
Conceptualizando y midiendo la democracia: Una evaluación de índices alternativos

GERARDO L. MUNCK Y JAY VERKUILEN*

El estudio de la democracia —una preocupación central de la política comparativa y las relaciones internacionales— ha usado crecientemente métodos estadísticos de inferencia causal sofisticados. Esta tendencia es positiva y las contribuciones de esta literatura cuantitativa son importantes. Sin embargo, con unas cuantas excepciones notables,¹ los investigadores cuantitativos han prestado poca atención a la calidad de los datos sobre la democracia que analizan. De hecho, las evaluaciones que se han llevado a cabo en general se restringen a discusiones bastante informales sobre diferentes bases de datos y a exámenes un tanto superficiales de correlaciones entre datos agregados.² En

* Gerardo L. Munck es Profesor del Departamento de Ciencia Política de la Universidad de Illinois en Urbana-Champaign, 361 Lincoln Hall, 702 South Wright St., Urbana, Ill., 61801, EE. Correo electrónico: g-munck@uiuc.edu.

Jay Verkuilen es estudiante de doctorado en la misma institución. Correo electrónico: jayv@ntx1.cso.uiuc.edu.

Queremos agradecer a Chris Achen, David Collier, Michael Coppedge, James Kuklinski, Mark Lichbach, James Mahoney, Scott Mainwaring, Sebastián Mazzuca, Aníbal Pérez-Liñán, Robert Pahre, Cindy Skach, Richard Snyder, y dos lectores anónimos por sus detallados y útiles comentarios.

El manuscrito se recibió en noviembre de 2001 y fue aceptado para su publicación en febrero de 2002. Traducción del inglés de Susana Moreno Parada.

¹ Bollen (1980, 1986, 1991, 1993); Bollen y Paxton (2000), y Foweraker y Krznaric (2000). Véanse también Gleditsch y Ward (1997) y Coppedge (1999).

² Para discusiones de índices alternativos sobre la democracia y correlaciones entre datos agregados, véanse Álvarez *et al.* (1996, pp. 18-21), Arat (1991, pp. 22, 23 y 28), Bollen (1980, p. 381), Coppedge (1997, p. 180), Coppedge y Reinicke (1991, pp. 51-52), Gasiorowski (1996, pp. 477-478), Hadenius (1992, pp. 41, 43, 71 y 159-163), Jagggers y Gurr (1995, pp. 473-476), y Vanhanen (1993, pp. 317-319; 1997, pp. 31-40).

buena medida, los problemas de la inferencia causal han eclipsado los problemas igualmente importantes de la conceptualización y la medición.

Intentando rectificar este descuido, este artículo ofrece una evaluación sistemática de las bases de datos sobre la democracia que son utilizadas con mayor frecuencia en la investigación estadística actual.³ Un primer paso en esa dirección es provisto por el *cuadro 1*, que compara estas bases de datos en términos de su alcance empírico. Ésta no es una cuestión banal. En realidad, la frecuente delimitación de bases de datos al periodo posterior a la II Guerra Mundial y la exclusión de ciertas regiones del mundo restringe las teorías que pueden probarse con estos datos. Sin embargo, una comparación y evaluación exhaustiva de estas bases de datos debe ir más allá de su alcance empírico y abordar una amplia gama de cuestiones metodológicas. Este hecho complica cualquier esfuerzo por evaluar los datos sobre la democracia.

El problema fundamental radica en que las cuestiones metodológicas relevantes para la generación de datos y que tienen un efecto directo sobre la calidad de los datos acerca de la democracia han sido abordadas sólo parcialmente en la literatura metodológica. Si bien esta literatura ofrece algunas pistas importantes respecto a la conceptualización y la medición, también tiene algunas lagunas considerables. Además, si bien la generación de datos es afectada por decisiones concernientes a un número considerable de asuntos interrelacionados, se ha hecho poco por ofrecer un enfoque integrado que muestre la manera en que esos asuntos están conectados. Así, tanto para hacer explícitos como para justificar los criterios que usamos para evaluar los diferentes índices sobre la democracia, este artículo aborda la tarea netamente metodológica de construir un esquema comprensivo e integrado para el análisis de datos.

El esquema de trabajo que proponemos, resumido en el *cuadro 2* y desarrollado a lo largo del artículo, distingue entre tres *retos* que se abordan sucesivamente: la conceptualización, la medición y la agregación. Asimismo, identifica las *tareas* específicas que los analistas confrontan al abordar esos retos y los *estándares de evaluación* que le corresponden a cada una. Como pre-

³ Para una breve pero útil discusión de algunos índices anteriores que han caído en desuso, véase Bollen (1980, pp. 373-375 y 379-384) y Arat (1991, p. 28).

CUADRO 1: BASES DE DATOS SOBRE DEMOCRACIA:
ALCANCE EMPÍRICO

Nombre ^a	Unidad I: País	Unidad II: Año ^b
ACLP: Álvarez, Cheibub, Limongi y Przeworski (1996, pp. 23-30)	141	1950-1990
Arat (1991, pp. 136-166)	152	1948-1982
Bollen (1980, pp. 387-388; 1991, pp. 16-19; 1993, pp. 1227)	113	1960
	123	1965
	153	1980
Coppedge y Reinicke (1991, pp. 63-66)	170	1985
Freedom House (2000)	Todo el mundo (el número varía)	1972-a la fecha
Gasiorowski (1996, pp. 480-482)	97	Independencia-1992 ^c
Hadenius (1992, pp. 61-69)	132	1988
Polity IV (Marshall y Jaggers, 2001b)	161	1800-1999
Vanhanen (2000b)	187	1810-1998

^a Las citas ofrecidas en este cuadro contienen los datos actuales.

^b Estos datos utilizan países como su unidad de análisis y registran un valor por año. Por lo tanto, aunque desagregamos estos dos aspectos, las unidades de análisis son en realidad país-año.

^c La mayoría de las bases de datos comienzan a clasificar países a partir de un año común, incluyendo casos nuevos a medida que países obtuvieron su independencia. Gasiorowski es una excepción, pues comienza a clasificar países no a partir de un año común sino más bien desde el momento en que logran la independencia. Por lo tanto, su punto de partida varía mucho, de 1747 a 1980.

CUADRO 2: ESQUEMA PARA EL ANÁLISIS DE LA CONCEPTUALIZACIÓN, LA MEDICIÓN Y LA AGREGACIÓN DE DATOS

<i>Reto</i>	<i>Tarea</i>	<i>Estándar de evaluación</i>
Conceptualización	Identificación de atributos.	Especificación del concepto: evitar definiciones maximalistas (la inclusión de atributos teóricamente irrelevantes) o definiciones minimalistas (la exclusión de atributos teóricamente relevantes).
	Organización vertical de los atributos según el nivel de abstracción.	Lógica conceptual: aislar las hojas del árbol conceptual y evitar los problemas de redundancia y confluencia.
Medición	Selección de indicadores.	Validez: usar indicadores múltiples y establecer la equivalencia de dichos indicadores por medio de distintas unidades; usar indicadores que minimicen el error de medición y que puedan ser verificados a través de múltiples fuentes. Confiabilidad.
	Selección del nivel de medición.	Validez: maximizar la homogeneidad dentro de las clases de medición con el número mínimo de distinciones necesarias. Confiabilidad.
	Registración y divulgación de reglas de medición, proceso de medición y datos desagregados.	Replicabilidad.
Agregación	Selección del nivel de agregación.	Validez: balancear la búsqueda de parsimonia con la preocupación por la dimensionalidad subyacente y la diferenciación.
	Selección de la regla de agregación.	Validez: asegurar la correspondencia entre la teoría de la relación entre atributos y la regla de agregación elegida. Robustez de los datos agregados.
	Registración y divulgación de reglas de agregación y datos agregados.	Replicabilidad.

tendemos mostrar, este esquema constituye tanto una contribución significativa a la literatura metodológica como una manera útil de estructurar nuestra evaluación de las bases de datos sobre la democracia.

La organización de este artículo sigue directamente nuestro esquema. La primera sección aborda el reto de la conceptualización, la segunda se aboca al reto de la medición y la tercera al reto de la agregación. En cada sección, primero elaboramos el esquema que proponemos y presentamos las pautas metodológicas clave que los analistas deben considerar. Luego evaluamos las bases de datos sobre la democracia a la luz de estas pautas. En la última sección, ofrecemos una evaluación global de estas bases de datos.

EL RETO DE LA CONCEPTUALIZACIÓN: ATRIBUTOS Y ORGANIZACIÓN LÓGICA

El reto inicial en la construcción de una base de datos es la *identificación de los atributos* que son constitutivos del concepto en cuestión. Esta tarea, por la cual se especifica el significado del concepto, afecta a todo el proceso de generación de datos, pues es el punto de referencia de todas las decisiones siguientes. De este modo, un impulso natural y comprensible es encontrar criterios objetivos y fijos que sirvan de guía a esta tarea. Sin embargo, no existe regla indiscutida alguna que pueda usarse para decidir cuáles atributos deben incluirse en la definición de un concepto determinado. De hecho, puesto que la conceptualización está íntimamente vinculada con la teoría y es una actividad abierta, en evolución que en última instancia es evaluada en términos de la utilidad de las teorías que ayuda a formular (Kaplan, 1964, pp. 51-53 y 71-78), “no tiene sentido discutir acerca de cuál es una definición ‘correcta’” (Guttman, 1994, p. 12; véase también p. 295). En efecto, afirmaciones que alegan que los debates sobre cómo especificar un concepto pueden ser resueltos son inherentemente sospechosas, y la sugerencia metodológica más útil —si bien flexible— que puede ofrecerse es que los estudiosos deberían evitar los extremos, ya sea al incluir demasiado o demasiado poco en una definición en relación con sus objetivos teóricos.

La tendencia a especificar el significado de un concepto de modo que incluya demasiados atributos —el problema de las *definiciones maximalistas*— tiene dos inconvenientes potenciales. Por un lado, sobrecargar un concepto puede reducir su utilidad al convertirlo en un concepto que no tenga referentes empíricos. Un ejemplo es la inclusión de la noción de justicia social como atributo de la democracia. Por otro lado, aun si se define un concepto de modo que puedan encontrarse ejemplos empíricos, las definiciones maximalistas suelen ser tan sobrecargadas que tienen poca utilidad analítica. Por ejemplo, si se considera que el sistema económico de mercado es un atributo de la democracia, el vínculo entre mercados y democracia no es abordable como una cuestión empírica. El problema con tales definiciones, como afirman Álvarez, Cheibub, Limongi y Przeworski (1996, pp. 18 y 20), es que no permiten el análisis de preguntas que pueden ser “demasiado interesantes para ser resueltas por medio de una decisión acerca de definiciones”.

El esfuerzo por evitar el problema de las definiciones maximalistas suele llevar a la propuesta de definiciones minimalistas, que tienen la ventaja obvia de que es fácil encontrar ejemplos de un concepto y de que permiten el estudio de numerosas cuestiones empíricas. Sin embargo, el minimalismo también tiene sus problemas. De hecho, si un concepto es tan minimalista que todos los casos automáticamente se convierten en instancias de este concepto, los investigadores deben agregar atributos al concepto para darle más contenido y, así, abordar mejor los asuntos teóricos relevantes y discriminar mejor entre los casos. Entonces, además de evitar el problema de las definiciones maximalistas, los analistas deben ser conscientes del problema de las *definiciones minimalistas*, la omisión de un atributo relevante en la definición de un concepto.

Los índices sobre la democracia han abordado este primer paso en la construcción de un índice —la identificación de los atributos— con considerable agudeza. De hecho, la decisión de hacer uso, si bien en diferente grado, de la influyente idea de Dahl (1971, pp. 4-6) de que la democracia consiste de dos atributos —contestación o competencia y participación o inclusión— ha contribuido a asegurar que esas mediciones de democracia estén claramente enfocados en atributos teóricamente relevantes. Mas, a pesar de estos aspectos po-

sitivos, un estudio sistemático de los atributos utilizados en los índices sobre la democracia (véase el cuadro 3) revela que son vulnerables a muchas críticas.

La mayoría de los constructores de índices adoptan una definición de democracia procedimentalista y, de este modo, evitan el problema de las definiciones maximalistas. La única excepción es Freedom House, que restringe severamente la utilidad analítica de su índice debido a la inclusión de atributos tales como “derechos socioeconómicos,” “libertad de desigualdades socioeconómicas brutas,” “derechos de propiedad” y “libertad de guerra” (Gastil, 1991, pp. 32-33, y Ryan, 1994, pp. 10-11), los cuales son más provechosamente entendidos como atributos de algún otro concepto. En cambio, el problema de las definiciones minimalistas está muy generalizado.

Una omisión importante que afecta a varios índices se refiere a uno de los atributos que Dahl destaca: la participación. Esta omisión es un problema particularmente grave para el índice Polity creado por Gurr y sus colegas (Marshall y Jaggers, 2001a). De hecho, puesto que esta base de datos se remonta hasta 1800, esta omisión los lleva a ignorar un rasgo clave de la experiencia con la democratización en el siglo XIX y principios del siglo XX en comparación con los fines del siglo XX: la expansión gradual del derecho al voto. En contraste, este descuido es menos significativo en los casos de los índices propuestos por Álvarez, Cheibub, Limongi y Przeworski o ACLP (1996) y Coppédge y Reinicke (1991). De hecho, la justificación que estos autores ofrecen —que a ellos les preocupa juntar datos sólo para el periodo posterior a la II Guerra Mundial, que puede suponerse la existencia del sufragio universal a partir de 1945 y, por tanto, que la competencia es el aspecto más importante del proceso electoral— es bastante razonable (Álvarez *et al.*, 1996, pp. 5 y 19; Coppédge y Reinicke, 1991, p. 51 y Coppédge, 1997, p. 181). No obstante, la exclusión del atributo de la participación sigue siendo problemática.⁴ Si bien restricciones *de jure* al derecho al voto son raras en las democracias actuales, toda

⁴ Otros dos índices omiten este atributo. Si bien Freedom House se refiere al “derecho de todos los adultos al voto” cuando define los derechos políticos, no incluye este aspecto en la lista de derechos políticos (Ryan, 1994, p. 10). Asimismo, Bollen (1980, pp. 372 y 376) subraya la importancia del sufragio universal, pero luego no parece retener este aspecto de las elecciones en sus atributos.

CUADRO 3: BASES DE DATOS SOBRE LA DEMOCRACIA: UN RESUMEN

Nombre del índice	Atributos	Componentes de atributos	Nivel de medición	Regla de agregación
ACLP, Álvarez, Cheibub, Limongi y Przeworski (1996)	Contestación Cargos	[Elección del Ejecutivo Elección del Legislativo	Nominal Nominal Nominal	Multiplicativa, al nivel de componentes y atributos.
Arat (1991)	Participación	[Selección del Ejecutivo Selección del Legislativo Efectividad del Legislativo Competitividad del proceso de nominación	Ordinal Ordinal Ordinal Ordinal	Aditiva, al nivel de componentes; aditiva y multiplicativa combinada, al nivel de atributos.
	Inclusividad Competitividad	[Legitimidad de partidos Competitividad de partidos	Ordinal Ordinal Ordinal	
	Coerción		Intervalar	
Bollen (1980)	Libertades políticas	[Libertad de Prensa Libertad de oposición Sancciones del gobierno	Intervalar Intervalar Intervalar Intervalar	Medidas de factores (promedios ponderados).
	Soberanía popular	[Imparcialidad electoral Selección del Ejecutivo Selección y efectividad del legislativo	Intervalar Intervalar Intervalar Intervalar	
Coppedge y Reinicke (1991)	Contestación	[Elecciones libres e imparciales Libertad de asociación Libertad de expresión Pluralismo en los medios de comunicación	Ordinal Ordinal Ordinal Ordinal	Escala Guttman (jerárquica), al nivel de componentes.
Freedom House (Ryan, 1994)	Derechos políticos Derechos civiles	[9 componentes 13 componentes*	Ordinal Ordinal	Aditiva, al nivel de componentes.

CUADRO 3. Continuación...

Nombre del índice	Atributos	Componentes de atributos	Nivel de medición	Regla de agregación
Gastorowski (1996)	Competitividad Inclusividad Libertades civiles y políticas		Ordinal con categoría residual**	Ninguna.
Hadenius (1992)	Elecciones	[Sufragio Cargos elegidos Elecciones significativas*** [apertura, imparcialidad y efectividad]	Intervalar Intervalar Ordinal	Combined additive and multiplicative (of weighted scores) at the level of components; additive, at the level of attributes.
Polity IV (Marshall y Jaggers, 2001a)	Libertades políticas	[Libertad de asociación Libertad de expresión Libertad de coacción	Ordinal Ordinal Ordinal	Aditiva (con puntuaciones ponderadas).
	Competitividad de la participación. Regulación de la par- ticipación		Ordinal	
	Competitividad del reclutamiento del Ejecutivo.		Ordinal	
	Apertura del reclu- tamiento del Ejecutivo. Limitaciones al poder del Jefe del Ejecutivo		Ordinal	
Vahanen (2000a)	Competición. Participación.		Intervalar Intervalar	Multiplicativa.

* Véase las listas de los componentes que utiliza Freedom House en Gastil (1991, pp. 26 y 32-33) y Ryan (1994, pp. 10-11).
 ** Gastorowski ofrece una definición que desagrega su concepto principal, pero no desarrolla medidas para sus atributos. Por lo tanto, su opción del nivel de medición corresponde a su concepto principal.
 *** Los atributos entre corchetes constituyen un tercer nivel de desagregación y, por tanto, implican "subcomponentes de atributos".

una batería de otras restricciones, por lo general informales, impiden el uso efectivo del derecho formal al voto y distorsionan significativamente el valor de los votos (Elklit, 1994; Hadenius, 1992, p. 40). Por lo tanto, no incluir la participación en sus distintas facetas es un problema incluso para el estudio de la democracia en épocas recientes.⁵

Más allá de este atributo obviamente relevante de la participación o inclusión, son notables otras omisiones importantes. Uno de los aspectos distintivos de la base de datos de ACLP (1996, pp. 4-5) es que incluye un atributo llamado Cargos, que se refiere al grado en que las oficinas públicas son ocupadas por medio de elecciones y no por otros procedimientos. Ésta es una decisión buena. Después de todo, el concepto de democracia parece inextricablemente vinculado a la noción del acceso al poder y —es crucial señalarlo— el conjunto de cargos de gobierno que son ocupados mediante elecciones ha variado independientemente del grado en que las elecciones han sido competitivas e inclusivas (Gehrlich, 1973). Así, la importancia de Cargos como un atributo de la democracia sugiere que los índices que se han inspirado únicamente en Dahl y que incluyeron sólo los atributos de competición y/o participación (Coppedge y Reinicke, Gasiorowski y Vanhanen) han omitido un atributo importante.⁶

Siguiendo este tren de ideas, la sugerencia de que Cargos es un atributo relevante suscita la pregunta sobre la pertinencia de otros atributos que no están vinculados tan estrictamente al proceso electoral. Así, algunos autores han sugerido que simplemente pensar en si los cargos públicos son llenados por medio de elecciones no es suficiente para plantear el verdadero corazón del asunto —¿quién verdaderamente ejerce el poder?— y

⁵ Esos aspectos de la participación a veces son incluidos en los índices bajo la forma del atributo Imparcialidad del proceso electoral. Tal es el caso con Bollen y Hadenius. Incluso Coppedge y Reinicke (1991, p. 49), quienes afirman que a ellos sólo les interesa la competencia, incluyen este aspecto de la participación en el índice que proponen. Sin embargo, la mayoría de los índices no abordan estos importantes aspectos.

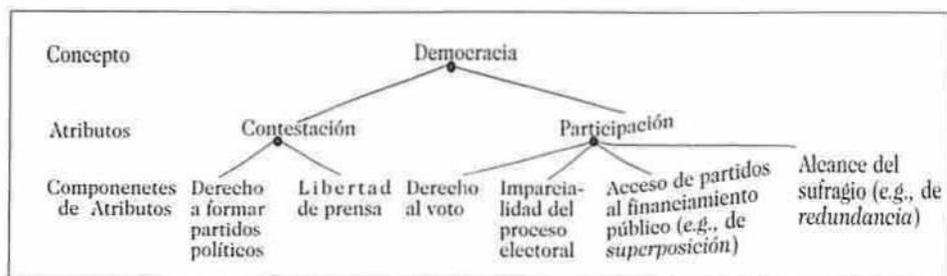
⁶ Otros han incluido un atributo que se asemeja a lo que ACLP quieren decir por Cargos, pero han usado otros nombres. Arat y Bollen se refieren a la Selección del Ejecutivo y del Legislativo. Hadenius habla del número de bancas ocupadas por medio de elecciones. Y el índice Polity IV se refiere un tanto confusamente a la Competitividad y la Apertura del reclutamiento del Ejecutivo.

por tanto han incluido en sus índices otro atributo, llamado Efectividad del Legislativo por Arat y Bollen, Efectividad de las elecciones por Hadenius, y Limitaciones al poder del Jefe del Ejecutivo en Polity IV. Si bien este atributo puede ser difícil de medir,⁷ su relevancia es difícil de disputar. Por lo tanto, los índices que no incluyen este atributo, que por conveniencia puede llamarse “el poder de fijar la agenda de los funcionarios públicos,” padecen de una omisión importante. En suma, el problema de las definiciones minimalistas está muy generalizado en los índices sobre la democracia.

Después del paso inicial de identificar *cuáles* atributos se consideran partes constitutivas de un concepto, los analistas deben también considerar *cómo* se relacionan entre sí estos atributos y, más específicamente, tomar pasos explícitos para garantizar la *organización vertical de los atributos según el nivel de abstracción*. Aunque es raramente abordado en las discusiones habituales sobre metodología, esta tarea tiene un impacto en la generación de datos pues afecta los dos retos siguientes de medición y agregación. Primero, la especificación del significado de un concepto frecuentemente implica la identificación de atributos que varían en términos de su nivel de abstracción. Por lo tanto, en la medida en que esos atributos comienzan a formar un puente entre el nivel abstracto en que se formulan, inicialmente, los conceptos y el nivel concreto de las observaciones, la identificación de los atributos conceptuales afecta y puede ayudar a los analistas a responder al siguiente y distinto reto de la medición. Sin embargo, para lograr este beneficio los diversos atributos deben estar organizados verticalmente según sus niveles de abstracción. De hecho, es al distinguir los atributos según sus niveles de abstracción —que por conveniencia les damos nombres diferentes (atributos, componentes de atributos, subcomponente de atributos, etc.)— que los analistas aíslan los atributos más concretos, llamados las *hojas del árbol conceptual*, que sirven como punto de partida para los esfuerzos de medición (véase la *figura 1*).

⁷ Álvarez *et al.* (1996, p. 20) justifican la exclusión del atributo de Efectividad del Legislativo sobre la base de que no contamos con datos suficientemente confiables sobre este atributo

FIGURA 1. LA ESTRUCTURA LÓGICA DE LOS CONCEPTOS



NOTA: Este ejemplo tiene dos niveles de abstracción, llamados atributos y componentes de atributos. Se podría introducir un tercer nivel de abstracción, llamado subcomponentes de atributos, y aún más niveles. Sin embargo, sin importar cuántos niveles de abstracción se introduzcan, los atributos del último nivel de abstracción, genéricamente llamados *hojas*, se usan como punto de partida para la tarea de medición. En este ejemplo, el Derecho a formar partidos políticos es una hoja.

Segundo, la identificación de múltiples atributos de un concepto esencialmente consiste de un proceso de desagregación, que inmediatamente plantea la pregunta de cómo los datos desagregados podrían ser agregados. El reto de la agregación sólo puede llevarse a cabo una vez que se han asignado medidas a cada hoja es decir, después de haber abordado el reto de la medición, e implica un conjunto complejo de cuestiones que veremos más adelante. No obstante, cualquier análisis de la agregación presupone que los atributos de un concepto están organizados de una manera que cumple las dos reglas básicas de la lógica conceptual. Por un lado, al organizar los atributos de un concepto verticalmente es necesario que los atributos menos abstractos se ubiquen en la rama adecuada del árbol conceptual; es decir, inmediatamente subordinados del atributo más abstracto al cual le da cuerpo y hacen más concreto. De otro modo, este atributo se enlazaría a atributos que son manifestaciones de un atributo más general diferente, creando un problema de superposición (*conflation*). Por otro lado, los atributos en el mismo nivel de abstracción deben ser manifestaciones de aspectos mutuamente excluyentes del atributo en el nivel de abstracción inmediatamente superior. De otro modo, el análisis cae presa del problema lógico distintivo de la *redundancia* (véanse ejemplos en la *figura 1*).

Respecto a esta segunda tarea relacionada con el reto de la conceptualización —la organización vertical de los atributos según el nivel de abstracción— todos los índices sobre la democracia distinguen cuidadosamente el nivel de abstracción de sus atributos y, por tanto, aíslan con claridad las hojas de sus árboles conceptuales (véanse las columnas 2 y 3 del *cuadro 3*). Sin embargo, esos índices no evitan problemas básicos de la lógica conceptual. El problema de la redundancia es evidente en dos índices. Polity IV cae presa de este problema porque identifica un par de atributos (Competitividad y Regulación de la participación) que captan sólo un aspecto de la democracia, la medida en que las elecciones son competitivas; y otro par de atributos (Competitividad y Apertura del reclutamiento del Ejecutivo) que también corresponden a un solo tema, si los cargos públicos son ocupados por elección popular o por otro procedimiento. Asimismo, es difícil distinguir el subcomponente que Hadenius llama Apertura de las elecciones de los tres componentes en que desagrega su atributo Libertades políticas (véase el *cuadro 3*).

El problema de la superposición es aún más común. Arat abre la puerta a este problema al combinar cuatro componentes bajo un solo atributo general, que llama Participación, cuando estos componentes se vinculan lógicamente a dos atributos diferentes: cargos y el poder de fijar la agenda de los funcionarios públicos. Lo mismo sucede con Bollen (1980, p. 376), quien incluye bajo su atributo Soberanía popular a dos componentes (Selección del Ejecutivo y del Legislativo) que captan y, por lo tanto, desagregan muy útilmente un solo atributo, si los cargos clave son llenados por elecciones, pero quien también incluye un tercer componente (Imparcialidad de las elecciones) que parece estar más vinculado con otro atributo, como la participación. Asimismo, el índice de Hadenius podría culparse por incluir en su atributo Elecciones un conjunto de componentes y subcomponentes que están claramente relacionados con el proceso electoral (Sufragio, Apertura e Imparcialidad), pero también otros componentes y subcomponentes (Cargos elegidos, Efectividad) que se pueden considerar, más claramente, como aspectos de otros atributos, como cargos y el poder de fijar la agenda de los funcionarios públicos. Por último, el índice de Freedom House incluye tantos componentes en sus dos atributos

Derechos políticos y Derechos civiles (9 y 13, respectivamente) y con tan poca consideración acerca de la relación entre componentes y entre componentes y atributos —los componentes se presentan como poco más que una lista *ad hoc* (Ryan, 1994, p. 10)— que no sorprende que se hayan amontonado juntos un gran número de aspectos distintivos o, a lo sumo, vagamente relacionados (Bollen, 1986, p. 584).

En verdad, los constructores de los índices sobre la democracia suelen ser bastante conscientes de cuestiones metodológicas. De este modo, todos presentan explícitamente sus definiciones de democracia, destacan los atributos que han identificado y distinguen claramente estos atributos según su nivel de abstracción. Además, unos pocos índices abordan algunas de las tareas específicas de una forma ejemplar. En este sentido, Hadenius demuestra gran perspicacia al identificar los atributos que son constitutivos del concepto de democracia, y lo mismo puede decirse de ACLP con respecto a cómo organizan lógicamente sus distintos atributos.⁸ No obstante, hay lugar para mejoramiento en lo que concierne a la especificación del concepto y la lógica conceptual.

EL RETO DE LA MEDICIÓN: INDICADORES Y NIVELES DE MEDICIÓN

El segundo reto en la generación de datos es la formación de medidas, que vinculan los atributos conceptuales identificados y organizados lógicamente durante el paso anterior con las observaciones. El reto de la medición toma como punto de partida los atributos que se encuentran en el nivel más bajo de abstracción, llamados hojas. Sin embargo, es importante señalar que aun cuando los conceptos estén muy bien desarrollados, esas hojas rara vez son observables en sí mismas. De ahí que, por usar la terminología acuñada por los psicometristas, es necesario formar modelos de medición que relacionen “variables latentes” no observables con “variables observables” o indicadores

⁸ Algunos índices que no desagregan mucho el concepto de democracia —los de Vanhanen y Gasiorowski— evitan los problemas de lógica conceptual, pero sólo porque renuncian a la posibilidad de desarrollar el concepto analíticamente y tender un puente entre el concepto abstracto de democracia y sus atributos más concretos. Los costos de esta opción son muy elevados.

(Bollen, 1989, cap. 6). Este es un reto sumamente complejo, pues exige tener en cuenta muchas cuestiones. Sin embargo, es muy justificable darle relevancia a dos tareas —la selección de indicadores y el nivel de medición— y a un estándar: la *validez* de las medidas, es decir, el grado en que las medidas propuestas realmente miden lo que se supone deben medir (Carmines y Zeller, 1979; Bollen, 1989; Adcock y Collier, 2001). Por eso empezamos por discutir estas cuestiones antes de pasar a algunas adicionales.

La primera decisión en la formación de las medidas es la *selección de indicadores* que operacionalizan las hojas de un árbol conceptual. Esta es una de las metas más escurridizas de las ciencias sociales, ya que no existen reglas indiscutidas para elegir indicadores válidos. Sin embargo, se pueden derivar ciertas lecciones al considerar el impacto de dos problemas comunes sobre la validez de las medidas. Un problema muy común es la tendencia a no reconocer las múltiples manifestaciones empíricas de un atributo conceptual y a no usar indicadores múltiples de una manera adecuada. Este es probablemente uno de los problemas más difíciles de evitar en la construcción de bases de datos grandes. Pero es difícil exagerar la importancia de esta cuestión. Por un lado, cuanto más uno busca formar medidas con el fin de hacer comparaciones a través de el tiempo y el espacio, más necesario es evitar los sesgos potenciales asociados con indicadores únicos usando indicadores múltiples. Por otro lado, mientras más se utilicen indicadores múltiples, más crece la necesidad de establecer la equivalencia de los indicadores diversos en contextos diferentes y la dificultad de esta tarea. Por lo tanto, una pauta importante para maximizar la validez de los indicadores es elegir indicadores múltiples, pero de modo que se aborde explícitamente la necesidad de establecer la equivalencia de dichos indicadores a través de distintas unidades o casos (Przeworski y Teune, 1970, caps. 5 y 6).

Otro problema común asociado con la selección de indicadores es la falta de apreciación de la naturaleza inexorable del error de medición. Como regla general, la elección de los indicadores está natural e inevitablemente guiada en parte por la disponibilidad o accesibilidad de los datos. Por lo tanto, es comprensible que esos aspectos prácticos afecten la selección de indicadores. Pero

esto representa un grave problema porque el registro que deja la historia está inherentemente sesgado. Por ejemplo, las diferencias en los niveles de violaciones reportadas puede deberse más a cambios culturales que al número real de violaciones. Asimismo, el aumento de evidencia de corrupción podría ser más un reflejo de un aumento en la libertad de prensa que de un incremento real en la corrupción. Este problema recalca la necesidad de que los analistas estén conscientes de cualquier fuente sistemática de error de medición y, específicamente, que maximicen la validez de sus indicadores eligiendo aquellos que probablemente sean menos afectados por sesgos y que pueden ser verificados a través de fuentes múltiples (Bollen, 1986, pp. 578-587, 1993).

Los índices de democracia demuestran diversos grados de atención a la necesidad de usar indicadores múltiples y de establecer la equivalencia de estos indicadores a través de distintas unidades. ACLP (1996, pp. 7-13) y Hadenius (1992, pp. 36-60) ofrecen una justificación detallada de sus indicadores que muestra gran sensibilidad al contexto. Sin embargo, en otros casos, aunque se presenten explícitamente los indicadores, la falta de un análisis detallado hace difícil entender cómo estos indicadores reflejan diferencias del contexto de los casos. Y en otras instancias, el uso de datos generados por otros —una práctica común— está asociado con una fuerte tendencia a simplemente esquivar la necesidad de justificar la elección de indicadores (Arat, 1991, cap. 2; Bollen, 1980, pp. 375-376).

Finalmente, uno de los ejemplos más problemáticos respecto a la elección de indicadores, un tanto irónicamente, es ofrecido por Vanhanen (1993, pp. 303-308 y 310), quien defiende el uso de “indicadores cuantitativos simples” y se opone a las medidas “demasiado complicadas y que tienen demasiados indicadores [...] que [...] dependen demasiado de evaluaciones subjetivas.” El problema está en que Vanhanen exagera el contraste entre los indicadores subjetivos y los objetivos y, en consecuencia, no le presta mucha atención a los juicios subjetivos que subyacen la selección de indicadores “objetivos” (véase, sin embargo, Vanhanen, 2000a, p. 255). No sorprende, entonces, que la decisión de Vanhanen de medir su atributo Competición en términos del porcentaje de votos que recibe el mayor partido, y su atributo Participación en términos

de la proporción del electorado que vota, haya sido criticada porque esos indicadores no sólo constituyen, cuando mucho, medidas deficientes de los atributos en cuestión, sino porque también introducen un sesgo sistemático al ejercicio de medición (Bollen, 1980, pp. 373-374; 1986, pp. 571-572; 1991, pp. 4 y 11 y Hadenius, 1992, pp. 41 y 43). En general, pues, los índices sobre la democracia demuestran insuficiente sensibilidad hacia los aspectos clave involucrados en la elección de indicadores.

La validez es de nuevo un asunto fundamental en lo que hace a la segunda tarea en la formación de medidas: la *selección del nivel de medición*. Esta tarea exige que los analistas ponderen consideraciones opuestas y tomen decisiones sensatas que reflejen un conocimiento profundo de los casos estudiados. Por consiguiente, no tiene fundamento la idea generalizada de que la selección del nivel de medición es algo que se decide únicamente haciendo referencia a suposiciones *a priori*. Tampoco es razonable afirmar que, de las opciones estándar entre las escalas nominales, ordinales, intervalar o proporcionales, la opción de un nivel de medición más cercano a una escala proporcional —convencionalmente entendida como el nivel de medición más alto porque hace las distinciones más precisas— debe ser preferida *a priori*. De hecho, la mejor pauta es la sugerencia más abierta de que la selección de nivel de medición debe:

- Ser dirigida por la meta de maximizar la homogeneidad dentro de las clases de medición con el número mínimo de distinciones necesarias, y
- Ser vista como un proceso que requiere tanto una justificación teórica como una prueba empírica (Gifi, 1990; Jacoby, 1991, 1999).

Desde esta perspectiva, la selección del nivel de medición puede ser vista como un intento por evitar los excesos de introducir o distinciones demasiado finas, lo cual produciría afirmaciones sobre la medición que simplemente no son plausibles a la luz de la información disponible y el grado en que puede minimizarse el error de medición, o demasiado gruesas, lo cual llevaría a que casos que estamos seguros que son bastante diferentes sean colocados juntos.

Esta no es una tarea sencilla ni mecánica. Por lo tanto, la elección del nivel de medición debe hacer uso del conocimiento, y estar sujeta al escrutinio, de los expertos. Además, debe estar condicionada por la disponibilidad de los datos y el grado probable de error de medición y, por lo tanto, no “exigir medidas que de hecho no podemos obtener” (Kaplan, 1964, p. 283). Por último, la elección del nivel de medición debe estar abierta a pruebas, en el sentido que los analistas deberían considerar las implicaciones de las diferentes suposiciones acerca del nivel de medición y usar una evaluación de esas implicaciones al justificar sus decisiones.

Pese a la importancia de esta decisión en el proceso global de generación de datos, los índices sobre la democracia probablemente prestan menos atención a los temas involucrados en la selección del nivel de medición que a la selección de los indicadores. Como muestra el *cuadro 3*, los diferentes índices usan escalas nominales, ordinales y intervalares. Sin embargo, con pocas excepciones, los defensores de diferentes niveles de medición rara vez van más allá de ofrecer aseveraciones acerca de la inherente superioridad del nivel de medición elegido y, por ende, no asumen adecuadamente la responsabilidad de justificar y probar empíricamente la validez de su elección (Collier y Adcock, 1999).⁹ Esta tendencia es particularmente llamativa en el caso de Bollen (1991, pp. 9 y 14), quien simplemente declara que: “el concepto de democracia política es continuo,” como si esto fuera patentemente obvio, y ACLP (1996, p. 21), quienes insisten en que el punto de vista de Bollen es “ridículo.” Desgraciadamente, la selección del nivel de medición es uno de los puntos más débiles de los índices sobre la democracia.

Más allá de la preocupación por maximizar la validez de las medidas, otros dos estándares básicos de evaluación merecen la atención en el contexto del reto de la medición. Uno corresponde a la *confiabilidad* de las medidas, es decir, la expectativa de que el mismo proceso de recopilación de datos siempre produzca los mismos datos. Los esfuerzos por determinar la confiabilidad de una medida, que

⁹ Un aspecto de la selección del nivel de medición incluiría pruebas que evalúan el impacto de usar diferentes puntos de corte (*cut-off points*) para organizar los casos en un continuo, como lo hizo Elkins (2000) con los datos reunidos por ACLP.

es típicamente evaluada por el grado de convergencia entre las medidas producidas por analistas usando los mismos procedimientos, son útiles en dos sentidos:

- *Primero*, si las pruebas de confiabilidad resultan débiles, sirven para alertar a los analistas acerca de problemas potenciales en el proceso de medición.
- *Segundo*, si las pruebas de confiabilidad generan resultados positivos, pueden ser interpretadas como una indicación de la confiabilidad de las medidas propuestas.

Al mismo tiempo, es importante señalar que estas pruebas no deben interpretarse como pruebas de la validez de las medidas. Una confiabilidad débil no ofrece pistas de cuáles medidas son más válidas, tan sólo de que existe desacuerdo acerca de cómo se deben medir los casos. A su vez, una confiabilidad fuerte puede surgir si todos los analistas comparten los mismos sesgos y, por tanto, no debe ser interpretada como una señal de validez de una medida. De hecho, una manera de obtener medidas muy confiables es adoptar sesgos similares, algo que se hace con demasiada frecuencia, aunque sea inconscientemente. Por tanto, si bien la confiabilidad es obviamente deseable, pues ofrece una señal del grado de consenso entre analistas, es importante reconocer que siempre existe la posibilidad de sesgos sistemáticos en la medición. Las medidas confiables no son necesariamente válidas.

Otro estándar de evaluación corresponde a la *replicabilidad* de las medidas, esto es, la capacidad de una comunidad de estudiosos de reproducir el proceso mediante el cual se generaron los datos. Este asunto tiene poco valor por sí solo; la replicabilidad nos preocupa porque las afirmaciones sobre la validez o la confiabilidad dependen de la replicabilidad de las medidas. Sin embargo, puesto que la medición involucra cuestiones que son inexorablemente subjetivos, suponiendo una variedad de juicios en lugar de criterios firmemente objetivos, es absolutamente vital que la comunidad de estudiosos conserve la habilidad de examinar y cuestionar las elecciones que afectan la generación de datos. De este modo, al abordar la formación de medidas, los analistas deben *registrar y divulgar*:

- Sus *reglas de medición*, que deben incluir, cuando menos, un listado de todos los indicadores, los niveles de medición seleccionados para cada indicador, e información lo suficientemente detallada para que estudiosos independientes puedan interpretar el significado de cada escala;
- El *proceso de medición*, que debe incluir un listado de las fuentes usadas en el proceso de medición, el número de analistas involucrados en la preparación de los datos y los resultados de cualesquiera pruebas del grado de convergencia entre las medidas producidas por los diversos analistas comprometidos, y
- Los *datos desagregados* generados para todos los indicadores.

Respecto a estas tareas, los índices de democracia merecen una evaluación mixta. En lo que se refiere a las reglas de medición, ACLP (1996, pp. 7-14), Hadenius (1992, pp. 36-40) y Polity IV (Marshall y Jaggers, 2001a) son modelos de claridad, especificando sus reglas de medición explícitamente y en bastante detalle. Otros también son demasiado explícitos acerca de sus reglas de medición, pero no ofrecen todos los pormenores y, por tanto, dejan mucho margen a la interpretación. Otros más, como los de Freedom House y Gasiorowski, nunca proporcionan un conjunto claro de reglas de medición y, por consiguiente, no ofrecen una base para un diálogo real acerca de cómo se midieron los casos.

Respecto al proceso de medición, los índices son bastante criticables. Todos los creadores de índices ofrecen alguna información sobre las fuentes consultadas en el proceso de medición. Pero el nivel de detalle es tal que un estudioso independiente dudosamente podría reconstruir precisamente qué información específica tenía en mente el analista al momento de clasificar un caso de una forma determinada. De hecho, el tipo de información ofrecida no va más allá de referencias a títulos de libros o fuentes generales como el *Registro de acontecimientos mundiales de Keesing (Keesing's Record of World Events)*, sin indicar qué información fue extraída de esas fuentes, precisamente dónde puede encontrarse esa información y qué atributos fueron medidos con base en esa información. Además, los índices existentes dejan mucho que desear en lo que se refiere a la información acerca de quién clasificó los casos, si varios

analistas estaban involucrados y, de ser así, si se hicieron pruebas de confiabilidad entre estos analistas. En algunos casos aislados, el problema es tan básico como saber cuántas personas fueron responsables de crear los datos. Aunque en la mayoría de casos se ofrece esta información, la práctica común de usar un solo analista incrementa la posibilidad de que los datos generados estén afectados por sesgos significativos. Finalmente, en algunos casos el beneficio potencial asociado con el uso de varios analistas es negado debido a que no se realizaron pruebas de confiabilidad entre los analistas (Ryan, 1994, pp. 7 y 11). De hecho, en sólo dos casos —el índice de Coppedge y Reinicke (1991, p. 55) y Polity IV (Marshall y Jaggers, 2001a, pp. 5-8)— se utilizaron varios analistas y se aplicaron pruebas de confiabilidad entre ellos.¹⁰

Por último, respecto a la disponibilidad de los datos desagregados, los índices sobre la democracia tienden a ser metodológicamente bastante sólidos. Algunos creadores de índices ofrecen solamente datos agregados.¹¹ Pero la mayoría ha divulgado tanto sus datos desagregados, como agregados y también ofrecido los datos desagregados a quienes lo soliciten, o puesto los datos desagregados al alcance del público por medio de Internet (véanse las fuentes del cuadro 1).

Los problemas que afectan a varios índices con respecto a una u otra tarea correspondiente a la formación de medidas palidece, sin embargo, frente a la respuesta insatisfactoria que dan a cada una de las tres tareas implicadas en la medición de un concepto los índices creados por Gasiorowski y Freedom House. El primer problema del índice de Gasiorowski es que nunca se realiza un intento por generar medidas al nivel de los atributos. Esto es, si bien Gasiorowski ofrece definiciones para los tres atributos de su índice, el esfuer-

¹⁰ Vanhanen evita muchos de esos problemas potenciales porque usa indicadores "objetivos." Además, ha hecho públicas las cronologías que usa para generar sus datos.

¹¹ En una comunicación personal, Arat ha indicado que pondría a la disposición sus datos desagregados si pudiese, pero que éstos fueron recopilados antes del uso generalizado de las computadoras. Por lo tanto no puede presentarlos en un formato computarizado. En el caso del índice de Freedom House, a pesar de que hemos solicitado acceso a los datos desagregados, no hemos podido disponer de ellos. En el caso de Gasiorowski (1996, pp. 480-482), los únicos datos que fueron generados son datos agregados. Después de terminado este artículo aprendimos que Bollen ha extendido el alcance de su base de datos para cubrir el periodo 1950-1990 y ha hecho públicos sus datos desagregados (Bollen, 2001).

zo de medición luego ignora estos atributos desagregados y se enfoca directamente en el nivel más agregado, negando la razón básica para desagregar un concepto. En el nivel agregado, Gasiorowski (1996, pp. 471-472) propone una escala ordinal de tres puntos —distinguiendo entre democracia, semidemocracia y autoritarismo— con una categoría residual para regímenes en transición. Esta opción está bien fundada en la literatura, pero Gasiorowski no presenta un análisis explícito de sus indicadores y sus reglas de medición. Finalmente, a pesar de que Gasiorowski identifica las fuentes que utiliza y hasta ha puesto a disposición del público los resúmenes narrativos que usa para clasificar sus casos, no existe manera de que un investigador independiente pueda replicar el proceso de medición que usa Gasiorowski, algo particularmente necesario dado que todas las medidas son el resultado del esfuerzo —y los sesgos— de una sola persona, el propio Gasiorowski (pp. 473-475).

El problema con el índice de Freedom House comienza con la selección de los indicadores. Aunque este índice refleja una conciencia de la necesidad de usar diferentes indicadores en distintos países (Gastil, 1991, pp. 25-26) esta sensibilidad a cuestiones de contexto no ha sido complementada por un esfuerzo paralelo por establecer la equivalencia de indicadores diferentes.¹² Los problemas continúan con respecto a la selección del nivel de medición. Cada uno de los componentes usados por Freedom House (Gastil, 1991, pp. 26 y 32-33; Ryan, 1994, pp. 10-11) es medido con una escala ordinal de cinco puntos. Esta decisión bien podría ser razonable, pero no se ofrece una justificación para adoptarla. De hecho, esta elección parece responder a un interés en la simetría más que a consideraciones basadas en la teoría o la estructura de los datos. Finalmente, oscureciendo todo el ejercicio, el proceso de medición no está abierto al escrutinio público. Puesto que Freedom House no ofrece sus reglas de medición, analistas independientes no tienen información acerca de los rasgos distintivos que hacen que un caso reciba una puntuación de 0, 1, 2, 3 o 4 puntos. Además, las fuentes de información no se identifican con la precisión suficiente para que los investi-

¹² Además, a pesar de que se usan múltiples fuentes, no hay señales de que se haya prestado atención al hecho de si la elección de los indicadores aumenta y no minimiza el error de medición atribuible al conjunto de fuentes del que depende el índice (Bollen, 1986, pp. 583-586).

gadores independientes puedan reanalizarlas. Y, para empeorar las cosas aún más, la decisión de Freedom House de no divulgar sus datos desagregados asegura que un debate público sobre los datos de Freedom House sea prácticamente imposible. Al fin de cuentas, los datos agregados ofrecidos por Freedom House deben ser aceptados en gran medida como asunto de pura fe.¹³

En resumen, los índices sobre la democracia no han abordado muy bien el reto de la medición. Pueden rescatarse algunos aspectos positivos. Con respecto a la selección de indicadores, valiosas sugerencias son ofrecidas por ACLP y Hadenius. Además, en lo que hace al registro y la divulgación de las reglas de medición, el proceso de medición y los datos desagregados, ACLP, Coppedge y Reinicke, y Polity IV establecen un estándar alto. Pero la tendencia más amplia es claramente negativa. Los casos de Gasiorowski y Freedom House son ejemplos de respuestas altamente problemáticas al reto de la medición. Más generalmente, se puede afirmar que los índices existentes exhiben varios problemas. Hacen poco para seleccionar indicadores que reflejen una sensibilidad al contexto, a problemas de equivalencia y a errores de medición. Suelen estar basadas en un enfoque poco sofisticado a la selección del nivel de medición. Y, por último, no toman los pasos adecuados para asegurar la replicabilidad. La necesidad de abordar el reto de la medición de una forma más cuidadosa es muy evidente.

EL RETO DE LA AGREGACIÓN: NIVELES Y REGLAS DE AGREGACIÓN

Una vez que el proceso de medición está completo con la asignación de medidas a cada una de las hojas del árbol conceptual, los analistas enfrentan un ter-

¹³ Cabe señalar otros problemas. El proceso de medición usado por Freedom House ha cambiado a través del tiempo. De 1977 a 1987, cuando Gastil (1991, pp. 22-23) estuvo a cargo del índice, un solo analista, Gastil, generó todos los datos. Durante este período, también parece que aunque había una lista de componentes, las medidas realmente se hacían al nivel de los dos atributos del índice. Después de 1989, las medidas eran hechas ya no por un individuo sino por un equipo y, al parecer, ya no al nivel de los atributos sino de los componentes (Ryan, 1994, pp. 7 y 11). Si bien ésta es una mejora, la lista básica de componentes utilizada en la construcción del índice sufrió varios cambios (compárese Gastil, 1991, pp. 26 y 32-33, y Ryan, 1994, p. 10). Por lo tanto, un problema con el índice de Freedom House es que la consistencia interna de la serie de datos es cuestionable.

cer reto: determinar si es deseable revertir el proceso de desagregación que se realizó durante la etapa de conceptualización y, si es que se decide ofrecer datos agregados, de qué manera se generarán.¹⁴ Pese a que esta etapa es muy importante, no ha recibido mucha atención en la literatura sobre metodología. Pero, como mostramos, este reto gira en torno a “tres tareas centrales”.

La primera tarea que debe enfrentarse —la *selección del nivel de agregación*— requiere el balanceamiento de dos fines deseables. Por un lado, la mera cantidad de atributos e información que puede estar asociada con un concepto denso, minuciosamente elaborado, puede hacer que la investigación realizada al nivel más desagregado se transforme en un ejercicio bastante engorroso. Así, es razonable que los analistas consideren que sea conveniente generar datos más resumidos, en el sentido de que tales datos serán analíticamente más manejables y facilitarán la elaboración de teorías y pruebas empíricas. Por otro lado, es necesario reconocer que subir a un nivel superior de agregación puede llevar a la pérdida de validez, ya que puede ocasionar la pérdida de información acerca de variación sistemática entre los casos. Así, es igualmente necesario reconocer los costos potenciales implicados en la elección de agregar datos desagregados. En pocas palabras, no hay reglas fáciles al que un analista pueda recurrir para resolver esta cuestión. Más bien, la selección del nivel de agregación es una elección explícita que debe ser justificada a la luz de la necesidad de balancear la búsqueda de parsimonia con la preocupación por la dimensionalidad subyacentes a los conceptos y la diferenciación entre los casos.

Si bien el reto de la agregación es relevante para todos los índices sobre la democracia que abordamos aquí con excepción de uno,¹⁵ en muchos casos se aborda de maneras bastante deficientes. La práctica común con respecto a la selección del nivel de agregación ha sido proceder como si lo único importante fuera la parsimonia, lo que justificaría la decisión de proceder al nivel de agregación más alto posible; esto es, la reducción de los datos desagregados a una

¹⁴ Todo este paso supone que se ha llevado a cabo cierta desagregación; es decir, que se ha identificado más de un atributo.

¹⁵ La excepción es el índice de Gasiorowski, que no mide los casos a un nivel desagregado.

sola medida por caso.¹⁶ Por lo tanto, los creadores de índices han hecho poco por evitar una pérdida de información. Aún más importante, no han hecho mucho por probar si los niveles inferiores de agregación miden lo que es realmente un fenómeno unidimensional y, de este modo, si la agregación puede realizarse sin forzar un fenómeno multidimensional dentro de una métrica común, una práctica que quita validez a las medidas resultantes. De hecho, con una notable excepción, no se ofrece una justificación teórica para la elección del nivel de agregación y no se hace ningún esfuerzo real por probar si la agregación al mayor nivel posible es apropiada. Sin duda, el afán por ofrecer una sola medida por caso responde al deseo de usar regresión múltiple y técnicas relacionadas para analizar los datos. Sin embargo, esto es como poner la carreta estadística delante del burro teórico.

La excepción es Coppedge y Reinicke (1991, pp. 52-53 y Coppedge, 1997, pp. 180-184), quienes abordan el proceso de agregación construyendo una escala Guttman. La ventaja de este tipo de escala es que permite que el proceso de agregación se realice sin perder información a medida que se pasa de un nivel inferior a uno superior de agregación y sin tener que asignar pesos a cada componente. El problema, sin embargo, es que una escala Guttman sólo puede construirse si los componentes múltiples se mueven juntos y miden la misma dimensión subyacente, lo cual no parece ser precisamente el caso con los componentes usados en el índice de Coppedge y Reinicke.¹⁷ A pesar de los límites de la utilidad de las escalas de Guttman

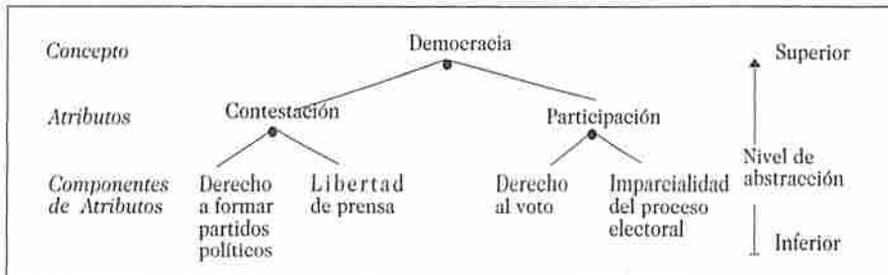
¹⁶ Dos excepciones parciales son los índices de Freedom House y Polity IV. El índice de Freedom House agrega sus datos sólo hasta el nivel de sus dos atributos —Derechos políticos y Derechos civiles— y así ofrece dos medidas para cada caso. Polity IV ofrece dos medidas, una para democracia y otra para autocracia. Sin embargo, estas dos medidas son generadas meramente por darles diferentes pesos a los mismos datos desagregados (Jaggers y Gurr, 1995, p. 472).

¹⁷ El hecho de que 33 de los 170 países incluidos en el índice de Coppedge y Reinicke (1991, pp. 52-53; Coppedge, 1997, pp. 181-183) no pueden ser localizados en su escala Guttman es notable. Según señaló el mismo Guttman (1977, p. 100), "escalabilidad no es algo que debe desearse o construirse," sino tratarse como una hipótesis. Además, enfatizó que al probar la "hipótesis de escalabilidad" no se pueden examinar varios componentes o atributos, ver cuáles forman escalas y luego eliminar los que no lo hacen; cálculos de probabilidades basados en tal procedimiento son plenamente inválidos (véase también Mokken, 1971, cap. 3). Después de todo, los atributos originales se eligieron por una razón teórica relevante, y excluirlas porque no son escalables tiene el potencial de sacarle provecho a la casualidad. Por lo tanto, el hecho de que Coppedge y Reinicke no hayan identificado una escala cumulativa sugiere la presencia de multidimensionalidad.

en un contexto multidimensional, Coppedge y Reinicke demuestran una sensibilidad ejemplar hacia la posible pérdida de información que puede ocurrir en el proceso de agregación y, más importante aún, acerca de la necesidad de probar más que simplemente afirmar la unidimensionalidad de los conceptos.

La segunda tarea que deben enfrentar los analistas, si se decide proceder a un nivel superior de agregación, es la *selección de la regla de agregación*. Esta es una tarea que supone, como requisito esencial, que los atributos de un concepto han sido organizados lógicamente de manera explícita, algo señalado arriba. De hecho, puesto que la selección de una regla de agregación exige identificar con claridad qué atributos deben agregarse y en qué orden (como se muestra en la *figura 2*), esta tarea depende de que se hayan resuelto con anterioridad los problemas de lógica conceptual. Pero la selección de una regla de agregación adecuada es una tarea distinta, que gira en torno a la formalización de la teoría acerca de los vínculos entre atributos.

FIGURA 2. EL PROCESO DE AGREGACIÓN



NOTA: El punto (*) representa un nodo. La agregación comienza al nivel inferior de abstracción, donde medidas son asignadas a las hojas, y sube hacia niveles superiores de abstracción. Además, la agregación exige el uso de reglas de agregación, que especifican el vínculo teórico entre los atributos que se encuentran en el mismo nivel de abstracción y están conectados por el mismo atributo general (por medio de un nodo). En este ejemplo, la selección de las reglas de agregación primero tendría que enfocarse en la relación entre el Derecho a formar partidos políticos y la Libertad de prensa, para definir una medida para Contestación, y entre Derecho al voto e Imparcialidad del proceso electoral, para generar una medida para Participación. A partir de entonces, si se decide pasar al siguiente nivel de agregación, representado aquí por Democracia, tendría que enfocarse en ese momento en la relación entre Contestación, y Participación.

Esta tarea involucra un proceso de dos pasos: En el primero, el analista debe hacer explícita la teoría referente a la relación entre los atributos. En el segundo, el analista debe asegurarse de que exista una correspondencia entre esta teoría y la regla de agregación seleccionada; es decir, que la regla de agregación sea realmente la expresión formal equivalente de la relación planteada.¹⁸ Por ejemplo, si se está discutiendo la agregación de dos atributos y la teoría indica que ambos tienen el mismo peso, simplemente se agregarían las medidas de ambos atributos. Si la teoría indica que ambos atributos son rasgos necesarios, se multiplicaría ambas medidas, y si la teoría indica que ambos atributos son rasgos suficientes, se tomaría la medida del atributo más alto. En este sentido, es esencial que los investigadores estén conscientes de las múltiples maneras en que pueden vincularse los atributos y evitar la tendencia a limitarse a usar opciones simples, como la adición, que se usa frecuentemente casi sin pensar.¹⁹

A pesar de la importancia de la teoría como guía en la selección de las reglas de agregación, así como en la selección de los niveles de agregación, sigue siendo fundamental subrayar que dichas opciones deben estar abiertas a pruebas. Por lo tanto, los analistas deben considerar qué resultados se conseguirán al aplicar reglas de agregación diferentes y así obtener un sentido de la *robustez* de los datos agregados; es decir, el grado en el cual cambios en la regla de agregación producen cambios proporcionales en los datos agregados. Además, para permitir que otros investigadores repliquen el proceso de agregación y lleven a cabo pruebas referentes a las reglas de agregación, los analistas deben *registrar y divulgar las reglas de agregación y los datos agregados*.

En lo que respecta a estas diversas tareas, las bases de datos sobre la democracia son otra vez menos que adecuadas. En el caso de Freedom House, la regla de agregación seleccionada es clara y explícita: se generan medidas

¹⁸ Este aspecto es análogo al problema de la especificación de la forma funcional en el análisis de regresiones.

¹⁹ Cuando la teoría no es lo suficientemente precisa para permitir una correspondencia clara con una regla de agregación específica, los analistas pueden recurrir a varias técnicas de análisis de datos, como el análisis de correspondencia, los componentes principales, el análisis de factores y la escalación dual.

para los dos atributos —Derechos políticos y Derechos civiles— sumando las medidas asignadas para cada uno de sus respectivos componentes.²⁰ Si bien esta operación parece muy inocente, está llena de problemas. Primero, dado que la muy larga lista de componentes usados por Freedom House no se presenta como un conjunto teóricamente conectado de componentes sino tan sólo como una lista *ad hoc* (Ryan, 1994, p. 10), no se ofrece una justificación teórica para elegir esta regla de agregación. Segundo, la implicación de una agregación mediante la suma —la asignación del mismo peso a cada componente— parece plenamente inadecuado a la luz del contenido de los componentes. Por dar un ejemplo, parece infundado darle al asunto de la centralización del poder (componente número 9 del atributo Derechos políticos) el mismo peso e importancia para la democracia que el poder real ejercido por los representantes electos (componente número 4 del atributo Derechos políticos, Ryan, 1994, p. 10). Tercero, a pesar de que los analistas independientes tienen buenas razones para cuestionar la regla de agregación utilizada por Freedom House, no pueden probar las implicaciones de las diferentes reglas de agregación ya que Freedom House no ha divulgado sus datos desagregados. En pocas palabras, a los numerosos problemas conceptuales y de medición que afectan al índice de Freedom House hay que agregar un descuido llamativo del reto de la agregación.

En este aspecto, sólo ligeramente mejores que el índice de Freedom House son los de Vanhanen y Polity IV. Vanhanen (2000a, pp. 255-257) propone una regla de agregación clara y sencilla: las medidas agregadas son generadas al multiplicar las medidas de sus dos atributos. Sin embargo, no se ofrece una justificación teórica para esta asignación de un mismo peso a cada atributo,²¹ y tampoco se realiza un esfuerzo por probar las implicaciones de las diferentes

²⁰ Posteriormente, las puntuaciones totales se transforman en una escala de 7 puntos, que a su vez se divide en tres categorías —libre, parcialmente libre, no libre— mediante un conjunto de decisiones bastante arbitrarias (Ryan, 1994, p. 11).

²¹ Al igual que la suma, la multiplicación da el mismo peso a cada atributo. Pero, a diferencia de la suma, la multiplicación le otorga un peso mayor a cada atributo. Es decir, si bien una medida baja de un componente del índice de Freedom House podría ser compensado con una medida más alta de otro componente, en el índice de Vanhanen la medida baja de un atributo no puede repararse con una medida más alta de otro atributo.

reglas de agregación. El único aspecto redimible de esta respuesta arbitraria y *ad hoc* al reto de agregación es que Vanhanen, a diferencia de Freedom House, al menos ofrece los datos de sus atributos desagregados. Por lo tanto, otros pueden realizar pruebas independientes de las diferentes maneras en que las reglas de agregación afectan las medidas agregadas.

A su vez, el índice Polity IV está basado en una regla de agregación explícita, pero bastante intrincada y problemática (Marshall y Jaggers, 2001a, pp. 11-14). Primero, los cinco atributos del índice tienen un peso diferente, porque se usan escalas distintas y asignan un número de puntos diferentes a cada atributo. Si bien darle distintos pesos a atributos es una forma legítima de reconocer la mayor o menor importancia teórica de diferentes atributos, el problema con Polity IV es que no se justifica el esquema de pesos. Segundo, las medidas asignadas a los cinco atributos son sumadas para generar o dos medidas (una para democracia y otra para autocracia) o una sola medida (el índice global de Polity IV) dando lugar aún a más problemas. No sólo prácticamente no se ofrece una justificación teórica para esta operación, sino que también es criticable debido a problemas de lógica conceptual. De hecho, como se vio antes, Polity IV incluye un par de atributos redundantes, lo que resulta en mucho conteo doble de los mismos elementos que nunca es reconocido o explicado. Una cualidad redimible de Polity IV, sin embargo, es que los datos desagregados están disponibles al público, asegurando así que los investigadores independientes puedan evaluar las implicaciones de las diferentes reglas de agregación y potencialmente sugerir reglas de agregación más adecuadas.

Otros índices ofrecen enfoques más lúcidos al proceso de agregación, pero de todos modos no están exentos de problemas. Arat (1991, p. 26) presenta una regla formal de agregación bastante complicada. Sin embargo, pese a que la regla de agregación es superficialmente razonable, no está explícitamente justificada. Además, la regla de agregación propuesta nunca es puesta a prueba y la oportunidad de que otros analistas realicen pruebas independientes está cerrada porque no están disponibles los datos desagregados. En cambio, ACLP (1996, p. 14) ofrece explícitamente las razones para considerar que un

caso es democrático sólo si el jefe del Ejecutivo y el Legislativo son elegidos mediante elecciones competitivas y, si bien no formalizan su discusión teórica de la conexión entre sus atributos, dejan claro que las medidas positivas de sus tres atributos son individualmente necesarias y conjuntamente suficientes para clasificar un régimen como democrático. No obstante, a pesar de que ofrecen toda la información necesaria para permitir que otros analistas consideren las implicaciones de usar diferentes reglas de agregación, ellos mismos no llevan a cabo esas pruebas.

Por lo tanto, en comparación con las otras bases de datos, el índice de Hadenius es especialmente notable. Él propone una regla de agregación muy complicada, pero la justifica explícita y ampliamente por medio de referencias a la teoría democrática y la formaliza. Asimismo, muestra ser sensible a las implicaciones de las diferentes reglas de agregación, y no sólo ofrece la información necesaria para que otros prueben las implicaciones de las diferentes reglas de agregación, sino que también lleva a cabo pruebas de la robustez de la regla de agregación que él propone (Hadenius, 1992, pp. 61 y 70-71). De hecho, a la luz del bajo estándar impuesto por otros índices, Hadenius ofrece una discusión sobre su regla de agregación que es bastante ejemplar.

En resumen, con unas cuantas excepciones notables, los índices sobre la democracia han mostrado un nivel bastante bajo de sofisticación en lo que se refiere al proceso de agregación. El problema más grande es que la mayoría de los constructores de índices han supuesto simplemente que es apropiado y deseable agregar datos hasta que tengan un índice unidimensional. Pero otros problemas están presentes también. Los constructores de índices han tendido a usar reglas de agregación de manera bastante *ad hoc*, sin ofrecer una teoría explícita acerca de la relación entre los atributos y sin asegurarse de la correspondencia entre la comprensión teórica de cómo están conectados los atributos y la regla de agregación seleccionada. De igual modo, casi no se intenta poner a prueba y evaluar las implicaciones de las diferentes reglas de agregación. El reto de la agregación es, sin lugar a dudas, un punto débil de muchos índices sobre la democracia.

CONCLUSIÓN: UNA EVALUACIÓN GLOBAL DE BASES DE DATOS
SOBRE DEMOCRACIA

Esta reseña de índices sobre la democracia resalta dos puntos clave: Primero, los creadores de índices han demostrado grados muy divergentes de sofisticación al abordar los retos de conceptualización, medición y agregación. Por destacar sólo los puntos fuertes y las debilidades más marcadas, hay que reconocer los esfuerzos de ACLP (1996), quienes son particularmente cuidadosos en cuanto a la selección de indicadores y especialmente claros y detallados con respecto a las reglas de medición; a Coppédge y Reinicke (1991), quienes demuestran una preocupación saludable por establecer la confiabilidad de sus datos y se destacan por ser los únicos que reflejan sensibilidad hacia la cuestión de los niveles de agregación; y a Hadenius (1992), quien ofrece una convincente conceptualización de democracia, una adecuada selección de indicadores, y un complejo uso de reglas de agregación. Las bases de datos que desafortunadamente son tan problemáticas que merecen una mención explícita son los compilados por Freedom House (2000), Gasiorowski (1996) y Vanhanen (2000a, 2000b), que presentan problemas en las tres áreas de conceptualización, medición y agregación (véase el cuadro 4).

Segundo, este artículo muestra que ningún índice ofrece una respuesta satisfactoria a los tres retos de conceptualización, medición y agregación. De hecho, incluso los índices más fuertes padecen de algunas dificultades de importancia. Así, el índice de ACLP está basado en una concepción bastante estrecha de la democracia y es muy débil en lo que se refiere a la selección del nivel de medición; el índice de Coppédge y Reinicke también ofrece una concepción bastante estrecha de la democracia; y el índice de Hadenius padece de numerosos problemas de lógica conceptual. Además, los mejores índices también tienen un alcance bastante restringido (véase el cuadro 1), mientras que los índices que tienen un alcance más amplio, con la excepción parcial de Polity IV, no son muy fuertes. En pocas palabras, a pesar de que estos índices constituyen una contribución importante, queda mucho por mejorar en lo que respecta a la calidad de los datos sobre la democracia.

CUADRO 4: BASES DE DATOS SOBRE LA DEMOCRACIA: UNA EVALUACIÓN

Nombre	Fuerzas	Debilidades
ACLP: Álvarez Cheibub, Limongi, y Przeworski	Identificación de atributos: cargos. Lógica conceptual. Selección adecuada de indicadores. Reglas de medición claras y detalladas.	Definición minimalista: omisión de la participación y el poder de fijar la agenda.
Arat	Identificación de atributos: cargos y poder de fijar la agenda.	Lógica conceptual: problema de superposición (conflation).
Bollen	Identificación de atributos: cargos, poder de fijar la agenda e imparcialidad.	Definición minimalista: omisión de la participación. Lógica conceptual: problema de superposición. Alcance empírico (temporal) restringido.
Coppedge y Reinicke	Identificación de atributos: imparcialidad. Prueba de confiabilidad entre analistas. Procedimiento de agregación complejo.	Definición minimalista: omisión de la participación, los cargos y el poder de fijar agenda. Alcance empírico (temporal) restringido.
Freedom House	Alcance empírico (espacial) amplio.	Definición maximalista. Lógica conceptual: prob. de superposición. Múltiples problemas de medición. Procedimiento de agregación inadecuado.
Gasiorowski	Alcance empírico amplio.	Definición minimalista: omisión de cargos y el poder de fijar la agenda. Múltiples problemas de medición.
Hadenius	Identificación de atributos: cargos, poder de fijar la agenda e imparcialidad. Selección adecuada de indicadores. Reglas de medición claras y detalladas. Procedimiento de agregación complejo.	Lógica conceptual: problemas de redundancia y superposición. Alcance empírico (temporal) restringido.
Polity IV	Identificación de atributos: cargos y poder de fijar la agenda. Reglas de medición claras y detalladas. Prueba de confiabilidad entre analistas. Alcance empírico amplio.	Definición minimalista: omisión de la participación. Lógica conceptual: problema de redundancia. Procedimiento de agregación inadecuado.
Vanhanen	Reglas de medición claras. Alcance empírico amplio. Replicabilidad.	Definición minimalista: omisión de cargos y el poder de fijar la agenda. Indicadores cuestionables. Procedimientos de agregación inadecuado.

A la luz de esta evaluación, puede parecer irónico que la comparación más común entre los índices, hecha por medio de pruebas de correlación sobre datos agregados, haya mostrado consistentemente un nivel muy elevado de correlación entre los índices.²² Estas comparaciones son valiosas y obviamente no pueden descartarse a la ligera. No obstante las diferencias en materia de conceptualización, medición y agregación que hemos notado, estas pruebas parecen mostrar que los índices bajo consideración ofrecen aproximaciones a una misma realidad subyacente. Sin embargo, es importante interpretar esas pruebas adecuadamente. De hecho, en este sentido, deben señalarse tres puntos:

- *Primero*, hasta cierto punto, esas altas correlaciones casi no sorprenden porque, a pesar de todas las diferencias entre estos índices, se han basado, en algunos casos en un grado muy alto, en las mismas fuentes de información e incluso en los mismos datos procesados.²³ Por lo tanto, el alto nivel de correlación puede significar que todos los índices están reflejando el mismo sesgo, que es producto de la contaminación de los sesgos de unas pocas fuentes y datos procesados.
- *Segundo*, como comienza a sugerir el primer punto, esas pruebas de correlación no dan un sentido de la validez de los datos sino tan sólo de su confiabilidad, lo cual es un tema secundario. Este punto lo dejó claro, desde hace mucho, Bollen (1986, pp. 587-588), quien afirmó que "pueden obtenerse medidas muy consistentes (esto es, confiables) que no son válidas" y advirtió que: "la confiabilidad no debe confundirse con la validez." Y algunos creadores de índice, como ACLP (1996, p. 21), claramente se refieren a las pruebas de correlación como un medio para establecer la confiabilidad de su índice. Pero, desafortunadamente, esta distinción es ignorada por otros, quienes usan estas pruebas de co-

²² Véanse las fuentes citadas en la nota 2.

²³ La evidencia más obvia de esto es el uso común de los datos procesados por Arthur Banks (Álvarez et al., 1996, p. 7; Arat, 1991, pp. 30-31; Bollen, 1980, p. 376, 1991, p. 10; Gasiorowski, 1996, p. 473; Gastil, 1978, pp. 8-9 y Hadenius, 1992, p. 177).

rrelación para hacer afirmaciones acerca de la validez.²⁴ De hecho, incluso el mismo Bollen (1980, pp. 380-381; véase también 1986, p. 589) es culpable de crear esta confusión al afirmar que el alto grado de correlación entre su índice y los otros es evidencia de la validez de su propio índice. Por lo tanto, es fundamental subrayar que el alto grado de correlación entre los índices sobre la democracia no resuelve cuestiones acerca de su validez.

- *Tercero*, es importante enfatizar que todas las pruebas de correlación se han realizado con datos altamente agregados y dejan sin resolver el aspecto fundamental de la multidimensionalidad potencial de los datos. Para demostrar este punto, utilizamos el método de componentes principales no-lineal para examinar sistemáticamente las diferencias entre las seis series existentes con una duración relativamente larga y que cubren muchos de los mismos casos: el índice de ACLP, el índice de Gasiorowski, los índices de libertades civiles y derechos políticos de Freedom House, y los índices de democracia y autocracia de Polity IV.²⁵ Como muestra esta prueba (véase la *figura 3*), si bien el índice de ACLP, el índice de Gasiorowski y los dos índices de Polity IV son todos consistentes y los dos índices de Freedom House son semejantes entre sí, existe una diferencia notable entre el índice de ACLP, el índice de Gasiorowski y los dos índices de Polity IV, por un lado, y los dos índices de Freedom House, por el otro, en lo que se refiere a la segunda dimensión. En pocas palabras, estos resultados sugieren que las tablas de correlación que generalmente se presentan como prueba del alto nivel de acuerdo entre índices en efecto pueden ocultar algunas diferencias sistemáticas reales. Por lo tanto, es importante no interpretar mal esas pruebas de correlación y usarlas como una base para descartar los

²⁴ Arat (1991, p. 27), Coppedge y Reinicke (1991, p. 57) y Jaggars y Gurr (1995, p. 473).

²⁵ Utilizamos un método de componentes principales no-lineal porque descomposiciones lineales tienen el potencial de inflar la dimensionalidad de los resultados y porque no son muy útiles en casos de índices, como éstos, que son categóricos. El número de observaciones comunes en cada año varía de 71 a 78. El cálculo se realizó con el módulo de SPSS 10.0 Categorías.

FIGURA 3: CARGAS DE COMPONENTES DE LA COMPARACIÓN DE ÍNDICES SOBRE LA DEMOCRACIA, 1973-1990

	<i>Dimensión 1</i>	<i>Dimensión 2</i>
Álvarez, Cheibub, Limongi y Przeworski	-927	-.180
Gasiorowski	.914	.259
Polity-Autoeracia	-.962	-.274
Polity-Democracia	.953	.251
Freedom House-Libertades civiles	-.569	.801
Freedom House-Derechos políticos	-.556	.809
Porcentaje de varianza	69%	26%

NOTA: Los signos de las cargas son consistentes con la dirección de los datos originales.

numerosos puntos problemáticos que este artículo ha planteado acerca de los índices existentes. De hecho, esas pruebas no ofrecen ninguna base para descartar nuestro análisis y para evitar el debate acerca de cómo mejorar los datos sobre la democracia que este artículo sugiere es sumamente necesario.

Cabe hacer énfasis en que la evaluación crítica de este artículo no busca desalentar los esfuerzos de evaluaciones causales utilizando bases de datos grandes. De hecho, así como hemos subrayado cómo las múltiples decisiones que afectan la generación de datos implica un delicado acto de balancear distintos fines deseables, también consideramos excesivo declarar una moratoria sobre pruebas estadísticas hasta que se hayan resuelto los problemas que hemos señalado. Nos parece que tener una base de datos sobre la democracia, por defectuoso que sea, es mejor que no tener ninguno y que los estudiosos deben usar lo que tienen a su disposición. Pero queremos enfatizar que el cuidadoso desarrollo de medidas constituye una base fundamental para el análisis de inferencias causales y que es una tarea crítica en sí misma.

La necesidad de análisis detallados de medidas, como el que ofrece este artículo, no siempre es reconocida. De hecho, los analistas muchas veces pasan por alto el hecho de que la estadística matemática —que desarrolla la relación entre teoría, datos e inferencia— presupone que la relación entre teoría, datos

y observación ha sido bien establecida. Por lo tanto, uno no puede despreciar la tarea de medición con la esperanza de que la estadística matemática de alguna manera ofrezca una solución a un problema para el cual no está diseñado (Jacoby, 1991). En este sentido, la meta básica y principal contribución de este artículo puede expresarse de la siguiente manera: Al ofrecer un esquema comprensivo e integrado para la generación y/o el análisis de datos, llama la atención hacia las cuestiones complejas que trata un aspecto de la investigación que apuntala la inferencia causal. Además, al aplicar este esquema a las medidas de la democracia y, por lo tanto, respondiendo al llamado de Bollen (1986, p. 589) por "mejores análisis de las medidas existentes," este artículo ha buscado identificar áreas particulares en donde podrían enfocarse provechosamente los esfuerzos por mejorar la calidad de los datos sobre la democracia. En última instancia, el valor de los análisis de las medidas tiene que ser evaluado en términos de la capacidad de generar mejores datos y no sólo de evaluar los datos existentes. Sin embargo, es importante reconocer el valor independiente de las evaluaciones de bases de datos existentes, especialmente en el caso de bases de datos, como los índices sobre la democracia aquí discutidos, que son frecuentemente utilizados en ejercicios de evaluación causal tanto en el campo de las relaciones internacionales como el de la política comparada, pero que aún no han sido objeto de atención suficiente. 

REFERENCIAS BIBLIOGRÁFICAS

- Adcock, Robert y Collier, David (2001), "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," en *American Political Science Review*, vol. 95, núm. 3, pp. 529-546.
- Álvarez, Michael; Cheibub, José Antonio; Limongi, Fernando y Przeworski, Adam (1996), "Classifying Political Regimes," en *Studies in Comparative International Development*, vol. 31, núm. 2, pp. 1-37.
- Arat, Zehra F. (1991), *Democracy and Human Rights in Developing Countries*, Boulder, CO, Lynne Rienner.

- Bollen, Kenneth A. (1980), "Issues in the Comparative Measurement of Political Democracy," en *American Sociological Review*, vol. 45, núm. 2, pp. 370-390.
- (1986), "Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984," en *Human Rights Quarterly*, vol. 8, núm. 4, pp. 567-591.
- (1989), *Structural Equations with Latent Variables*, Nueva York, Wiley.
- (1991), "Political Democracy: Conceptual and Measurement Traps," en Alex Inkeles (ed.), *On Measuring Democracy: Its Consequences and Concomitants*, New Brunswick, NJ, Transaction, pp. 3-20.
- (1993), "Liberal Democracy: Validity and Method Factors in Cross-National Measures," en *American Journal of Political Science*, vol. 37, núm. 4, pp. 1207-1230.
- (2001), *Cross-National Indicators of Liberal Democracy, 1950-1990* (<http://www.icpsr.umich.edu:8080/ABSTRACTS/02532.xml?format=ICPSR>).
- Bollen, Kenneth A. y Paxton, Pamela (2000), "Subjective Measures of Liberal Democracy," en *Comparative Political Studies*, vol. 33, núm. 1, pp. 58-86.
- Carmines, Edward G. y Zeller, Richard A. (1979), *Reliability and Validity Assessment*, Beverly Hills, CA, Sage.
- Collier, David y Adeock, Robert (1999), "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts," en *Annual Review of Political Science*, vol. 2, pp. 537-565.
- Coppedge, Michael, (1997), "Modernization and Thresholds of Democracy: Evidence for a Common Path and Process," en Manus I. Midlarsky (ed.), *Inequality, Democracy, and Economic Development*, Nueva York, Cambridge University Press, pp. 177-201.
- (1999), "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics," en *Comparative Politics*, vol. 31, núm. 4, pp. 465-476.
- Coppedge, Michel y Reinicke, Wolfgang H. (1991), "Measuring Polyarchy," en Alex Inkeles (ed.), *On Measuring Democracy: Its Consequences and Concomitants*, New Brunswick, NJ, Transaction, pp. 47-68.
- Dahl, Robert (1971), *Polyarchy*, New Haven, CT, Yale University Press.
- Elkins, Zachary (2000), "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations," en *American Journal of Political Science*, vol. 44, núm. 2, pp. 287-294.

- Elklit, Jørgen (1994), "Is the Degree of Electoral Democracy Measurable? Experiences from Bulgaria, Kenya, Latvia, Mongolia and Nepal," en David Beetham (ed.), *Defining and Measuring Democracy*, Thousand Oaks, CA, Sage, pp. 89-111.
- Foweraker, Joe y Krznic, Roman (2000), "Measuring Liberal Democratic Performance: An Empirical and Conceptual Critique," en *Political Studies*, vol. 48, núm. 4, pp. 759-787.
- Freedom House (2000), *Annual Survey of Freedom Country Scores, 1972-73 to 1999-00* (<http://freedomhouse.org/ratings/index.htm>).
- Gasiorowski, Mark J. (1996), "An Overview of the Political Regime Change Dataset," en *Comparative Political Studies*, vol. 29, núm. 4, pp. 469-483.
- Gastil, Raymond D., ed. (1978), *Freedom in the World: Political Rights and Civil Liberties, 1978*, Boston, G.K. Hall.
- (1991), "The Comparative Survey of Freedom: Experiences and Suggestions," en Alex Inkeles (ed.), *On Measuring Democracy: Its Consequences and Concomitants*, New Brunswick, NJ, Transaction, pp. 21-46.
- Gehrlich, Peter (1973), "The Institutionalization of European Parliaments," en Allan Kornberg (ed.), *European Parliaments in Comparative Perspective*, Nueva York, D. McKay, pp. 94-113.
- Gifi, Albert (1990), *Nonlinear Multidimensional Analysis*, Nueva York, Wiley.
- Gleditsch, Kristian S. y Ward, Michael D. (1997), "Double Take: A Reexamination of Democracy and Autocracy in Modern Politics," en *Journal of Conflict Resolution*, vol. 41, núm. 3, pp. 361-383.
- Gurr, Ted Robert; Jagers, Keith y Moore, Will H. (1991), "The Transformation of the Western State: The Growth of Democracy, Autocracy, and State Power since 1800," en Alex Inkeles (ed.), *On Measuring Democracy: Its Consequences and Concomitants*, New Brunswick, NJ, Transaction, pp. 69-104.
- Guttman, Louis (1977), "What is not What in Statistics," en *Statistician*, vol. 26, núm. 2, pp. 81-107.
- (1994), *Louis Guttman on Theory and Methodology: Selected Writings*, Brookfield, VT, Dartmouth Publishing.
- Hadenius, Axel (1992), *Democracy and Development*, Cambridge, Cambridge University Press.

- Jacoby, William G. (1991), *Data Theory and Dimensional Analysis*, Newbury Park, CA, Sage.
- (1999), "Levels of Measurement and Political Research: An Optimistic View," en *American Journal of Political Science*, vol. 43, núm. 1, pp. 271-301.
- Jäggers, Keith y Gurr, Ted Robert (1995), "Tracking Democracy's Third Wave with the Polity III Data," en *Journal of Peace Research*, vol. 32, núm. 4, pp. 469-482.
- Kaplan, Abraham (1964), *The Conduct of Inquiry. Methodology for Behavioral Science*, Scranton, PA, Chandler.
- Marshall, Monty G. y Jäggers, Keith (2001a), *Polity IV Project. Political Regime Characteristics and Transitions, 1800-1999. Dataset Users Manual*. (<http://www.bsos.umd.edu/cidcm/polity/>).
- Marshall, Monty G. y Jäggers, Keith (2001b), *Polity IV Project. Political Regime Characteristics and Transitions, 1800-1999. The Polity IV dataset*. (<http://www.bsos.umd.edu/cidcm/polity/>).
- Mokken, Robert J. (1971), *A Theory and Procedure of Scale Analysis with Applications in Political Research*, Berlín, Walter de Gruyter.
- Przeworski, Adam y Teune, Henry (1970). *The Logic of Comparative Social Inquiry*, Nueva York, Wiley.
- Ryan, Joseph E. (1994), "Survey Methodology," en *Freedom Review*, vol. 25, núm. 1, pp. 9-13.
- Vanhanen, Tatu (1993), "Construction and Use of an Index of Democracy," en David G. Westendorff y Dharam Ghai (eds.), *Monitoring Social Progress in the 1990s. Data Constraints, Concerns and Priorities*, Aldershot, UNRISD/Avebury, pp. 301-321.
- (1997), *Prospects of Democracy: A Study of 172 Countries*, Nueva York, Routledge.
- (2000a), "A New Dataset for Measuring Democracy, 1810-1998," en *Journal of Peace Research*, vol. 37, núm. 2, pp. 251-265.
- (2000b), *The Polyarchy Dataset: Vanhanen's Index of Democracy* (<http://www.svt.ntnu.no/iss/data/vanhanen>).