menos que as suposições sejam válidas.

Equações Estruturais e Democracia

Os esforços para avaliar e aperfeiçoar medidas transnacionais de democracia com SEM-L tem centralizado as suas preocupações com os erros de medição, como mostra o trabalho de Shen e Williamson (2005). Como parte de um amplo estudo focado nas percepções sobre corrupção, estes autores usaram equações estruturais que incorporam um modelo de mensuração de democracia, que permitiu a eles avaliar a proporção de erros nos indicadores da Freedom House sobre direitos políticos, liberdades civis e liberdade de imprensa¹⁰. Estas três medias são consideradas os indicadores de democracia em uma sociedade, e a cada indicador é permitido ter uma modesta margem de erro. As análises de Shen e Williamson, portanto, demonstram confiança nos rankings da Freedom Houe.

Bollen e Paxton (2000), seguindo de perto o trabalho já publicado por Bollen (1993), sugerem uma aplicação diferente para o SEM-L. Ao invés de incorporarem um modelo de mensuração da democracia em um universo causal muito maior, eles focaram a sua atenção em avaliar a qualidade das medidas em diferentes indicadores transnacionais de democracia. Além dos rankings elaborados pela Freedom House, eles incorporaram os dados organizados por Arthur Banks (1971 e 1979). Mais do que isso, eles avaliaram a possibilidade de ameaça a validade dos dados devido às distorções introduzidas pelas instituições e autores que criaram cada medida. Por exemplo, eles concluíram que os indicadores produzidos pela Freedom House são

⁹ Como é possível se notar acima, variáveis ocultas(latentes) são, ass vezes chamada de “conceitos”, fatores ocultos(latentes), ou, simplesmente, fatores.

¹⁰ Disponível em <http://www.freedomhouse.org/report/freedom-world/freedom-world-2005>. Visualizado em 31 de marco de 2013.

distorcidos em favor de países de orientação católica, em detrimento dos países de orientação marxista (p. 74-77), ao mesmo tempo em que os dados organizados por Banks trazem uma distorção oposta. Em comparação com Shen e Williamson, os quais estimaram uma margem de erro de 7%, a análise de Bollen e Paxton sugere uma margem na ordem de 15 a 20% para os dados organizados pela Freedom House. O modelo tenta reduzir os erros de medição através da introdução de uma média ponderada para os indicadores disponíveis.

Treier e Jackman (2008) analisam os erros de medição nos dados produzidos pelo Polity¹¹. Eles usaram um modelo de item-response, baseado em um fator oculto de democracia unidimensional hipotetizado (p. 204 e 205). A análise coloca a mostra uma substancial quantidade de erros de medição. Em uma tentativa de remediar esses erros, Pemstein, Meserve e Melton (2010) usam ferramentas similares para criar uma ponderação aperfeiçoada das medidas existentes, com o objetivo de gerar um novo indicador com reduzida distorção e confiabilidade aprimorada. Estes dois estudos mostram a oportunidade de termos uma visão cumulativa sobre erros de medição.

Críticas sobre o SEM-L por metodologistas

Essas excelentes aplicações do SEM-L foram realizadas por renomados pesquisadores, porém ainda assim, um certo ceticismo sobre este modelo precisa claramente ser mantido em vista. Alguns dos mais contundentes comentários estão na introdução deste artigo. Este modelo é visto como “predominantemente senso comum” (FREEDMAN, 1987, p. 102), como um “desastre” (CLIFF, 1983, p. 116), e como uma “ciência patológica” (MICHELL, 2008, p. 10). Mais especificamente, os críticos desta abordagem, trouxeram à tona um número bastante grande de não testadas e, até mesmo, não testáveis suposições sobre as quais ela se apoia enquanto metodologia. O que exatamente significa, enquanto um postulado teórico e empírico, presumir uma combinação específica de distribuição estatística, para uma coleção de variáveis ocultas? Que tipo de evidência poderia demonstrar que tal presunção, ou suposição, está certa ou errada?¹² Psicometristas tem dedicado grande atenção, alguns sendo extremamente críticos, ao problema das suposições. Michell (2008) sugere que

¹¹ Marshall (2013).

¹² A teoria estatística que está por trás do modelo de equações estruturais, em geral, exige que o fator oculto(latente) de democracia seja normalmente distribuído. Se esta hipótese for falsa, as estimativas e inferências podem ser enganosas. Claro, o mesmo eh verdadeiros para a maioria dos procedimentos estatísticos. No entanto, quando a hipótese distributiva fundamental envolve uma variável que, em princípio, não pode ser observada fica difícil imaginar como este tipo de suposição poderia ser detectada de forma confiável.

no seu campo, “a hipótese central (que atributos psicológicos são quantitativos) é admitida como verdade na ausência de evidências, sobre as quais ela possa estar apoiada... Psicometristas afirmam saber aquilo que eles não sabem, ao mesmo tempo em que erguem barreiras para preservar sua ignorância (MICHELL, 2008, p. 10).

The fato, resultados de análise empírica somente testam a hipótese central, sobre medidas e causalidade, uma vez que o modelo de suposições seja preciso. Assim como em qualquer análise estatística, estes resultados são basicamente um produto do modelo, muito mais do que um teste. Entre os elementos chave deste modelo estão suposições sobre variáveis não observadas: o seu número, distribuição, dimensão e a estrutura das relações de medição com as variáveis observadas. Sem este tipo de suposições, o modelo se torna impossível.

A teoria item response (IRT) tenta contornar este tio de problema. No entanto, apesar das diferenças em ênfase e procedimentos, as duas técnicas têm fundamentalmente suposições similares (RECKASE, 2009; TAKANE e DE LEEUW, 1987; TREIER e JACKMAN, 2008, p. 205-206). Assim, muito embora a IRT dê atenção a uma gama de questões interessantes, negligenciadas na maioria das análises que usam o modelo SEM-L, ela não consegue fugir das preocupações discutidas aqui.

Não obstante as críticas, o modelo de equação estrutural contribui para o processo de medição de várias maneiras. Ele (a) reúne os diversos procedimentos de validação de medição desenvolvidos por psicometristas, (b) enriquece o trabalho de medição encorajando os pesquisadores a focarem sua atenção diretamente sobre as conexões causais e descritivas entre os indicadores, (c) fornece estimativas sobre erros aleatórios e distorções, e (d) procura fazer inferências causais que não estejam contaminadas por problemas de medição. Desta forma ele dá aos pesquisadores evidências sobre a qualidade dos indicadores, muito embora o valor destas evidências esteja condicionado as suposições mencionadas acima.

Modelo Níveis de Medição (LoM, sigla em inglês)

Visão Geral

A preocupação central deste modelo está relacionada com as restrições lógicas sobre as técnicas estatísticas mais apropriadas para cada nível de medição. Ela decorre de uma longa história de trabalho sobre medição se afastando das contribuições fundamentais de Stevens (1946, 1951 e 1975; para uma das mais recentes sínteses, ver GILL, 2006, p. 300-3004), assim como da tradição axiomática de medição (KRANTZ *et al.*, 1971; SUPPES *et al.*, 1989). Devido as presumidas restrições, ele

também se preocupa com os métodos de ampliação – isto é, mover para o alto LoM – entre nominal, ordinal, intervalo, e os dados de razão nominal, ampliando assim o leque de técnicas estatísticas adequadas¹³. Desta forma, começando com a escala nominal, nós podemos perguntar, “Quais os atributos estas categorias devem ter, e que tipo de técnicas analíticas podem ser aplicadas para que os pesquisadores possam considera-las nominais?” Se ordem é estabelecida, que critérios adicionais devem ser atendidos para estabelecer uma unidade de medida, afim de se obter uma escala intervalar ou de razão?

Uma preocupação recorrente aqui é sobre como se alcançar uma medição ordinal, a qual é o objetivo de Guttman Scaling (1950; ENGELHARD, 2008). Esta técnica é usada para testar uma possível ordem subjacente, e quando tais testes são satisfeitos, os pesquisadores podem converter categorias nominais em uma escala ordinal.

A Guttman Scaling é aplicada em situações nas quais uma série de critérios – por exemplo, atributos da democracia- são hipotetizados para refletir diferentes posições ao longo de uma presumida dimensão. Alguns dos casos sob análise podem cumprir aos mais exigentes critérios, o que corresponde ao mais alto valor, ao passo que outros podem apenas cumprir os critérios menos exigentes. A Guttman Scaling postula que se os casos que cumpriram os mais exigentes critérios, também cumpriram os menos exigentes, então, uma única dimensão ordenada foi estabelecida.

Uma outra meta do LoM modelo, além da medida ordinal, é estabelecer uma significativa unidade de medida e por conseguinte aditividade, com alguns teóricos, inclusive, afirmando que estes são os critérios mínimos a serem atingidos para um determinado indicador qualificar-se enquanto uma medida (CAMPBELL, 1920; GRANT, 2004; KARIYA e FINKELSTEIN, 2000). Por isso, uma das preocupações centrais desta abordagem tem sido oferecer provas de que estes critérios foram alcançados. Técnicas empíricas, orientadas nesta direção inclui os modelos de medida Racsh (ANDRICH, 1988; BOND e FOX, 2012; FISCHER e MOLENAAR, 1995).

¹³ Quando as técnicas do modelo de equações estruturais são usadas para estimar classificações fatoriais, com base em variáveis categóricas, ou mesmo dicotômicas, uma transformação entre diferentes níveis de medição ocorre. Algumas técnicas combinam preocupações destes dois modelos; exemplos são discutidos abaixo. No entanto, muitas técnicas que produzem classificações fatoriais baseadas em variáveis categóricas não estão preocupadas com questões relacionadas com ordem, igualdade, e unidades de medida consistentes que derivam do modelo níveis de medição.

LoM e a Democracia

Uma questão conceitual precisa ser resolvida antes de discutir LoM. Existe um acirrado debate, o qual tenta definir se democracia versus não democracia deve ser sempre uma dicotomia (PRZEWORSKI *et al.*, 2000; SARTORI, 2009), ou uma continuidade (BOLLEN e JACKMAN, 1989)¹⁴. Esta importante discussão diz respeito ao conceito de fundo de democracia e o conceito sistematizado adotado por esses autores. Como enfatizado acima, este tipo de disputas envolve importantes questões de ordem conceitual e normativa, e elas são vistas aqui separadamente de questões de validade de medidas.

No que concerne ao LoM e a busca pela validade de medidas, Coppedge e Reinicke (1990) procuram ir além dos dados nominais para criar uma escala ordinal de poliarquia usando a escala Guttman. No seu índice 1-7, sendo 1 o mais alto grau de democracia, consequentemente 7 o mais baixo, o ranking dos países é considerado cumulativo se, por exemplo, todos os países posicionados na categoria 6 possuem todas as características democráticas dos países posicionados na categoria 7, assim como, mais algumas características adicionais. O mesmo deve ser verdadeiro para toda a extensão da escala. De fato, Coppedge e Reinicke conseguem localizar 137 dos 170 países da escala. Os outros 33 países são ambíguos, nos termos que, a particular combinação de características democráticas e autoritárias não atinge este padrão cumulativo. Por isso, uma regra clara para pesar as diferentes características se faz necessária para atingir o principal objetivo do LoM de estabelecer uma ordem rigorosa. A escala Guttman, portanto, é apenas uma forma parcial de ordenamento.

Baker e Koesel (2001) estendem Coppedge e Reinicke (1990) abordagem gerando escalas Guttman para quatro componentes de uma poliarquia: Eleições, livre expressão, inclusão, e equilíbrio governamental. Usando uma base de dados sobre as classificações anuais para os países do leste europeu 1992-2000, os autores foram capazes de classificar de forma inequívoca praticamente todos os países- ano para os três primeiros componentes considerados, estabelecendo com sucesso três escalas ordinais. Para a dimensão equilíbrio governamental, os resultados foram mais ambíguos, com 68 dos 117 países caindo em categorias mistas. Por isso, não foi possível incluir esta dimensão no processo de ordenamento. Ao contrário de Coppedge e Reinicke, Baker e Roesel não fazem tentativa alguma para estabelecer um escore poliárquico para cada país.

Usando uma perspectiva mais qualitativa, Munck e Verkuilen (2002) descrevem as técnicas de ordenação e agregação necessárias para captar os atributos

¹⁴ Veja uma revisão sobre o debate em Collier e Adcock (1999).

chave que definem uma democracia. Eles insistem que pesquisadores criam escalas com o menor número possível de categorias necessárias para alcançar caso equivalência dentro de cada categoria –uma ênfase que se encaixa em um dos focos do LoM de estabelecer uma significativa relação igual-desigual (MUNCK, 2009; MUNCK e VERKUILEN, 2002, p. 17)¹⁵. Além disso, para agregar indicadores de diferentes traços de um regime e transforma-los em uma medida geral de democracia, estes autores argumentam que:

“Primeiro, os analistas devem explicitar a teoria de sustentação da relação entre os atributos. Segundo, os analistas devem assegurar que há uma correspondência entre esta teoria e a regra de agregação escolhida, isto é, que a regra de agregação seja verdadeiramente a expressão formal equivalente da relação postulada” (MUNCK e VERKUILEN, 2002, p. 24).

Os autores, desta forma, compartilham a ênfase, descrita cima, em transformar sub indicadores em uma classificação final de democracia de uma forma que preserva ordem e igualdade, que são preocupações centrais do LoM. Com um espírito similar, Coppedge e Gerring (2011; veja também COPPEDGE, 2010) oferecem uma inovadora solução para o desafio de chegar a uma ordem, e a uma regra de agregação que preserva a igualdade: publicar indicadores desagregados, permitindo que os pesquisadores adotem a regra de agregação que servir melhor para o seus propósitos.

A análise de Treier e Jackman (2008) dos indicadores Polity, discutida anteriormente, na sessão sobre o “SEM-L”, também se serve do LoM. Dito isso, se faz necessário um alerta sobre o tratamento de dados ordinais como se estivessem contidos neles intervalos iguais. O modelo deles estima uma distância subjacente entre as categorias dos indicadores ordinais dos quais a medida Polity é composta. Baseados nestas estimativas, os autores concluem que, “é possível observar a partir das distâncias entre limiares que muitos incrementos diferenciais especificados nos cálculos do Polity não são válidos” (p. 16). Algumas categorias adjacentes são estimadas como sendo virtualmente idênticas em termos de escala subjacente, enquanto que outras são radicalmente distantes nesta escala. Assim, uma análise mais rigorosa deve ser deixada para o nível ordinal de medição.

Críticas do LoM por Metodólogos

Já consta, na introdução deste artigo, algumas das mais contundentes críticas a este modelo: Por exemplo, MoL produz “uma análise de dados ruim, é uma má

¹⁵ Veja também Foweraker e Krznaric (2000).

ciência” além de ser “enganosa” (VELLEMAN e WILKINSON, 1993, p. 70, 72). Tukey (1986) também centra severas críticas ao LoM. Ele vê como uma perda de tempo, uma atenção demasiada em estabelecer ordinalidade ou distâncias iguais entre as classificações, porque técnicas de padrão estatísticos funcionam muito bem mesmo que os indicadores usados parcialmente falharem para este critério (BAKER, HARDYCK e PETRINOVICH, 1966). Estas preocupações, as quais rejeitam a principal premissa do LoM, são a peça central do modelo pragmático, o qual vai ser o nosso próximo tópico.

Modelo Pragmático

Visão Geral

Tukey (1986), como é possível observar, defende uma abordagem muito mais flexível do que puritana dos processos mensuratórios, evitando desta forma uma “demasiado simplificada, assim como, purista” visão. Ele argumenta que o puritanismo é “perigoso”, e com esse espírito sensível, a sua lista de maus conselhos metodológicos inclui o sarcástico “não pense, use estatística” (TUKEY, 1986, p. 202, 243, 244 e 247). Ele sugere que, no lugar da aplicação de padrões rígidos, “um conjunto de dados pode produzir a sua própria análise” (TUKEY, 1986, p. 207).

O modelo pragmático postula que os analistas deveriam considerar padrões, para uma boa mensuração, que estejam fora dos limites das estruturas tradicionais. Estes critérios podem, muitas vezes, substituir as preocupações dos cânones tradicionais de mensuração, porém este modelo, não raras vezes trabalha no sentido contrário, a partir de uma aplicação específica, para a escolha dos indicadores. Devido a ampla gama de agendas muito importantes e ferramentas analíticas em Ciências Sociais e Estatística, isto pode sugerir uma variedade de prioridades divergentes em medição.

O modelo pragmático desempenha um importante papel em pesquisa quantitativa, e declarações sobre isso vem sendo feitas já desde muitas décadas atrás. Entre os primeiros a se manifestarem está a contundente crítica de Lord (1953) contra a estrutura de níveis de medida de Stevens (1946 e 1951) e as correspondentes permissíveis operações estatísticas. Lord faz piada da ideia de que um dado grupo de dados seja inerente a um determinado nível de medida, afirmando que “os números não se lembram de onde eles vem” (p. 751). Dependendo das circunstâncias, aquilo que inicia como uma escala nominal pode naturalmente ser tratado como ordinal, ou algumas vezes, mesmo como intervalo.

Em contribuições subsequentes, Tukey (1986, p. 237-243) argumenta que pesquisa científica deve ser guiada pela experiência: SE um determinado procedimento parece funcionar bem para o problema em questão, ele deve ser adotado, sendo ou não justificado, por exemplo, por um LoM argumento. Em resposta a uma síntese feita por Luce (1959) sob o ponto de vista do LoM, com relação aos tipos de escalas e operações estatísticas admissíveis, Tukey, de forma incisiva, responde que “as limitações discutidas por Luce não controlam que tipo de estatística poderia ser usada sensatamente, mas apenas que tipo poderia ser usada de forma puritanista” (p. 244).

Abelson e Tukey (1963) usam análise estatística de um conjunto de dados simulados para analisar os resultados do tratamento de dados ordinais como se fosse um intervalo. Esta abordagem parece razoável, dado que as correlações entre os valores atribuídos e os valores verdadeiros são, na sua maioria, altas e esta abordagem parece não dar muita atenção às inferências a partir dos dados. Por isso, se há um argumento viável que uma determinada variável satisfaz os critérios para a medição de intervalo, poderá também haver uma forma praticas para se operar neste nível de medição.

A escala Multidimensional (MDS, sigla em Inglês) seguidamente está vinculada ao modelo pragmático. Este método pretende representar as semelhanças e diferenças entre casos em termos do que é normalmente um espaço bidimensional. É verdade que modelos tem sido desenvolvidos para justificar esta técnica e procedimentos de escala podem, as vezes, ser estatisticamente consistentes (BRADY, 1985). Ainda assim, em geral estes procedimentos são justificados mais pela sua utilidade intuitiva e valor heurístico das exposições que produzem do que por serem um modelo estatístico formal. Na verdade, o MDS é seguidamente aplicado em contextos exploratórios onde nenhum outro modelo estatístico tenha sido postulado, ficando assim, os argumentos pragmáticos, como a única justificativa disponível para os procedimentos. Como Kruskal e Wish (1978, p. 26-27) afirmam, “a mais importante justificativa é que o MDS funciona e eh útil”.

A técnica chamada alternating least square, optimal scaling (ALSOS) (JACOBY, 1999; YOUNG, 1981) nos oferece uma outra versão do modelo pragmático. Aqui as variáveis categóricas recebem um dado valor inicial e são tratadas como nível de intervalo em uma análise de regressão. As variáveis então, recebem um novo valor com o objetivo de minimizar as variações não explicadas na regressão. Este processo é repetido até se alcançar um resultado satisfatório. Com efeito, as variáveis categóricas são convertidas em variáveis de nível de intervalo e se escolhe os valores

que produzem os melhores resultados na análise de regressão¹⁶. A tradicional preocupação com o LoM é desta forma abandonada, e uma variação de validação nomológica é empurrada para um extremo.

O Modelo Pragmático e a Democracia

Uma aplicação padrão do modelo pragmático acontece quando os autores que criaram uma nova medida transnacional justificam esta medida baseados na sua alta correlação com existentes, geralmente aceitas, medidas – isto é, validação convergente. Em contraste com o SEM-L, nenhum modelo de medição é postulado; ao invés disso, a validação convergente é adotada com uma prática (e muitas vezes não suficientemente teorizada) checagem sobre a validade dos indicadores. Por exemplo, Przeworski *et al.* (2000) invocam um critério pragmático para justificar o seu bastante conhecido indicador de democracia entre 1950 e 1990, um indicador com base em uma classificação dicotômica de democracia/não democracia. Eles argumentam que, “apesar de todas as diferenças conceituais e observacionais, as várias abordagens produzem classificações de regimes altamente similares. Desta forma, não há razão para se pensar que os resultados dependem de uma maneira específica pela qual os regimes foram classificados” (p. 55)¹⁷. Assim, não obstante a sua insistência em termos teóricos de que a democracia deveria ser medida de uma forma distinta – isto é, como uma dicotomia – Przeworski *et al.* consideram a sua medida um sucesso, em parte, com base no critério pragmático de se correlacionar com outros indicadores estabelecidos que não usam a dicotomia¹⁸. Esta abordagem, em um sentido, abandona o seu argumento sobre dicotomia.

Ao justificar a sua medida tricotômica de democracia, semidemocracia e não-democracia para 19 países da América Latina, entre 1945 e 1999, Mainwaring, Brinks e Pérez-Liñán (2001, p. 48) fazem um argumento paralelo. Eles também dependeram, em parte, da validação pragmática baseada em fortes correlações com outros indicadores que não são tricotômicos, assim, em um certo sentido, deixando

¹⁶ Esta técnica adota a abordagem da validação nomológica (ADCOCK e COLLIER, 2001).

¹⁷ As evidências apresentadas em defesa desta proposição envolve uma aplicação de logit, com a nova medida como variável dependente e outras medidas como variável independente. Eles concluem que já existentes medidas de democracia podem prever corretamente entre 85% e 94% das classificações em Przeworski *et al.* (2000, p. 56-57).

¹⁸ Przeworski *et al.* Prefere as suas medidas ao invés de outras, principalmente por razões conceituais. Embora as questões conceituais, muitas vezes, desempenham um papel central nas decisões sobre medidas, a literatura sobre medição se distingue da literatura substantiva, e das discussões sobre conceitualização, por sua ênfase em ferramentas relacionadas com a tarefa de se mover de um determinado conceito para uma, mais exata, classificação para o caso.

de lado o seu argumento em favor de trabalhar com três categorias (MAINWARING, BRINKS e PÉREZ-LIÑÁN, 2001, p. 53). Estes autores deram este passo apesar de terem feito uma argumentação bastante forte sobre o porquê de seus argumentos serem diferentes dos outros. Eles evocam um adicional argumento pragmático, a saber: “Dado o nosso custo e tempo limitado, teria sido muito difícil construir uma medida exata para cada país e cada ano desde 1945” (p. 50). Custo e tempo são certamente considerações pragmáticas.

Casper e Tufis (2003; ver também CHEIBUB, GHANDI e VREELAND, 2010) ilustram uma forma alternativa do modelo pragmático, no qual indicadores alternativos os quais parecem altamente semelhantes entre si, de fato, produzem resultados diferentes quando dos testes das hipóteses. Estes autores usam um grupo padrão de variáveis independentes, mas introduzem diferentes indicadores de democracia como variável dependente, os quais, muitas vezes produzem resultados significativamente diferentes. Por exemplo, “Nós podemos observar que, muito embora educação fundamental seja significantemente associada com democracia, quando usamos Poliarquia {i.e., a medida Coppedge}, a sua importância cai, quando usamos Polity” (p. 198-200). Assim, eles usam uma variante de validação nomológica para explorar a plausibilidade relativa dos indicadores alternativos.

Com base nestes resultados, Casper e Tufis (2003) continuam defendendo que a validação convergente pode ser um critério inadequado para estabelecer que dois diferentes indicadores medem um conceito. Ao invés disso, estes autores aplicam a alternativa da abordagem pragmática, a qual diz que, medidas equivalentes deveriam produzir semelhantes inferências causais. Estes dois critérios pragmáticos – os quais podem estar em conflito um com o outro, como no presente exemplo – devem ser levados a sério. E, se o objetivo dos pesquisadores for uma viável inferência causal, os resultados de Casper e Tufis parecem ter importância ainda maior.

O modelo pragmático também tem sido usado para providenciar evidências empíricas sobre o mérito de uma classificação ordenada, como oposição as medidas dicotômicas. Elkins (2000) usa dicotômicas e ordenadas versões¹⁹ a partir do indicador de democracia do Przeworski *et al.* Como uma variável alternativa independente prevendo inter-democratic Peace e estabilidade de regime. Ele tenta avaliar se isto é possível, usando esses indicadores, repetir importantes resultados sobre relações causais que muitos pesquisadores veem como bem estabelecida em literatura anterior. Em ambas situações, Elkins constata que as medidas ordenadas

¹⁹ Elkins criou uma versão graduada das variáveis, a partir do indicador de democracia em Przeworski *et al.* Veja Elkins (2000, p. 296).

produzem uma mais sutil capacidade de repetição de resultados anteriores, e também, seguidamente mais convincentes estatisticamente. Desta forma, através da validação nomológica, Elkins consegue concluir, baseado no critério pragmático de produzir inferências causais mais consistentes com teorias anteriores, que a não dicotomia é preferível e desejável.

Outros dois exemplos do modelo pragmático são encontrados em Bollen. O seu artigo de 1980 desafia a inclusão do índice de comparecimento eleitoral como um indicador de poliarquia mostrando que desta forma, a poliarquia é fracamente, ou negativamente associada com outros indicadores de democracia. A sua análise é um passo importante no questionamento sobre ser ou não apropriado incluir o comparecimento eleitoral como um componente de medida. Além do mais, muito embora Bollen (1993) confie primeiramente no SEM-L, a análise ultrapassa a barreira do modelo, em direção ao critério pragmático, quando ele desenvolve um novo grupo de classificação de democracia em 1980. Ele usa o critério pragmático para justificar a sua decisão de rejeitar o fator pontuação para construir a sua própria classificação, confiando, ao invés disto, na média de três indicadores já existentes. Ele justifica este passo argumentando que, quanto mais simples as técnicas mais serão produzidos pontuações estáveis de um ano para o outro. Assim, o que inicia como um exemplar estudo no SEM-L termina como um estudo no campo pragmático que produz um indicador mais útil.

Críticas ao Modelo Pragmático por Metodologistas

O risco do modelo pragmático é que, ele pode se transformar em um tratamento ad hoc de inferências descritivas, uma falta de atenção a sistemáticas medições, e no pior dos cenários, a escolha das medidas apenas porque elas confirmam as hipóteses que estão sendo investigadas. Se o processo de medição obedecer a outros objetivos, a análise pode perder completamente o contato com a descrição, desta forma abandonando as suas firmes ligações com o mundo empírico. Qualquer pessoa que tenha procurado explicar para outros a ideia de validação nomológica – para reiterar a abordagem que começa com uma hipótese explicativa estabelecida e usá-la como um ponto de referência para a avaliação de uma medida – sem dúvida nenhuma tem, mais do que ocasionalmente, encontrado intenso ceticismo.

Não obstante estas questões, o modelo pragmático serve como um útil lembrete de que, qualquer resultado estatístico interessante é digno de exploração adicional, mesmo se os pressupostos de avaliação por trás da análise parecerem difíceis ou impossíveis de se justificar. Tomar um exemplo que reflete a prática padrão em

Ciências Políticas, se uma regressão usando uma escala nominal não transformada e potencialmente subteorizada, como uma variável independente produz uma sequência estatisticamente significativa. Então, em princípio, o pesquisador descobriu um quebra-cabeças que merece ser mais explorado.

Modelo Baseado em Estudo de Casos

Visão Geral

De acordo com o modelo baseado em estudo de casos, os indicadores devem representar, da forma mais precisa possível, os detalhes mais relevantes analiticamente, do caso, e esta correspondência deve ser testada através de uma avaliação direta das classificações a luz de fontes primárias e secundárias. Esta abordagem confia em um profundo estudo dos casos, construindo base de dados (BOWMAN, LEHOUCQ e MAHONEY, 2005; COLLIER, 1999) e tendo uma ampla visão das informações relevantes (COPPEDGE, 1999). Ela vai além das outras abordagens, as quais, obviamente também, confiam em classificações baseadas no conhecimento de caso, na coleta de muito mais detalhada de informações sobre cada caso e avaliando a correspondência entre profundo conhecimento e a classificação para o caso a partir de um determinado indicador. Em uma análise baseada em estudo de caso, envolvendo uma quantitativa e qualitativa comparação transnacional, o indicador poder ser uma medida categórica, ou uma medida com alto nível de precisão. As falhas na classificação para adequadamente refletir as informações detalhadas sobre o caso exige uma recontagem do caso e pode levantar questões mais amplas sobre o indicador geral.

No modelo de estudo de casos, um erro de medição não é visto tipicamente como um componente aleatório de um indicador, o qual, no modelo de equação estrutural, seria resolvido através de meios estatísticos. Pelo contrário, ele é considerado um erro de classificação isolado, por informações não adequadas, ou distorção por parte do pesquisador. Erros de classificação devem ser corrigidos através de uma estreita e presumivelmente “imparcial” atenção aos fatos, sobre eventos específicos, lugares e pessoas.

Elementos do modelo de estudo de casos podem ser encontrados em uma variedade de abordagens metodológicas. Em estudos médium-N transnacional, pesquisadores podem ter um alto nível de conhecimento do caso e ao mesmo tempo usar ferramentas estatísticas, identificadas com outros modelos de medição²⁰. Do

²⁰ Por exemplo, Mainwaring, Brinks e Pérez-Liñán (2001).

mesmo modo, alguns trabalhos com metodologia qualitativa dá ao estudo de casos o orgulho do local de medição. Por exemplo, o estudo de casos comparativo (GEORGE e BENNETT, 2005), a Ciência Política e literatura sociológica em análise histórico-comparativa (MAHONEY e RUESCHEMEYER, 2003), e o método de comparação contextualizada (LOCKE e TELLEN, 1995). Todas elas enfatizam a importância de se confiar em detalhes muito profundos de caso como fundação da pesquisa comparativa.

No modelo de estudo de casos, o nível de detalhamento e conhecimento dos casos em estudo podem variar grandemente. Primeiro, seguindo o que pode ser a “melhor prática” pesquisadores trabalhando com um pequeno N, cuidadosamente discutem as evidências que justificam as classificações para casos particulares. Muitas das pesquisas histórico-comparativas incluem excelente exemplos de explícita e detalhada exposição de critérios e evidências envolvendo classificação, com as evidências comumente em um quadro de narrativa, ou uma quase-narrativa. É por isso que esse tipo de estudo seguidamente se parece com livros, mais do que artigos. Segundo, outros estudos, usando um N maior, no entanto, providenciam discussões relativamente detalhadas de como, pelo menos, alguns casos são codificados. Terceiro, ainda em outros exemplos, o conhecimento de caso, por parte dos autores estava presente no processo de classificação, mas a apresentação (talvez por espaço restrito no veículo da publicação do artigo) providenciam poucas evidências das escolhas específicas feitas na codificação. Enquanto este tipo de decisão pode significar um profundo conhecimento de caso, as falhas em apresentar os detalhes relevantes pode impedir os leitores de verificar as medidas derivadas de tal conhecimento.

Esta perspectiva pode servir como uma útil, e parcialmente compatível, alternativa para as necessárias, não testadas, pressuposições no modelo de equações estruturais, para as abstrações envolvidas no modelo níveis de medição, e para o tratamento, por vezes, ad hoc dos indicadores no modelo pragmático. De uma forma mais ampla, o modelo de medição baseado em estudo de casos, desempenha um papel muito útil para demonstrar os limites das medições, de uma maneira geral, pois elas tem mostrado que capturam de forma inadequada os detalhes dos casos.

O Modelo Baseado em Estudos de Casos e a Democracia

Um espectro de abordagens baseadas em estudo de casos tem sido usadas em debates sobre medidas de democracia. Assim como reconhecemos as excelentes praticas do modelo de equação estrutural, nós podemos apontar para um outro extremo deste espectro, onde podemos encontrar uma altamente detalhada

apresentação de evidências para justificar as medidas adotadas. Esta extremidade está representada pelo modelo baseado em estudo de casos. Por exemplo, no seu estudo histórico comparativo de dois diferentes períodos e regimes, na Costa Rica e Guatemala (i.e., N=4), Yashar (1997), oferece vários capítulos de evidências para sustentar a sua identificação dos episódios de democratização nos anos 40 e 50, assim como as realizações dos regimes democrático e autoritário desde a metade do século XX. Mahoney (2001), analisando um grupo de 5 países da América Central, da mesma forma, apresenta um detalhado estudo de caso para identificar episódios democratizantes durante a primeira metade do século XX, assim como uma comparação de resultados dos períodos democrático e autoritário durante a segunda metade do século.

Há também estudos se utilizando de um universo (N) maior. Como exemplo pode ser citado R. B. Collier (1999) que usa dados histórico comparativos para sustentar a codificação de 38 democratizações em 27 países. Uma série de relevantes detalhes são apresentadas para cada caso. No entanto, a quantidade de informações apresentadas para justificar a codificação de cada caso se mostra limitada, se comparada com os exemplos da América Central acima, onde o universo (N) utilizado era menor. Em um passo adiante, ainda dentro do mesmo espectro, Rueschemeyer, Stevens e Stevens (1992) analisam um histórico de regime de aproximadamente um século para 40 países. Enquanto estes autores, às vezes, apresentam informações específicas para justificar a codificação de um regime, ou um período de transição, o escopo desses estudos inevitavelmente limitam os detalhes dos casos, os quais os autores podem usar sem que outros aspectos analíticos sejam negligenciados.

Uma outra rota para um universo (N) maior pode ser vista em estudos que analisam um número relativamente pequeno de países durante um longo período de tempo, obtendo desta forma um alto nível de especialização sobre cada país. Bowman, Lehoucq e Mahoney (2005), em seu artigo “Case Expertise” foca sua atenção em 5 países para gerar uma série temporal N de 500, e Mainwaring, Brinks e Pérez-Liñán (2001) analisa 19 países para gerar uma série temporal N de 855. Ambos estudos analisam detalhadamente cada caso e apontam importantes divergências com relação a medidas anteriores. A credibilidade dos indicadores destes estudos vem de uma variedade de fontes. Em graus variados, os autores discutem como determinados casos foram classificados. Bowman, por um lado, usa evidências sobre determinados países-anos da história de Costa Rica, El Salvador, Honduras, Guatemala, e Nicarágua (p. 946-949) para documentar o problema de fontes secundárias imprecisas, parciais, ou enganosas na construção de indicadores transnacionais de

democracia (p. 940). Por exemplo o indicador Polity dá para a Costa Rica uma classificação de democracia completa para cada ano do século XX, apesar da relevância que tiveram golpes e intervenções militares durante a primeira metade do século XX. Em um outro contexto regional, Berg-Schlosser (2004) traz uma análise similar, porém com foco na África.

Mainwaring, Brinks e Pérez-Liñán (2001), por outro lado, se coloca na outra extremidade do espectro do modelo baseado em estudo de casos: Eles trazem evidências menos detalhadas de cada caso, o que não se constitui surpresa, já que eles analisam 19 países da América Latina por 45 anos. Os autores explicam os seus critérios de codificação para classificar país-ano, e eles apresentam evidências ilustrativas para justificar algumas das suas decisões sobre medidas. No entanto, com tantos casos, não é factível apresentar o nível de detalhes que é característico do modelo histórico comparativo, ou mesmo o número de informações específicas por país apresentado por Bowman, Lehoucq e Mahoney (2005).

A credibilidade das medidas de Mainwaring, Brinks e Pérez-Liñán (2001) também deriva da reputação de pesquisadores academicamente reconhecidos por seu profundo conhecimento sobre os países estudados, o que acaba evidenciando que os casos foram classificados de forma cuidadosa. Além disso, Mainwaring, Brinks e Pérez-Liñán (2001) e Bowman, Lehoucq e Mahoney (2005) fazem incursões fora do modelo baseado em estudos de casos, na direção do modelo pragmático, nas quais eles usam a abordagem de validação convergente para comparar os novos indicadores com espectro de medidas já existentes.

Uma diferente aplicação do conhecimento de casos pode ser encontrada em um análise de O'Donnell (1996) sobre a história do regime argentino entre 1983 e a metade dos anos 90. Ele rejeita uma avaliação transnacional sobre consolidação democrática, a qual se baseava no número e na severidade de crises que um dado regime democrático tenha sobrevivido. O argumento seria que o regime superou mais obstáculos, portanto ele estaria mais consolidado. Em aproximadamente uma década após a sua transição para o regime democrático em 1983, a Argentina já havia passado por um processo politicamente traumático de levar a justiça agentes militares, por violação dos direitos humanos, durante a ditadura, múltiplas tentativas de golpe pelos militares e uma prolongada crise econômica. Porque estes obstáculos foram mais severos do que aqueles superados por diversos países do sul da Europa, em aproximadamente o mesmo período de tempo, os indicadores transnacionais sugeririam que a democracia argentina estava mais consolidada do que em Portugal e Espanha.

Ainda assim, O'Donnell (1996), com base em um profundo conhecimento sobre os processos políticos na Argentina, argumenta que as instituições básicas de democracia eram mais tênues e menos solidas neste país do que no sul da Europa. Assim, informações adicionais sobre um pequeno número de casos leva a rejeição do indicador transnacional, o qual O'Donnell caracteriza como *reductio ad absurdum*.

Outros exemplos de estudos baseados em estudo de casos que buscam preservar, ou mesmo melhorar a validade das medidas, são apresentados na discussão levada a cabo por Collier e Levitsky (1997) de como pesquisadores criam novos subtipos de democracia. Estes autores mostram que analistas têm criado categorias nominais para mais adequadamente servir a um determinado tipo de caso, ou grupo de casos, desta forma evitando um alargamento do conceito (SARTORI, 1970). Deste modo, dentro da, em evolução, literatura comparada sobre democracia nos anos 80 e 90 do século XX, pesquisadores adaptaram, de forma indutiva, suas categorias baseados em informações sobre estudo de casos, refletindo a interação do conhecimento do caso e do entendimento conceitual sobre democracia. Uma faceta desta adaptação envolve a adição de um adjetivo a democracia, ou seja, subtipos democráticos derivados, simplesmente anexando um adjetivo ao substantivo. Por exemplo, com base no julgamento que a democracia em países como Chile, depois de 1990 era limitada pela persistente influência dos militares nos processos políticos, pesquisadores criaram os subtipos democracia cerceada ou tutelada. Onde as liberdades civis eram consideravelmente frágeis, pesquisadores criaram a categoria de democracia iliberal (COLLIER E LEVITSKY, 1997, p. 440). Estes subtipos democráticos, em efeito, permite aos pesquisadores criar um degrau em uma escala ordinal de democracia, nos quais casos ambíguos são situados em um relacionamento menos do que vis-a-vis com as concepções dos analistas de completa democracia.

Inovações relacionadas, também, tentam evitar uma classificação inválida de casos –e assim, mais uma vez, tentam evitar uma alargamento conceitual- envolvendo mudanças no que pode ser chamado de conceito fundamental de democracia, isto é, formas mais amplas de instituições políticas, das quais a democracia é uma instância específica. Democracia, pode tipicamente ser pensada como um tipo de regime, mas há muitas variações para este conceito. Por exemplo, dado a aparente fragilidade do regime democrático no Brasil no final dos anos 80, pesquisadores costumavam usar os termos situação democrática, ou governo democrático, sugerindo um baixo nível de institucionalização, comparado com o rótulo de regime democrático. Em contraste, um outro pesquisador concluiu que embora o Brasil tivesse um regime democrático, a proteção aos cidadãos e os direitos civis eram tão fracos que o país era caracterizado como um, ainda, não existente estado democrático (O'DONNELL,

1996, p. 447). Em todos estes exemplos citados, as categorias usadas na classificação dos casos foram adaptadas para incorporar as percepções tiradas de casos conhecidos em profundidade.

Críticas do Modelo Baseado em Estudo de Casos por Metodologistas

A crítica fundamental a esse modelo centra-se no preço que o estudo de casos precisa pagar no trade off entre a riqueza de detalhes advinda do conhecimento do caso e a alavancagem de uma medição precisa através de um trabalho sistemático e cuidadoso, com um N maior. Certamente se faz necessário evitar a fácil conclusão de que maior é geralmente e automaticamente alcançado com um N grande, devido as complicadas questões de específico contexto que aparecem em qualquer processo de medida. No entanto, uma análise que é feita a partir de uma N grande, certamente abre a possibilidade para se conseguir uma medida que é válida e também mais facilmente generalizada.

Para pesquisadores que usam métodos quantitativos, os estudos baseados em casos com um pequeno N tem também a desvantagem de que as suas ferramentas padrão de inferência causal simplesmente não podem ser aplicadas. O principal foco dos pesquisadores no modelo de estudo de casos está, normalmente, sobre o pequeno número de casos examinados para melhorar as medidas, e as esperadas melhorias são, presumivelmente, de grande valia para eles. O mesmo, não se pode dizer sobre os analistas que se utilizam de grandes N.

No entanto, é preciso se enfatizar que com recursos suficientes, uma pesquisa focada em um grande N pode estar combinada com um intenso e profundo exame de casos específicos. Quaisquer que sejam as outras críticas aos *surveys* feitos pela Freedom House, eles usam uma grande equipe de pesquisadores para elaborar as suas classificações. Em 2011, isto envolveu um incrível número de 70 analistas, os quais trouxeram para a tarefa um considerável conhecimento sobre os países²¹. Assim, o estudo intenso e detalhado sobre um determinado país não precisa estar restrito a um estudo de caso com um pequeno N.

Conclusões

Os pesquisadores adotam diversas abordagens no processo de validar medidas. Este artigo dá a estrutura a esta diversidade explorando quatro modelos contrastantes.

²¹ Disponível em <http://www.freedomhouse.org/report/freedom-world-2011/survey-team>. Visualizado em 31 de março de 2013.

Uma vez que os analistas tenham reconhecido estas alternativas, como eles deveriam abordar as escolhas sobre o processo de validação de suas medidas?

Parte da resposta, a partir do ponto de vista do metodologista em cada modelo, é que alguns dos outros modelos estão simplesmente no caminho errado. Por exemplo, autores do modelo pragmático podem estar convencidos de que o LoM tem desperdiçado décadas em uma tarefa não produtiva. Autores de vários outros modelos se preocupam com o fato de os pragmáticos confiarem excessivamente em se a medida serve para uma específica aplicação, e portanto não são cuidadosos o suficiente para estabelecer se uma medida reflete o conceito que está sendo aplicado e os detalhes dos casos específicos.

Para os metodologistas do modelo de estudo de casos, e potencialmente também os do modelo pragmático e LoM, os pressupostos do modelo SEM-L podem parecer inverossímeis, e esses pesquisadores talvez questionem se o grande investimento em especialização técnica traz necessariamente resultados compatíveis. Os autores do modelo baseado em estudo de casos podem, da mesma forma, estar convencidos que pesquisadores em outros modelos rotineiramente desperdiçam tempo trabalhando com grandes grupos de dados, para os quais é impossível se obter suficiente conhecimento de caso para se alcançar uma medição que faça sentido. Por outro lado, metodologistas comprometidos com o LoM e SEM-L podem achar que o modelo baseado em estudo de casos tende a ser *ad hoc*, e tem uma idiossincrática, ao invés de sistemática abordagem para lidar com os erros, além de falhar rotineiramente no processo de generalização dos casos. Assim, até um certo grau, os autores de cada modelo veem alguns dos outros modelos como fundamentalmente falhos.

A partir de uma menos cética perspectiva, os objetivos que norteiam cada modelo podem ser vistos como componentes de normas gerais razoáveis – e amplamente aceitas – que autores de qualquer abordagem deveriam usar para o estabelecimento de medições válidas. Em um mundo ideal, uma medição de alta qualidade deve ter um relativamente baixo índice de erros, um convincente argumento para o ordenamento de suas categorias, espaçamento entre suas unidades de medida, relevância em alcançar os mais importantes objetivos da pesquisa, e uma estreita correspondência com os detalhes empíricos dos casos estudados. Uma visão que, em geral, aceita estas deferentes abordagens tem certamente seus méritos garantidos.

Algumas ferramentas de validação relacionam estes modelos entre si. Por exemplo, o modelo Item-response usado por Treir e Jackman (2008) trata os objetivos do LoM estimando a distância entre categorias em uma escala ordinal, assim até o nível intervalo. Além disso, esses modelos também abordam o objetivo em

SEM-L de estimar a quantidade de medição em vários indicadores. Da mesma forma, a análise baseada em estudo de casos de Bowman, Lehoucq e Mahoney (2005), fala diretamente às preocupações da abordagem de equações estruturais através da identificação de substanciais erros de medição em indicadores de casos cruzados de democracia. Onde um modelo estatístico pode apontar a existência de erros, a abordagem baseada em estudo de casos identifica erros específicos. A aparente prevalência de erros nas medições de democracia na América Central é, desta forma, um resultado congruente com, e ilumina, alguns resultados derivados do modelo de equações estruturais. Análises baseadas em estudo de casos que prestam muita atenção na ordem e na relativa distância entre os casos, podem também ser um suplemento útil para versões mais padronizadas do LoM.

Todas as quatro abordagens são colocadas juntas por Bollen e Jackman (1989; ver também BOLLEN, 1990). Estes autores apresentam um argumento empírico, mais do que teórico, que os pesquisadores deveriam usar medidas contínuas, as quais excluem o conceito de estabilidade de regime. Primeiro, eles apontam que as medidas dicotômicas não cumprem o critério fundamental da medição de nível nominal: um apropriado grau de igualdade entre os casos com classificações similares (BOLLEN e JACKMAN, 1989). Eles então elaboram modelos estatísticos para testar a contribuição da experiência democrática, em oposição a estabilidade do regime e erros de medição para cada par de indicadores concorrentes. A seguir eles mostram que o seu indicador preferido desempenha um melhor papel em reafirmar a hipótese familiar que diz que as democracias são menos economicamente desiguais do que regimes autoritários, usando então, o que temos chamado de validação nomológica. Finalmente, eles adentram por uma análise detalhada de casos para argumentar que os indicadores dicotômicos invariavelmente violentam a história de um importante conjunto de países.

Através do estabelecimento de pontes metodológicas deste tipo, analistas podem certamente fazer incursões por múltiplos modelos para avaliar e melhorar medidas de democracia. Em geral, os pesquisadores devem reconhecer os debates que ocorrem sobre as fraquezas de seus métodos preferidos de validação e explorar a contribuição de métodos que utilizam várias abordagens ao mesmo tempo para minimizar estas fraquezas.

■ Jason Seawright é Professor Associado da Northwestern University. E-mail: j-seawright@northwestern.edu.

■ David Collier é Professor da University of California, Berkeley. E-mail: dcollier@berkeley.edu.

Referências

- ABELSON, Robert P.; TUKEY, John W. Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics*, v. 34, n. 4, p. 1347-1369, 1963.
- ADCOCK, Robert; COLLIER, David. Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, v. 95, n. 3, p. 529-546, set. 2001.
- ALDRICH, John H.; NELSON, Forrest D. *Linear probability, logit, and probit models*. Beverly Hills, USA: Sage, 1984.
- ANDRICH, David. *Rasch models for measurement*. Newbury Park, USA: Sage, 1988.
- BAKER, Bela O.; HARDYCK, Curtis D.; PETRINOVICH, Lewis F. Weak measurements vs. strong statistics: An empirical critique of S.S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*, v. 26, n. 2, p. 291-309, jul. 1966.
- BAKER, Peter J.; KOESEL, Karrie J. Measuring "polyarchy plus": Tracking the quality of democratization in Eastern Europe. *American Political Science Association, Annual Meeting*, San Francisco, USA, 2001.
- BANKS, Arthur S. *Cross-polity time-series data*. Cambridge, USA: MIT Press, 1971.
- BANKS, Arthur S. *Cross-polity time-series data archive user's manual*. Binghamton: State University of New York at Binghamton, 1979.
- BERG-SCHLOSSER, Dirk. Indicators of democracy and good governance as measures of the quality of democracy in Africa: A critical appraisal. *Acta Politica*, v. 39, n. 3, p. 248-278, set. 2004.
- BLALOCK, Hubert. *Conceptualization and measurement in the social sciences*. Beverly Hills, USA: Sage, 1982.
- BOLLEN, Kenneth A. *Structural equations with latent variables*. New York, USA: John Wiley, 1989.
- BOLLEN, Kenneth A. Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development*, v. 25, n. 1, p. 7-24, Spring 1990.
- BOLLEN, Kenneth A. Liberal democracy: Validity and method factors in cross-national measures. *American Journal of Political Science*, v. 37, n. 4, p. 1207-1230, nov. 1993.
- BOLLEN, Kenneth A.; JACKMAN, Robert. Democracy, stability, and dichotomies. *American Sociological Review*, v. 54, n. 4, p. 612-621, ago. 1989.
- BOLLEN, Kenneth A.; PAXTON, Pamela. Subjective measures of liberal democracy. *Comparative Political Studies*, v. 33, n. 1, p. 58-86, fev. 2000.

- BOLLEN, Kenneth A.; RABE-HESKETH, Sophia; SKRONDAL, Anders. Structural equation models. In: BOX-STEFFENSMEIER, Janet M.; BRADY, Henry E.; COLLIER, David (Eds.). *The Oxford handbook of political science*. New York, NY: Oxford University Press, 2008. p. 432-455.
- BOND, Trevor G.; FOX, Christine M. *Applying the Rasch model: Fundamental measurement in the human science*. New York, USA: Routledge, 2012.
- BOWMAN, Kirk; LEHOUCQ, Fabrice; MAHONEY, James Measuring political democracy: Case expertise, data adequacy, and Central America. *Comparative Political Studies*, v. 38, n. 8, p. 939-970, out. 2005.
- BRADY, Henry E. Statistical consistency and hypothesis testing for nonmetric multidimensional scaling. *Psychometrika*, v. 50, n. 4, 509-537, dez. 1985.
- CAMPBELL, Norman Robert. (1920). Physical number. In: CAMPBELL, Norman Robert (Ed.). *Physics: The elements*. Cambridge, USA: Cambridge University Press, 1920. p. 295-309.
- CASPER, Gretchen; TUFIS, Carliu. Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated datasets. *Political Analysis*, v. 11, n. 2, p. 196-203, 2003.
- CHEIBUB, José Antonio; GHANDI, Jennifer; VREELAND, James Raymond. Democracy and dictatorship revisited. *Public Choice*, v. 143, n. 1-2, p. 67-101, 2010.
- CLIFF, Norman. Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, v. 18, n. 1, p. 115-126, jan. 1983.
- COLLIER, David. Data, field Work, and extracting new ideas at close range. *American Political Science Association*, Newsletter for the Organized Section in Comparative Politics, v. 10, n. 1, p. 1-2, 4-6, Winter 1999.
- COLLIER, David; ADCOCK, Robert. Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science*, v. 2, p. 537-565, 1999.
- COLLIER, David; LAPORTE, Jody; SEAWRIGHT, Jason. Putting typologies to work. *Political Research Quarterly*, v. 65, n. 1, p. 217-232, mar. 2012.
- COLLIER, David; LEVITSKY, Steven. Democracy with adjectives: Conceptual innovation in comparative research. *World Politics*, v. 49, n. 3, p. 430-451, abr. 1997.
- COLLIER, Ruth Berins. *Paths toward democracy: The working class and elites in Western Europe and South America*. Cambridge, USA: Cambridge University Press, 1999.
- COPPEDGE, Michael. Thickening thin concepts and theories: Combining large N and small N in comparative politics. *Comparative Politics*, v. 31, n. 4, p. 465-476, jul. 1999.
- COPPEDGE, Michael. Varieties of democracy project. *Kellogg Institute*, University of Notre Dame, 2010. Disponível em: <<http://kellogg.nd.edu/varieties.pdf>>.
- COPPEDGE, Michael; GERRING, John. Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics*, v. 9, n. 2, p. 247-267, jun. 2011.
- COPPEDGE, Michael; REINICKE, Wolfgang H. Measuring polyarchy. *Studies in Comparative International Development*, v. 25, n. 1, p. 51-72, Spring 1990.
- DAHL, Robert Alan. *Polyarchy: Participation and opposition*. New Haven, USA: Yale University Press, 1971.
- DUNCAN, Otis Dudley. *Notes on social management: Historical and critical*. New York, USA: Russell Sage. 1984.

- ELKINS, Zachary. Gradations of democracy? Empirical tests of alternative conceptualizations. *American Journal of Political Science*, v. 44, n. 2, p. 287-294, abr. 2000.
- ENGELHARD, George. Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement: Interdisciplinary Research and Perspectives*, v. 6, n. 3, p. 155-189, 2008.
- FABRIGAR, Leandre R.; WEGENER, Duane T. *Exploratory factor analysis*. Oxford, UK: Oxford University Press, 2011.
- FISCHER, Gerhard H.; MOLENAAR, Ivo W. (Eds.). *Rasch models: Foundations, recent developments, and applications*. New York, USA: Springer-Verlag, 1995.
- POWERAKER, Joe; KRZANARIC, Roman. Measuring liberal democratic performance: An empirical and conceptual critique. *Political Studies*, v. 48, n. 4, p. 759-787, set. 2000.
- FREEDMAN, David A. As others see us: A case study in path analysis. *Journal of Educational and Behavioral Statistics*, v. 12, n. 2, p. 101-128, Summer 1987.
- GALLIE, Walter Bryce. Essentially contested concepts. *Proceedings of the Aristotelian Society*, v. 56, p. 167-198, 1956.
- GEORGE, Alexander L.; BENNETT, Andrew. *Case studies and theory development in the social sciences*. Cambridge, USA: MIT Press, 2005.
- GILL, Jeff. *Essential mathematics for political and social research*. Cambridge, USA: Cambridge University Press, 2006.
- GRANT, Tobin. Unifying political metrology: A probabilistic model of measurement. *Society for Political Methodology*, Annual Meeting, Palo Alto, USA, jul. 2004.
- GREENE, William H. *Econometric analysis*. Upper Saddle River, USA: Prentice Hall, 2000.
- GUTTMAN, Louis. The basis for scalogram analysis. In: STOUFFER, Samuel et al. (Eds.). *Studies in social psychology in World War II: Vol. IV. Measurement and prediction*. Princeton, USA: Princeton University Press, 1950. p. 60-90.
- HARDY, Melissa A. *Regression with dummy variables*. Newbury Park, USA: Sage, 1993.
- HAYDUK, Leslie A. *Structural equation modeling with LISREL: Essential and advances*. Baltimore, USA: Johns Hopkins University Press, 1987.
- JACOBY, William G. Levels of measurement and political research: An optimistic view. *American Journal of Political Science*, v. 43, n. 1, p. 271-301, jan. 1999.
- KAPLAN, David W. *Structural equation modeling: Foundations and extensions*. Thousand Oaks, USA: Sage, 2008.
- KARIYA, Komyo; FINKELSTEIN, Ludwik. *A discussion: Measurement science*. Amsterdam, Netherlands: IOS Press, 2000.
- KIM, Jae-On; MUELLER, Charles W. *Factor analysis: Statistical methods and practical issues*. Beverly Hills, USA: Sage, 1978a.
- KIM, Jae-On; MUELLER, Charles W. *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, USA: Sage, 1978b.
- KRANTZ, David H. et al. *Foundations of measurement, Volume 1: Additive and polynomial representations*. New York, USA: Academic Press, 1971.
- KRUSKAL, Joseph B.; WISH, Myron. *Multidimensional scaling*. Beverly Hills, USA: Sage, 1978.
- LOCKE, Richard M.; THELEN, Kathleen. Apples and oranges revisited: Contextualized comparisons and the study of comparative labor politics. *Politics & Society*, v. 23, n. 3, p. 337-367, 1995.
- LORD, Frederic. On the statistical treatment of football numbers. *American Psychologist*, v. 8, n. 12, p. 750-751, dez. 1953.

- LUCE, R. Duncan. On the possible psychophysical laws. *Psychological Review*, v. 66, n. 2, p. 81-95, mar. 1959.
- MAHONEY, James. *The legacies of liberalism: Path dependence and political regimes in Central America*. Baltimore, USA: Johns Hopkins University Press, 2001.
- MAHONEY, James; RUESCHEMEYER, Dietrich. *Comparative historical analysis in the social sciences*. Cambridge, USA: Cambridge University Press, 2003.
- MAINWARING, Scott; BRINKS, Daniel; PÉREZ-LIÑÁN, Aníbal. Classifying political regimes in Latin America, 1945–1999. *Studies in Comparative International Development*, v. 36, n. 1, p. 37–65, 2001.
- MARSHALL, Monty G. *Polity IV project: Political regime characteristics and transitions, 1800–2011*. 2013. Disponível em: <<http://www.systemicpeace.org/polity/polity4.htm>>.
- MICHELL, Joel. Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, v. 6, n. 1-2, p. 7-24, 2008.
- MUELLER, Ralph. *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York, USA: Springer, 1996.
- MUNCK, Gerardo L. *Measuring democracy: A bridge between scholarship and politics*. Baltimore, USA: Johns Hopkins University Press, 2009.
- MUNCK, Gerardo L.; VERKUILEN, Jay. Conceptualizing and measuring democracy: Evaluating alternative indices. *Comparative Political Studies*, v. 35, n. 1, p. 5-34, fev. 2002.
- O'DONNELL, Guillermo. Illusions about consolidation. *Journal of Democracy*, v. 7, n. 2, p. 34-51, 1996.
- O'DONNELL, Guillermo; SCHMITTER, Philippe C. *Transitions from authoritarian rule: Tentative conclusions about uncertain democracies*. Baltimore, USA: Johns Hopkins University Press, 1986.
- PAMPEL, Fred C. *Logistic regression: A primer*. Thousand Oaks, USA: Sage, 2000.
- PEMSTEIN, Dan; MESERVE, Stephen A.; MELTON, James. Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis*, v. 18, n. 4, p. 426-449, 2010.
- PRZEWORSKI, Adam et al. *Democracy and development: Political institutions and well-being in the World, 1950–1990*. Cambridge, USA: Cambridge University Press, 2000.
- RECKASE, Mark D. *Multidimensional item response theory: Statistics for social and behavioral sciences*. New York, USA: Springer, 2009.
- RUESCHEMEYER, Dietrich; STEVENS, Evelyn Huber; STEVENS, John D. *Capitalist development and democracy*. Chicago, USA: University of Chicago Press, 1992.
- SARTORI, Giovanni. Concept misformation in comparative politics. *American Political Science Review*, v. 64, n. 4, p. 1033-1053, nov. 1970.
- SARTORI, Giovanni. Democracy: What is vs. how much. In: COLLIER, David; GERRING, John (Eds.). *Concepts and method in social science*. New York, USA: Routledge, 2009. p. 165-167.
- SCHUMPETER, Joseph A. *Capitalism, socialism, and democracy*. New York, USA: Harper & Row, 1950.
- SHEN, Ce; WILLIAMSON, John B. Corruption, democracy, economic freedom, and state strength: A cross-national analysis. *International Journal of Comparative Sociology*, v. 46, n. 4, p. 327-345, 2005.
- STEVENS, Stanley Smith. On the theory of scales of measurement. *Science*, v. 103, n. 2684, 677-680, jun. 1946.

- STEVENS, Stanley Smith. Mathematics, measurement, and psychophysics. In: STEVENS, Stanley Smith (Ed.). *Handbook of experimental psychology*. New York, USA: John Wiley, 1951. p. 1-49.
- STEVENS, Stanley Smith. *Psychophysics: Introduction to its perceptual, neural and social prospects*. New York, USA: John Wiley, 1975.
- SUPPES, Patrick et al. *Foundations of Measurement, Volumes II and III*. San Diego, USA: Academic Press, 1989.
- TAKANE, Yoshio; DE LEEUW, Jan. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, v. 52, n. 3, p. 393-408, set. 1987.
- TREIER, Shaw; JACKMAN, Simon. Democracy as a latent variable. *American Journal of Political Science*, v. 52, n. 1, p. 201-217, jan. 2008.
- TUKEY, John W. Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments. In: JONES, Lyle V. (Ed.). *The collected works of John W. Tukey: Philosophy and principles of data analysis 1949-1964*. London, England: Chapman & Hall, 1986. p. 187-208.
- VELLEMAN, Paul F.; WILKINSON, Leland. Nominal, ordinal, interval and ratio typologies are misleading. *American Statistician*, v. 47, n. 1, p. 65-72, 1993.
- YASHAR, Deborah J. *Demanding democracy: Reform and reaction in Costa Rica and Guatemala, 1870s-1950s*. Cambridge, USA: Cambridge University Press, 1997.
- YOUNG, Forrest. Quantitative analysis of qualitative data. *Psychometrika*, v. 46, n. 4, p. 357-388, dez. 1981.

Publicado originalmente como:

SEAWRIGHT, Jason; COLLIER, David. Rival Strategies of Validation: Tools for Evaluating Measures of Democracy. *Comparative Political Studies*, v. 47, n. 1, p. 111-138. 2014

*Texto recebido em 09 de março de 2015.
Aprovado em 06 de abril de 2015.*

