

La medición en la investigación comparativa

A. Przeworski y H. Teune

Tomado de The Logic of Comparative Social Inquiry; Cap. 5, pp. 91-112. John Wiley & Sons, EUA, 1970. Traducción de Guillermina Feher.

El lenguaje cotidiano posee una estructura comparativa implícita. Decir que la economía de los Estados Unidos es “competitiva” implica que este país puede diferenciarse de otros de acuerdo con esta propiedad o que, por lo menos, es posible que una nación carezca de una economía competitiva. Los enunciados descriptivos comparativos son cada vez más frecuentes en las ciencias sociales, pese a que su uso no vaya acompañado de una puesta en práctica cuidadosa. Tanto las poblaciones como los sistemas están sujetos a la comparación. El contraste de puntuaciones de muestras de poblaciones nacionales constituye una forma más sistemática que la de las caracterizaciones previas, y más generales, conferidas a tales poblaciones en la literatura de carácter nacional. Las caracterizaciones de los sistemas sociales o políticos no se rigen tanto por rudimentos del método científico, como sucede con aquellas sobre poblaciones nacionales cuyos criterios de muestreo están disponibles. Los criterios argumentativos y autoritarios predominan en la evaluación de propiedades “autoritarias”, “totalitarias”, “democráticas”, “tradicionales”, “modernas” y “en vías de desarrollo” de los sistemas políticos y sociales.

Dicho tipo de comparaciones procede, por lo general, sin un lenguaje explícito de comparación, lo que conlleva que las discusiones de comparaciones específicas se engloben en un conjunto vago de criterios conocidos como “igual”, “idéntico”, “similar”, “paralelo”, “equivalente” y “de la misma extracción”. El lenguaje científico debe ser explícito, siendo la medición un lenguaje de este tipo, que a su vez puede emplearse como un metalenguaje de comparación. Sin embargo, cabe preguntarse: ¿cuáles son los ajustes que deberán realizarse en el lenguaje de la medición con el objeto de afrontar los problemas persistentes de mensurabilidad detectados en la investigación comparativa? Los problemas de esta naturaleza surgen en la investigación comparativa, principalmente, por la necesidad de incorporar características contextuales de sistemas complejos en el lenguaje de medición.

En el presente capítulo se discutirá si es el lenguaje de medición científica, imprescindible para la formulación de teorías, puede incorporarse en el contexto de sistemas específicos; y, si así fuere, se verá cómo el contexto social puede integrarse a los postulados de medición sin destruir la uniformidad y generalidad del lenguaje de la medición.

El lenguaje de la comparación

La similitud y la universalidad son preceptos que deben existir al formularse una teoría. Dicha regla se deriva de las justificaciones empíricas de la investigación científica. En la

medición, el precepto consistirá en la diferencia y habrá de derivarse del fundamento fenomenológico de la ciencia. Sin embargo, si bien la ciencia es general y cuantitativa en sus conclusiones, es específica y cualitativa en sus observaciones.

Los científicos sociales realizan sus observaciones en sistemas sociales específicos. Estos se codifican directamente en un lenguaje con reglas directas de interpretación, tal como “esto es una manifestación”, o bien, se incorporan indirectamente a un lenguaje mediante reglas de inferencia, tal como “las manifestaciones en este sistema son una forma de descontento social”. El uso de reglas de inferencia se encuentra obstaculizada por dificultades que, con frecuencia, se perciben como inherentes a la medición comparativa.

El argumento que se opone a la utilización de reglas generales de inferencia en la medición comparativa plantea que el único marco adecuado para evaluar las características de los fenómenos sociales debe derivarse de los sistemas en los que se efectuaron las observaciones. Morris – Jones afirma que, para el mundo occidental algunos modos de comportamiento político de la India son claramente corruptos, mientras que en el uso idiomático del sistema político indio la misma conducta es una expresión de lealtad. Estos argumentos deben considerarse con seriedad. En el caso de ciertas observaciones específicas, un eructo es un eructo y el nepotismo es nepotismo; más, en el interior de un marco de inferencia, un eructo es un “insulto” o un “halago” mientras que el nepotismo es “corrupción” o “responsabilidad”. Un lenguaje de medición directa requiere únicamente de una gramática y de reglas de interrelación empírica que permanezcan sin variación, para toda cultura o sociedad. El lenguaje de medición inferencial requerirá, además de enunciados generales que definan el significado, de una observación específica en términos de su contexto sistémico.

El que dos o más fenómenos sean “comparables” dependerá de que sus propiedades se hayan expresado en un lenguaje estándar. Un lenguaje de medición define clases de fenómenos, al proporcionar criterios específicos que permiten decidir si una observación puede asignarse a una clase particular, y asimismo ordena las relaciones existentes entre dichas clases. Se tratará de un lenguaje estándar cuando pueda aplicarse consistentemente a cualquier individuo o unidad social. A fin de comparar la modernidad del campesino indio y del hombre de negocios estadounidense, se asignarán a ambos algunos de los términos de un lenguaje estándar. La clasificación de las observaciones en categorías, su graduación o conteo sirven para expresar las observaciones en un lenguaje de medición. Dictaminar sobre si el sistema político de un país es democrático mientras que el de otro es totalitario, ordenar países en términos del grado de su desarrollo político, o contar el número de personas que participan en elecciones formales constituye una medida de una propiedad del sistema político. Tales observaciones serán, en efecto, comparables al expresarse en un lenguaje estándar. El metalenguaje que determina la comparabilidad es el lenguaje de la medición.

Como resultado de esta interpretación de la medición en la investigación comparativa, algunos criterios de comparabilidad llegarán a estar disponibles. En primer lugar, el lenguaje en que se expresan las observaciones debe contener reglas de interpretación empírica y debe ser aplicable de manera uniforme a todas las observaciones. En segundo lugar, este lenguaje debe especificar las clases o magnitudes que pueden asignarse a las observaciones. Por último, deberán enunciarse las relaciones entre clases y, por tanto, las

transformaciones admisibles. Cuando estos criterios se cumplen, será factible afirmar que, un mayor número de indios que de estadounidenses concuerda con una cuestión, o que los resultados electorales en los Estados Unidos son menores que en Noruega. Un lenguaje de medición que exprese la relación entre clases de objetos implica la selección de un modelo de medición. Cualquier lenguaje es indiferente a la realidad a la que se aplica. Un lenguaje de graduación puede aplicarse tanto a una jerarquía de clases sociales como a una de ángeles; en este caso, las clases de este lenguaje estarían vacías empíricamente. La cuestión de la similitud o el isomorfismo de un lenguaje respecto a un fenómeno observado suele estar abierto a la discusión. Las reglas para la interpretación empírica deben ser lo suficientemente explícitas para que el proceso de asignación de observaciones a categorías no sea ambiguo. Los enunciados de medición deben mantenerse estables a través de las diferentes observaciones del mismo fenómeno en el mismo lenguaje.

El uso de un lenguaje estándar no acarrea problemas inherentes para la investigación comparativa. Será de esperarse, naturalmente, que existan problemas al asignar observaciones en términos de un lenguaje formal, y tales dificultades podrán ser de hecho más numerosas en una sociedad que en otra. El problema de la medición en la investigación comparativa surge cuando la validez de las reglas de inferencia es relativa a los sistemas en los que se realizaron las observaciones. Aún se encuentra en discusión la estructuración del marco que permita hacer inferencias respecto a la conducta cortés observada en una persona que hace una reverencia en Egipto. Para el caso de la medición directa, el problema estriba en la aplicación estable de las reglas de interpretación empírica de un lenguaje dado. Este es el problema de la confiabilidad. Sin embargo, cuando se realizan inferencias, el problema de la confiabilidad se conjuga con el de la validez. A menos que logremos contar con reglas estándar, los contextos culturales o sociales en los que se efectúan las observaciones distorsionarán la validez de las inferencias.

Medición directa versus medición inferida

Existen dos procedimientos básicos mediante los cuales se incorporan las observaciones en un lenguaje estándar: el que requiere de la inferencia y el que no la requiere. Para que un fenómeno ingrese a un sistema lógico, deberá cumplir con un conjunto de operaciones. En un caso, se definirá el fenómeno en términos de las operaciones que se utilizan para medirlo. El concepto y el procedimiento de evaluación se igualan. En la investigación de encuestas, la identificación con un partido se define en ocasiones como una respuesta a la pregunta acerca de si la persona se considera un demócrata arraigado, un demócrata débil, un independentista, un republicano indeciso o un republicano decidido. La mayor parte de los conceptos que se emplean en este tipo de investigación, en especial cuando las respuestas a ciertas preguntas se relacionan con otras, suele medirse de esta forma, además de que es comúnmente utilizado en estudios internacionales que usan datos agregados. Por ejemplo, si el concepto aparece como “el porcentaje de hablantes del idioma dominante” y la operación de medición consiste en sumar los individuos que hablan cada uno de los idiomas, la definición y la operación de medición se habrán igualado. Una definición

operacional es aquella en la que el nivel de generalidad o el “dominio” de un concepto es contérmino con el resultado de una operación de medición; este tipo de medición será directa.

Las definiciones que se igualan a las operaciones de medición pueden ser complejas y se forman, por lo general, al combinar varias medidas directas. Por ejemplo, la estabilidad política puede definirse como la suma equilibrada del número de cambios de jefes del poder ejecutivo, del número de golpes de estado intentados y del número de revueltas. Aunque suele llamarse índice a este tipo de definición, no lo es en realidad en tanto no indica, sino que define. La generalidad de un concepto no trasciende los resultados de operaciones particulares de medición.

Dado que en la medición directa las operaciones suelen ser sinónimos del concepto, sus resultados serán necesariamente válidos. Si la identidad política se define como la respuesta a una pregunta específica, la medición será tautológicamente válida. Sin embargo, ya que existe la posibilidad de que se hayan cometido errores al asignar los términos del lenguaje a las observaciones, la confiabilidad deberá cuestionarse y corroborarse.

La importancia teórica de los conceptos definidos en función de operaciones de medición específica se encuentra limitada, puesto que la generalidad de dichos conceptos es relativamente baja y, con frecuencia, característica respecto a un sistema social. Una teoría sobre la votación será menos general, y, por tanto, menos significativa, desde un punto de vista teórico, que una teoría sobre participación política. El acto de votar puede utilizarse como un fenómeno específico que indica cierto nivel de participación política, y no como un fenómeno que define un grado de participación en la política.

Debido a que un gran número de conceptos de relevancia teórica general no pueden definirse satisfactoriamente mediante una operación de medición específica, será necesario inferir las observaciones específicas a los fenómenos generales. A su vez, los fenómenos específicos se abordarán como indicadores o apuntadores. Las inferencias se basan sobre leyes generales de comportamiento. La interdependencia del volumen de los objetos físicos y el calor puede emplearse en la medición por inferencia. El volumen del mercurio se utiliza para medir la temperatura del aire ambiental bajo condiciones prescritas. El conocimiento de la ley general en psicología, que postula que las personas con cierta disposición, la manifestarán verbalmente, nos servirá para realizar inferencias acerca de las motivaciones sobre logros que tiene un individuo si se atiende al número de veces que alude a temas de logros. Ante el supuesto general de que un gobierno eficaz aumenta los servicios para la comunidad, el número de camas en un hospital será el fundamento para inferir la eficacia gubernamental.

Cuando la medición se basa en la inferencia, se supone, por lo común, que será más general que cualquier conjunto específico de indicadores que se haya empleado efectivamente en la medición. Sin embargo, cuando tanto el fenómeno nombrado por un concepto como el fenómeno que se aborda como indicador pueden observarse directamente, será factible realizar inferencias sobre la base de este único indicador. Por ejemplo, la clase social puede definirse mediante la posición relativa del individuo en el interior de su comunidad, expresada como un puntaje de participación/observación. Surgirá entonces, la pregunta acerca de la manera en que se podrán reducir las operaciones costosas de observación

mediante el uso de indicadores simples. Warner advirtió que pueden realizarse predicciones sobre una clase social que se ha definido de esta manera tomando como base algunos indicadores como escolaridad, lugar de residencia, ingreso y ocupación. En algunos casos los indicadores podrían reducirse aún más a una única manifestación de clase, como lo sería el material empleado para cubrir el techo de una casa residencial. La prueba psicométrica respecto a la “capacidad de producir focos de luz”, tema proverbial de envidia entre los científicos sociales, puede fácilmente validarse en relación con el desempeño real. Cuando el indicador y el concepto se pueden observar independiente y directamente, podrá establecerse la validez de los criterios de las inferencias. Este tipo de medición continuaría siendo directa, dado que las inferencias empleadas en él no se fundamentan sobre supuestos derivados de una teoría, sino a partir de inferencias que simplemente se evaluaron en términos de la dependencia estadística de los fenómenos directamente observables.

Una situación diferente surge cuando no es posible observar directamente el fenómeno denotado por un concepto. Será entonces necesario hacer uso de un conjunto de indicadores que, suponemos, se encontraría relacionado con el fenómeno si fuera observado. Las “propiedades de disposición” se corroboran de esta manera, mientras que los rasgos psicológicos como el autoritarismo, y los caracteres sistémicos, como la integración, se corroboran mediante inferencias.

La medición directa en la investigación comparativa

Las mediciones directas se basan en definiciones por decreto, y son de la forma: “sea S el resultado del procedimiento de observación X expresado en este lenguaje”. Por ejemplo, “sea el nivel de desarrollo económico el producto nacional per cápita, de acuerdo con las características registradas y expresadas en dólares estadounidenses”. En su estudio sobre Japón y los Estados Unidos, Kuroda definió “las aspiraciones al poder político” como una respuesta a la pregunta: “¿le gustaría ejercer mayor (o menor) influencia?”. Aun cuando la observación directa constituye el fundamento de toda medición, las definiciones por decreto se encontrarán limitadas a dichas observaciones. Cuando estas últimas se formulan en un lenguaje estándar, como anteriormente se definieron, el resultado será un conjunto de enunciados de medición que permitirá la comparación de todos los casos descritos en este lenguaje.

La medición directa requiere que el lenguaje de medición sea común a todas las observaciones, refleje las relaciones entre los fenómenos observados y sea de aplicación consistente. Los problemas del uso, incluyendo los lenguajes estándar más simples, se incrementan en un contexto comparativo. Repetiremos una vez más que, toda medición requiere de un lenguaje común con reglas de interpretación estándar. Por ejemplo, sea un “político” aquella persona que tiene el derecho formal de participar en la toma de decisiones autoritarias, obligatorias para cierta comunidad; y sea el “administrador” aquel que sólo puede tomar decisiones en asuntos prescritos por reglas y reglamentos. En la medida en que la jerga de los sistemas políticos particulares define estas posiciones de

manera diferente, la comparación entre la proporción de políticos y de administradores será altamente problemática. Una alternativa consistiría en el uso de estándares específicos para cada país. Estos estándares podrían fundamentarse en las estipulaciones constitucionales de la India o Yugoslavia. Sin embargo, la utilización de estos sistemas de reglas específicas imposibilitaría la comparación de tales porcentajes debido a que no se ha aplicado un lenguaje común. Los lenguajes específicos de medición sólo conducirían a enunciados descriptivos, tales como “en el contexto yugoslavo hay tantos administradores y tantos políticos”. Se tendrá entonces que, esta formulación sólo será comprensible si se conoce el contexto de Yugoslavia.

Los requerimientos de la medición directa satisfacen el deseo intuitivo de obtener la “verdadera” medida, la verdadera cifra de políticos y administradores. Pero la naturaleza concreta y específica de dicha medición constituye un obstáculo para la posibilidad de hacer generalizaciones entre naciones. En la medición directa no pueden efectuarse inferencias a partir de listas concretas de administradores y políticos en el contexto indio y yugoslavo, hacia el contexto general subyacente del papel político que desempeñan los administradores y políticos. Las definiciones por decreto son arbitrariamente verdaderas, aunque también son menos generales. El segundo requerimiento de la medición directa consiste en que el lenguaje debe expresar las relaciones entre clases de fenómenos. Suppes y Zinnes afirman que el proceso de medición debe “caracterizar las propiedades formales de las operaciones y relaciones empíricas empleadas en el procedimiento, y deberán además mostrar que son isomórficas en las operaciones y relaciones numéricas que se seleccionaron adecuadamente”.

Regresemos al ejemplo de los políticos y administradores. Una vez que se ha enunciado la definición de dichos conceptos en términos comunes a todos los sistemas, se procederá a escoger la gramática del lenguaje. Puede ser posible que un individuo no pueda desempeñar ambos roles en un país debido a prohibiciones legales o tradicionales, en tanto que en otra nación podría darse el caso también, de que no existieran restricciones de este tipo. En el primer ejemplo, la gramática más sencilla sería disyuntiva: si A, entonces no B, y si B, entonces no A.

En el segundo, la gramática requerida sería alternativa: A y no B, y B y no A, y A y B. De hecho, no sería legítimo emplear la primera gramática para ambos países, si bien la segunda sí sería adecuada para los dos. En términos de esta gramática general, sería cuestión de indiferencia el que un país prohibiera los puestos duales y el otro no. El uso de la gramática alternativa no distorsiona las características de las observaciones del sistema en el que una persona no puede ocupar ambos puestos. Si se aboliera la prohibición en el primer país y se estableciera en el segundo, la gramática seguiría siendo adecuada. La naturaleza de los datos y las reglas de interpretación contenidas en el lenguaje interactúan en la necesidad de una gramática particular.

Según el tercer requerimiento, las reglas de interpretación empírica, en términos de un lenguaje, deben aplicarse consistentemente a cualquier observación. En nuestro ejemplo, estas reglas se manifiestan a partir de las definiciones de políticos y administradores; más, en la medida en que no se aplican consistentemente, pierden su confiabilidad. Por otra parte, en razón de que su aplicación consistente tiende a concentrarse en un país, en lugar

de encontrarse distribuidas en todos, habrá ahí un error sistemático en las observaciones. Este error deberá eliminarse si pretendemos que los descubrimientos sean confiables.

Es posible utilizar diferentes reglas de interpretación y además traducirlas. Por ejemplo, si las posiciones A, B,..., N se codificaran para el rubro “político” en un conjunto de observaciones, y las A, B,..., N, O, P lo fueran para otro conjunto, no habría dificultades para hacer que ambos conjuntos de reglas fuesen intercambiables mediante el añadido de O y P, en el primer conjunto, o restando estas dos posiciones del segundo. A menos que ello se efectúe, ninguno de los dos conjuntos de observaciones podrá emplearse en el análisis comparativo.

La mayor parte de la discusión anterior se basó sobre la escala nominal de medición. No obstante, la medición directa puede formularse utilizando otros tipos de escalas. En las escalas ordinales, los objetos se ordenan sobre una magnitud subyacente. Cuando todos los objetos se categoricen conjuntamente, sin importar el sistema social al que pertenezcan, la medición se expresará en un lenguaje estándar, esto es, en un lenguaje ajeno a cualquier sistema dado. Las comparaciones son legítimas en el interior de este lenguaje. La aplicación típica de las escalas ordinales se ejemplifica por los estudios de exposición a los medios masivos de comunicación, en general, y a las noticias políticas, en particular. Se preguntó a los sujetos en diferentes países con qué frecuencia seguían las noticias políticas: “nunca”, “una vez al mes”, “por lo menos una vez a la semana” o “diariamente”. Las frecuencias de los individuos que se situaron en estas categorías ordenadas se registraron a continuación por cada uno de los sistemas y se compararon. Dichas comparaciones son legítimas sólo si la ordenación se realiza mediante magnitudes que cuenten con una interpretación física, más que con una meramente psicológica. En consecuencia, si la frecuencia de exposición se mide en términos de tiempo físico (días, semanas y meses), podrá suponerse que el lenguaje es estándar. Sin embargo, si las categorías emplean tiempo psicológico (“rara vez”, “de vez en cuando”, “frecuentemente”), la ordenación probablemente no será invariante con respecto a la cultura. El argumento implícito aquí requiere la autocomparación del sujeto con los demás miembros del grupo o con las normas de éste. Para un profesor universitario la lectura semanal de un diario podría caer bajo el rubro “rara vez” mientras que un campesino podría percibirla como “frecuentemente”. La inmutabilidad del orden de categorías que se emplea en las mediciones ordinales, no suele ser un tema explícito de discusión en las investigaciones transnacionales, razón por la cual una gran parte de las comparaciones transnacionales presenta numerosas dudas.

En las escalas de intervalo, el problema consiste en asignar números a diversas marcas de un instrumento de medición. El sistema numérico expresa intervalos iguales de la magnitud implícita. La “escala auto-anclante” de Cantril, diseñada para medir magnitudes de esperanzas y aspiraciones del individuo y del país, es una de las escalas de dicho tipo que más ampliamente se ha utilizado en investigaciones transnacionales. La escala se diseñó para permitir que cada sujeto, sin importar su cultura o país, definiera personalmente su lugar en el presente, dónde estuvo en el pasado y dónde piensa que se encontrará en el futuro. Se le presentan once categorías a manera de “escalera”, numeradas del “0” al “10”. Se le pregunta al sujeto: “Aquí ve usted una escalera. Supongamos que la parte superior de la escalera (señalando) representa la mejor vida posible para usted y la parte inferior (señalando) representa la peor vida posible para usted. ¿En qué parte de la escalera

(desplazando el dedo hacia arriba y hacia abajo rápidamente) piensa usted personalmente se encuentra en el presente?” La misma pregunta se utilizó para obtener la percepción del individuo acerca de la situación en la que se encontraba hace cinco años y en la que esperaba encontrarse en el futuro. Al describir el instrumento, Cantril afirmó:

Debe subrayarse hasta el cansancio que, las categorizaciones que los individuos se asignan, o asignan a su país, son totalmente subjetivas. Por tanto, una asignación de ‘6’ dada por una persona de ninguna manera indica lo mismo que el ‘6’ indicado por otra persona. Esta aclaración obvia, la menciono aquí porque la experiencia ha mostrado que algunos individuos no comprenden la lógica global de esta técnica y suponen que la escala es como una prueba de inteligencia en la que una categorización dada cuenta con una connotación precisa y, aparentemente en cierta medida, universal.

La escala “autoanclante” es un modelo de magnitud. Cada uno de los 11 puntos de la escalera recibirá un número que representa una cantidad única, conocida y constante, del todo de esperanzas y aspiraciones. A pesar de la advertencia, se le suele manejar como si tuviese propiedades de intervalo. Un ‘6’ como respuesta en la India, los Estados Unidos e Israel se maneja como un “6” en lenguaje estándar. De otra manera, no podría efectuarse una comparación entre las medias de los países. Las observaciones expresadas en el lenguaje de la escala constituyen mediciones directas en las que el lenguaje se aplica en forma consistente transculturalmente. Y, pese a ello, el lenguaje de medición no es estándar dado que es claro que un “6” puede expresar una cantidad en un país y una cantidad diferente en otro.

La diversidad de contextos sociales en los que se conducen las mediciones impone, por lo general, serias limitaciones a los procedimientos de medición directa, ya que ésta depende inmediatamente de observaciones específicas y por tanto es altamente sensible a las diferencias de los contextos sociales.

Mediciones inferenciales: validez del sistema e interferencia sistémica

Los postulados de la medición inferida se fundamentan en lenguajes sociales y conductuales. Al evaluar las disposiciones psicológicas, las respuestas a las preguntas se emplean para inferir magnitudes o actitudes. Así, por ejemplo, los lapsus lingüísticos se utilizan para inferir ansiedad. Al evaluar las características de las colectividades, las revueltas pueden usarse para inferir inestabilidad política y los comentarios fortuitos de los diarios oficiales proporcionan la base para inferir un cambio en la política. En este tipo de medición los conceptos se definen nuevamente por decreto; sin embargo, las observaciones directas se combinan con leyes generales con objeto de inferir la presencia o la magnitud de las propiedades. La definición del concepto no es coextensiva con ningún conjunto de observaciones directas, pues éstas, por el contrario, incorporan enunciados de medición indirectamente, mediante la inferencia formulada con propuestas generales.

En este contexto las observaciones directas indican significado, aunque no prescriben sus límites.

La medición inferida tiene los siguientes componentes:

Definición: Sea S definida como X .

Proposición: Todos los objetos, y sólo esos objetos que tengan la propiedad N , probablemente (o ciertamente) tendrán las propiedades x_1, x_2, \dots, x_k .

Procedimientos de medición directa: x_1, x_2, \dots, x_k pueden observarse de las siguientes maneras.

Regla de composición: x_1, x_2, \dots, x_k interactúan de tal forma que pueden combinarse de acuerdo con la(s) siguiente(s) regla(s).

Regla de inferencia a la propiedad: Cualquier objeto que se haya observado contenga una cierta cantidad de x_1, x_2, \dots, x_k tendrá probablemente (o ciertamente) a contar con una cantidad específica de la propiedad X .

El resultado de la combinación de las observaciones directas x_1, x_2, \dots, x_k constituye la base para inferir la magnitud de la propiedad X ¹. Una regla de inferencia que se utiliza por lo común es la monótona lineal, que postula que cualquier instrumento en la puntuación indicará un incremento correspondiente unívoco en la propiedad. No obstante, esta regla de correspondencia es sólo una entre tantas. La relación podría ser monótona y no lineal; esto es, podría tratarse de una función al escalón² o curvilínea. Algunos de los pasos mencionados no suelen explicitarse en la práctica. Aun así, cada uno de los componentes está presente en cualquier evaluación por inferencia. Un enunciado de medición es, por lo tanto, una proposición singular que enuncia que un objeto específico de observación posee una cantidad dada de una cierta propiedad. Por ejemplo, enunciados como “Panamá tiene un sistema político inestable” y “el sujeto 173 tiene un puntaje de autoritarismo de 55 unidades de intervalo” son enunciados de medición, ya que se infirieron a partir de enunciados que registran los resultados de las observaciones directas x_1, x_2, \dots, x_k sobre la base de las proposiciones generales y de la regla de inferencia. Tenemos por ende que un enunciado específico de medición se obtiene de la siguiente manera: Enunciado de medición directa: un objeto A obtiene la puntuación de a como resultado de la combinación de las mediciones directas x_1, x_2, \dots, x_k . Proposición general que incluye una regla de inferencia: todos los objetos, y sólo esos objetos, que tengan una cantidad de a' de propiedad x obtendrán una puntuación de a como el resultado combinado de las operaciones x_1, x_2, \dots, x_k . Enunciado de medición inferida: por consiguiente, el objeto A tiene la cantidad a' de la propiedad N . La conclusión de esta inferencia es el enunciado de medición. La relación entre el puntaje a y la cantidad a' de la propiedad x se encuentra sujeta a las reglas de inferencia arriba mencionadas.

Las proposiciones empleadas para las inferencias deben percibirse en términos probabilísticos más que en términos deterministas³. En el primer caso, cada observación que se maneje como indicador contendría un componente “verdadero” y uno de “error”. Si

¹ Para una discusión acerca de las reglas de composición véase May Brodbeck, “Methodological Individualisms: definition and Reduction”, *Philosophy of Science*, 25, 1958.

² Step function: función que crece o decrece abruptamente de un valor a otro (Esta nota al pie no forma parte del texto original).

³ Estos son modelos probabilistas. Véase W. A. Scott, “Attitude Measurement”, en Gardner Lindzey y Elliot Aronson, eds., *Handbook of Social Psychology*, Vol. 2, Addison – Wesley, Reading, Mass. 1968.

las observaciones son precisas, el error será una función del valor verdadero probable de la ley social sobre la que se efectuó la inferencia. Si se supone que concordar con una pregunta indica una actitud, aunque también refleje consentimiento, los enunciados de medición contendrán entonces un error. A pesar de que la medición inferida reconozca los errores de inferencia, el lenguaje de la medición directa no expresará este tipo de error. La respuesta a una pregunta se maneja como si se tratase de una medición precisa, más aquello que la respuesta indica se expresará en términos probabilísticos. El lenguaje probabilístico de medición puede fundamentarse en variar leyes o hipótesis sobre diversos indicadores. La hipótesis a corroborar es que el puntaje de un indicador particular contribuye a un puntaje compuesto que corresponde a la magnitud de una propiedad. Si se pudieran observar la propiedad y los múltiples indicadores, sería entonces posible observar conceptos generales de cierta importancia teórica, como lo serían las aspiraciones, la estabilidad o el status, los que sólo pueden ser “vistos” mediante indicadores. Este hecho presenta dificultades especiales para la investigación comparativa.

El problema básico de la medición radica en el grado de validez que hay en inferir enunciados de medición a partir de registros de observaciones directas. Cuando no se cuenta con criterios externos claros, el investigador debe valorar indirectamente la validez de los indicadores. El cotejamiento, por lo menos en un inicio, será de juicio, o de “validez de apariencia”⁴. Los juicios respecto a la validez de apariencia son la base para manejar el número de camas de un hospital como un indicador de la eficacia gubernamental. La selección de ciertos indicadores no es completamente arbitraria en tanto que refleja la etapa de desarrollo de una teoría en un área dada de las ciencias sociales. Si se aceptara un mapa conceptual general en las ciencias sociales, podríamos contar con reglas para seleccionar indicadores. En la medida en que los diversos grupos de científicos operan con distintos marcos conceptuales y diferentes teorías que relacionan los conceptos con los indicadores, la validez de apariencia continuará siendo una cuestión de criterio personal o grupal. La pregunta acerca de si los resultados de una votación constituyen un indicador válido de presiones igualitarias, es en gran medida irresoluble. A lo mucho se podrá argumentar que este fenómeno observable se ha empleado con frecuencia para indicar un concepto diferente, el de la participación política y que, por lo mismo, no puede utilizarse como un indicador de presiones igualitarias en el interior de la misma teoría. El criterio empírico de “validez discriminante” se emplea aquí para aislar este tipo de fenómenos observables que indican sólo un concepto general⁵.

Además de la validez de apariencia, existe otro criterio común que implica la “validez interna” de los propios indicadores. En este caso, la confiabilidad de los indicadores se usa para inferir su validez. La lógica de este procedimiento estipula que si todos los x_1, x_2, \dots, x_k son indicadores de x , esto es, si cada indicador se relaciona con la presencia o magnitud de x , entonces los x_1, x_2, \dots, x_k deberán interrelacionarse. En otras palabras, si x_1 se

⁴ “face validity”.

⁵La idea fundamental de la validez discriminante es que los indicadores de la misma propiedad, medida de diferentes formas, deberá tener una interrelación más elevada que la de los indicadores de diferentes fenómenos medidos de la misma forma. Véase D. T. Campbell y D. W. Fiske, “Convergent and Discriminant Validation by the Multitrait- multimethod Matrix”, *Psychology Bulletin*, 56, 1959 .

relaciona positivamente con x y lo mismo sucede para x_2 , su dependencia mutua respecto a x deberá reflejarse en su interdependencia. Si tanto las camas de los hospitales como la recolección de impuestos se vinculan con la eficacia gubernamental, entonces los gobiernos eficaces deberán contar tanto con un gran número de camas para los hospitales como con una gran cantidad de impuestos recaudados. Si este supuesto fuera falso (los indicadores no se encuentran mutuamente inter-correlacionados), sería claro que en conjunto indican más que un concepto general. De esta manera, las inferencias basadas sobre la propuesta que postula que todos los indicadores son una función de una única propiedad N serían inválidas. Si el supuesto fuera correcto, una prueba de la validez de cualquier indicador único estribaría en detectar la manera en que se correlaciona con los demás indicadores.

Respecto al procedimiento, un conjunto de indicadores se establece: 1) cuando se postula un conjunto de relaciones de indicadores específicos con la propiedad general que se esté midiendo, 2) al recolectar datos para las unidades en estudio, 3) inter-correlacionando los indicadores (podrá emplearse cualquier número de expresiones o relaciones), 4) excluyendo los que no presenten una correlación elevada respecto a los demás indicadores y 5) mediante la combinación de los valores de los indicadores seleccionados, ya sea sumando, ponderando, etc., para obtener una medida compuesta de la propiedad general. En el siguiente capítulo se proporcionarán ejemplos de este procedimiento.

En un contexto comparativo o transistémico, se entenderá por validez aquello que estemos midiendo en cada sistema, tomando en consideración qué es lo que pretendamos medir. El instrumento diseñado para medir la participación política debe referirse a los fenómenos que conforman la participación política en la Unión Soviética y en los Estados Unidos. Sin embargo, la participación política de un soviético y un estadounidense obviamente no se expresa en términos de una misma conducta. Un soviético políticamente activo dirigiría su atención al éxito de los proyectos económicos y asistiría a cursos que le permitieran desarrollar su conciencia ideológica, mientras que un estadounidense haría cosas tales como portar un botón en la solapa o donar dinero para un partido político. Los sistemas sociales afectan la conducta de los indicadores, de tal manera que los conceptos inferidos de los indicadores deberán modificarse en concordancia con el sistema en que se observaron.

Si se tratara del concepto “volumen de transacciones”, el flujo de dinero que circulara entre los individuos y las instituciones podría ser un indicador adecuado en un sistema social, en tanto que en otro quizás fuera imperativo contabilizar el número de intercambios no monetarios. Si se tratara del concepto “bienestar comunitario”, la construcción de autopistas podría ser un indicador válido para un sistema, mientras que la inoculación del ganado bien podría serlo para otro. Si el concepto fuese “igualitarismo”, el sentimiento hacia los negros podría ser un indicador válido en una cultura y su contrapartida en otra lo constituirían, los sentimientos dirigidos a castas menesterosas. Estos son ejemplos de lo que denominaremos “interferencias del sistema”, factor al que nos enfrentaremos en cualquier aspecto de la investigación comparativa.

Las interferencias del sistema inciden cuando los enunciados de medición, que se infieran de los mismos enunciados de medición directa, no son válidos para todos los sistemas que se encuentran en investigación. Por ejemplo, la aseveración de un sujeto que declarara se opondría a que su hijo contrajera matrimonio con la hija de un simpatizante de cierto partido político, no tendría la misma validez en cualquier país para inferir “partidarismo

político” como una definición de “distancia psicológica o entre partidos”⁶. Esta inferencia podría ser válida en los Estados Unidos, mientras que en Italia habría una actitud implícita de autoritarismo familiar, en vez de una distancia percibida entre partidos. En otras palabras, la respuesta “Sí, yo me opondría”, en los Estados Unidos podría indicar una alta probabilidad de distancia percibida entre partidos, en tanto que, en Italia, la objeción al matrimonio no la reflejaría.

Un sistema es un conjunto de elementos interdependientes, a tal grado que un cambio en cualquiera de los elementos modificará probablemente cualesquiera de los demás elementos. Pese a que los patrones de interacción entre los elementos de un sistema puedan describirse en términos generales, el comportamiento de cualquier componente o “rasgo” específico será determinado por otros componentes. Cuando los componentes o rasgos específicos se utilizan como indicadores de ciertos fenómenos generales, suelen extraerse del contexto del sistema al que pertenecen. Es obvio, por ejemplo, que el contexto sistémico determina si la votación es un indicador válido de participación. En un sistema político la votación puede resultar de una elección entre alternativas y, en otro, una manifestación de solidaridad. La interdependencia existente en el interior de un sistema constituye la médula de la postura funcionalista, que argumenta que los sistemas sociales o las culturas sólo pueden compararse en tanto unidades sociales, dado que la remoción de cualquier rasgo particular fuera de su contexto es no sólo engañosa sino inválida⁷.

En la investigación comparativa es necesario diversificar las proposiciones utilizadas para las inferencias con el objeto de que la validez de los enunciados de medición se preserve en cada uno de los sistemas. Ello sólo podrá realizarse al hacer referencia al contexto sistémico de las observaciones utilizadas, en un enunciado de medición inferida. Cuando se infiere el volumen de una substancia de la temperatura del aire que lo rodea, las reglas de inferencia deben ser relativas a la presión atmosférica. Aun cuando dichos ajustes no sean raros ni complicados para las ciencias naturales, los científicos sociales realizan sus observaciones, por lo común, en el interior del contexto de un sistema social y los problemas de las interferencias sistémicas casi nunca pueden evitarse. Incluso las observaciones de laboratorio, que se encuentran exentas de contexto en las ciencias del comportamiento, no están tan exentas, como alguna vez se pensara, y no es una cuestión de indiferencia atender en dónde se efectúen las observaciones.

Las consecuencias de las interferencias del sistema son sorprendentes sobre todo cuando se percibe que un fenómeno, que por lo común siempre está presente en un sistema, no puede detectarse en otro. Los estadounidenses no asisten a cursos de conciencia ideológica ni tampoco excavan fosas de estiércol, ni sacrifican animales en sus ceremonias religiosas. Aunque se trate de ejemplos de datos producidos por el sistema social, sin intervención de los esfuerzos de los científicos, el problema de la ausencia de fenómenos comparables no desaparece cuando el investigador genera sus propios datos. Ciertos interrogantes, tal como

⁶ La distancia psicológica entre partidos y su medida se discute en G. A. Almond y Sydeny Verba, op. cit.

⁷ A pesar de que el énfasis en la interrelación de los fenómenos sociales se remonte a los primeros funcionalistas, ello no significa que el uso del concepto “sistema” requiera de la acepción de “explicación funcional”. Nosotros utilizamos el concepto de “sistema” como cualquier conjunto distintivo de elementos interrelacionados y no solamente como un conjunto que contenga mecanismos de auto-regulación.

si los campesinos peruanos prefieren automóviles compactos o grandes, no pueden hacerse en ciertos sistemas sociales porque están desprovistos de significado y por tanto son inútiles como base de inferencia. En un sentido menos obvio, los investigadores con frecuencia crean fenómenos que no “existen”. Los indicadores aparecen entonces como un artificio de los procedimientos de observación, como sucede al plantear la pregunta de si un individuo “revisa más de dos veces si se está excediendo en sus tareas”, en una cultura en la que no se han definido las tareas individuales. El investigador casi siempre será capaz de obtener la varianza en el interior de un sistema cuando genere sus propios datos, incluso si sólo se trata de una varianza de error. La comprobación de que las inferencias en el lenguaje de medición, en efecto, se encuentren exentas de interferencias sistémicas será responsabilidad del investigador, puesto que él deberá sensibilizarse a las interferencias sistémicas, incluso en las poblaciones que presenten una homogeneidad evidente. Según ha observado Allardt, el desempleo en Finlandia puede indicar inseguridad en ámbitos rurales, mas no así en las ciudades. Finlandia no es un sistema único cuando se trata de emplear al desempleo como un indicador de inseguridad.⁸

Las mediciones transistémicas equivalentes

La cuestión de la equivalencia surge si y sólo si se presentan interferencias sistémicas y si la medición implica inferencia. Cuando el “jefe ejecutivo” se define como aquella persona que es el presidente, primer ministro o monarca, sólo cabe preguntarse si alguno de los tres puestos existe en un sistema dado. Si la votación se define como la introducción de una papeleta en una urna en cualquier proceso electoral formal desempeñado en términos legales, no tendría sentido el preguntarse si la introducción de la boleta es una medida “equivalente” al acto de votar. Cuando cualquier fenómeno se define en términos de la enumeración de sus manifestaciones observables (medición directa), no podrá siquiera formularse la pregunta de si la definición es equivalente a través de diferentes sistemas. Si se presentan interferencias de sistema y la medición se basa sobre la inferencia, las inferencias que sean válidas en un sistema no serán igualmente válidas en otro. El problema de la medición que se impone, al combinar la inferencia con las interferencias del sistema, estriba en que la validez de las mediciones de inferencia será específica para cada sistema social. Las inferencias que conllevan enunciados de medición deben validarse en cada sistema social. Debe, asimismo, considerarse la interacción de cada propiedad de los componentes de los sistemas con otras propiedades. En otras palabras, las inferencias en el lenguaje de medición deben considerar el significado, o el contexto, de un rasgo medido en términos del patrón de interacciones sistémicas de cada uno de los sistemas. Aun cuando en la formulación de teorías se preste énfasis a la generalidad más que a la precisión, en el caso

⁸ Eric Allardt, “Implications of Within Nation Variations and Regional Imbalances for Cross – National Research” en R. L. Merrit y Stein Rokkan, eds., *Comparing Nations: The Use of Quantitative Data in Cross – National Research*, Yale University Press, New Haven, Conn., 1966, pp. 337-58.

de la medición se tratará de la precisión, y no de la generalidad. Como se indicó en la primera parte, las propuestas específicas para el sistema que proporcionen mayor precisión deberán ser, por tanto, el punto dominante de la medición. Regresemos al ejemplo de la distancia psicológica entre partidos políticos. La conducta observable es la misma: en los diversos países los individuos responden a la misma pregunta en iguales condiciones. Como resultado de estas observaciones, cada individuo recibe un puntaje de partidismo político, esto es la medida en la que percibe la distancia entre los partidos.

País A

País B

La persona A respondió “Sí”.

La persona B respondió “Sí”.

Todas las personas, y sólo esas personas que perciben una gran distancia entre partidos, tenderán a responder (o ciertamente responderán) “sí”. Por tanto, es altamente probable que: La persona A sea un “partidario” y la persona B sea un “partidario”. La verdad de los últimos enunciados, los enunciados de medición inferida, depende de la verdad de la proposición general que estipula que las personas que afirmaron oponerse a que su hijo o hija contrajera matrimonio con un simpatizante de un partido distinto, tenderán a ser “partidarios”, sin importar el sistema al que pertenezcan. Pero si se presentan interferencias del sistema, la proposición general no será igualmente verdadera en ambas sociedades. La validez de la inferencia no es invariante entre los sistemas sociales.

Si estas mismas inferencias no son igualmente válidas para diferentes sociedades, ¿podrán compararse los enunciados de medición transistémicamente? Si las mismas inferencias conducen a diferentes resultados en diferentes sistemas sociales, ¿cómo podrán ajustarse las reglas de inferencia y los instrumentos de medición para producir resultados igualmente válidos en diferentes sistemas sociales? El problema es tan penetrante que ha obligado a algunos científicos sociales a negar la naturaleza nomotética de las ciencias sociales. Sin embargo, la validez de las inferencias puede hacerse relativa a cada sistema social y para ello deberán emplearse, obviamente, instrumentos basados sobre diferentes observaciones e inferencias. No obstante, si se diseñan distintos instrumentos para sistemas particulares, ¿cómo sabremos si lo que estamos midiendo es lo mismo para cada uno de los sistemas y es por tanto comparable? Además de su validez, los instrumentos deben ser confiables a través de los sistemas. En otras palabras, los instrumentos de medición deben proporcionar resultados consistentes, sin importar el sistema social en los que se hayan aplicado.

Resulta más sencillo corroborar la confiabilidad que la validez y existen numerosos métodos para calcularla. La mayor parte de los cálculos se basa sobre el mismo supuesto fundamental: la medición es confiable en la medida en la que la varianza de los puntajes “verdaderos” se aproxime a la varianza observada⁹. Si la varianza verdadera se define como

⁹ Para una breve discusión acerca de la confiabilidad y sus cálculos, véase F. N. Kerlinger, Holt, Rinehart, Foundation of Behavioral Research, Nueva York, 1964.

la varianza común, dispondremos de expresiones numéricas relativas a la confiabilidad. A pesar de que las distribuciones de las observaciones (el número real de observaciones) y los errores se conjuguen en el problema de calcular la confiabilidad, nosotros sólo deseamos saber si los resultados que proporciona el instrumento pueden ser reproducidos por este mismo. La medición que satisface las dos condiciones anteriormente mencionadas –validez en términos de cada sistema social y confiabilidad transistémica- será denominada “medición equivalente”. Debe subrayarse que, este concepto de equivalencia no se refiere a las observaciones sino a los resultados de las inferencias realizadas a partir de esas observaciones, esto es, los enunciados de medición inferida. Un instrumento es equivalente transistémicamente en la medida en la que los resultados que proporcione tal instrumento describan de manera confiable y con (casi) la misma validez un fenómeno particular en diferentes sistemas sociales. En un cuestionario no es relevante que los estímulos sean los mismos o que la pregunta se haya traducido adecuadamente de un lenguaje a otro. Lo que sí será importante, es que el planteamiento de la pregunta permita realizar inferencias válidas acerca de las mismas propiedades implícitas, sin importar el sistema social al que corresponda. La equivalencia de estímulos constituye un problema significativo, solamente, en la medida en que la medición no involucre inferencia¹⁰. En la medición inferida, la equivalencia de estímulos es de importancia secundaria. Deberá, por tanto, prestarse el énfasis a la equivalencia de las inferencias realizadas a partir de las observaciones.

Algunas nociones auxiliares son necesarias con el objeto de comprender este concepto de equivalencia. Si deseáramos medir la facilidad verbal de todas las personas en cada uno de los países del mundo, cada idioma podría, de un modo plausible, entrar en nuestra medición. El dominio del concepto, esto es, aquello a lo cual se refiere el concepto, estriba en la facilidad para usar todas las palabras, en cada uno de los países del mundo. La totalidad de palabras de todos los idiomas podría ser la población total de los indicadores, pero es evidente que la medición de la facilidad verbal que tengan los estadounidenses en Hindi no sería adecuado. Cada idioma cuenta con una población correspondiente de individuos, por lo que, cada idioma específico puede percibirse como un conjunto de indicadores que puede utilizarse para medir la facilidad verbal de un conjunto de individuos. Para construir un instrumento de medición de la facilidad verbal en cada idioma, se tomaría una muestra de palabras del idioma en cuestión, siendo ésta una subpoblación de indicadores. Si nos interesara medir la facilidad verbal de los estadounidenses, tomaríamos varias muestras de 100 palabras inglesas y probaríamos la cantidad que conoce cada individuo.

En términos generales, lo que medimos es una propiedad denotada por un concepto. Primero, definimos el dominio del concepto o, en otras palabras, especificamos la población de indicadores para cada uno de los sistemas. Dichos indicadores pueden no ser igualmente relevantes, o incluso existentes, en cada sistema social. Cuando no lo son, es necesario identificar la subpoblación de indicadores relativa a cada sistema social, para así

¹⁰ Por ende, según nuestro punto de vista, el énfasis tradicional que se da a la equivalencia de estímulos está equivocado. El entrenamiento de los entrevistadores, el tiempo, la manera en que se llevan a cabo las entrevistas, e incluso, la formulación de las preguntas, no deben ser idénticos. La equivalencia de estímulos debe percibirse meramente desde el punto de vista de la equivalencia de los enunciados resultantes de medición.

inferir muestras de indicadores a partir de la subpoblación que se empleará como el instrumento específico de cada sistema. Esta definición de equivalencia no requiere, en la práctica, la ubicación de la población total de indicadores. En el ejemplo anterior, el procedimiento necesitó, primeramente, la descripción de la población total de indicadores (todas las palabras en cada uno de los idiomas en uso) y, posteriormente, el examen acerca de qué subpoblación de este universo (idiomas específicos) se podría encontrar para cada uno de los conjuntos de individuos que pertenecen a los sistemas específicos. Si fuese imprescindible conocer el total de la población de indicadores, se requeriría efectuar una investigación substancial previa a la corroboración de cualquier propiedad. Es posible comenzar con sistemas sociales específicos y generar una muestra de indicadores para cada uno de ellos: “El autoritarismo en la India se indica por (...), en los Estados Unidos el autoritarismo se indica por...” Al finalizar el proceso, sería posible identificar los indicadores invariantes del conjunto total, transistémicamente, así como aquellos que sean relativos a sistemas particulares. Incluso en nuestro ejemplo, para la construcción de un instrumento para la facilidad verbal transistémica, podrá detectarse que, como consecuencia de la difusión, algunas palabras serán comunes a diversos idiomas.

Este procedimiento, para la construcción de mediciones equivalentes en diversos sistemas, permite que el contexto de un sistema se incorpore al proceso de medición. Enunciados inherentes al sistema, como: “el eructo es una expresión de deferencia en Egipto”, pasa a formar parte del instrumento de medición. Se considera que los indicadores hipotéticos de cada sistema constituyen una muestra de indicadores para cada sistema, en tanto que se supondrá que cada una de las muestras inherentes al sistema será una muestra de la población total de indicadores. Cuando algunos indicadores se consideren relevantes en todos los sistemas en estudio, podrán manejarse, al menos inicialmente, como si estuviesen “exentos de sistemas”. En aquellos casos en los que la muestra de indicadores contenga un conjunto superpuesto o exento de sistema, el procedimiento para establecer la equivalencia podrá consistir en combinar indicadores exentos de sistema, e inherentes al sistema, en una medición equivalente y, por ende, comparable. En el siguiente capítulo se discutirá este procedimiento.

Cuando las muestras de indicadores derivados de sistemas particulares no se sobreponen, nos encontramos ante una situación distinta. ¿Será aún posible manejar las puntuaciones basadas sobre dos muestras independientes de indicadores como si fueran equivalentes? Primeramente, si los puntajes se basaron sobre una muestra al azar de la población total de indicadores, sería posible traducir la puntuación de una muestra de indicadores a otra que se fundamentara a su vez en otra muestra. Si la población total de indicadores de habilidad numérica estuviese constituida por mil preguntas y se evaluara a un grupo de estudiantes con una muestra aleatoria de cien problemas, sus puntuaciones serían las mismas sin importar la muestra de la que se hayan obtenido las cien preguntas. Las muestras de los indicadores serían intercambiables. Si un estudiante recibiera un puntaje de X en una prueba de cien preguntas que se hubiese seleccionado al azar, recibiría el mismo puntaje en cualquier otra prueba de la misma naturaleza. Cada prueba puede, por tanto, intercambiarse con cualquier otra.

La prueba dura de la equivalencia de los instrumentos radica en la posibilidad de intercambiar sus resultados. Los instrumentos son equivalentes si y sólo si contamos con

algunas reglas que permitan intercambiar los enunciados de medición por ellos, sin alterar su valor lógico. En otras palabras, los instrumentos son equivalentes si la veracidad o la falsedad de los enunciados que de ellos se derivan son invariantes con respecto a su estructura. Si la puntuación resultante de la aplicación de un instrumento predice el puntaje resultante de la aplicación de otro instrumento, entonces los instrumentos son equivalentes. Dado que las mediciones involucran errores, debemos utilizar un lenguaje probabilístico: la puntuación de una prueba sólo requiere predecir probabilísticamente, más que determinar, el puntaje de otra prueba.

El concepto de equivalencia es intuitivamente claro en el interior de una misma cultura, por lo que es posible aplicar pruebas directas para corroborar si un par de instrumentos son, en efecto, equivalentes. ¿Pero podrán intercambiarse los instrumentos para evaluar la facilidad verbal en sueco y en inglés? ¿Existiría acaso la posibilidad lógica de que pudieran intercambiarse? La respuesta a la primera pregunta es claramente negativa. Para el caso de un estadounidense que no hable sueco, la puntuación de su facilidad verbal en sueco proporcionará una predicción trivial de su facilidad verbal; esto es, un puntaje de cero. La respuesta a la segunda pregunta es positiva. Existe la posibilidad lógica de que ambos instrumentos se intercambien, de tal manera que, una persona con cierta facilidad verbal obtuviera la misma puntuación de la misma forma, sin importar el idioma del “test”. Este enunciado se basa sobre el postulado de “substituibilidad” presentado en el primer capítulo: si el articulado¹¹ Sr. Peterson fuese estadounidense equivalente, podrían intercambiarse sus puntajes de facilidad verbal en sueco y en inglés. Ambos instrumentos, por consiguiente, serían equivalentes. El procedimiento para establecer equivalencias, para el caso en que los indicadores no sean comunes, se discutirá en el siguiente capítulo.

Resumen y conclusiones

El objeto de la ciencia consiste en reducir el agobiante número de observaciones específicas a enunciados generales. El primer paso para lograr este objetivo consiste en la ordenación de las observaciones específicas en un sistema de medición. Las observaciones específicas se efectúan en el interior de contextos de sistemas sociales particulares. Más, si se pretende que las teorías de las ciencias sociales sean generales, será necesario avanzar hacia conceptos generales, a partir de tales observaciones específicas relativas a contextos particulares. La medición transistémica deberá basarse, por consiguiente, en la mayoría de las situaciones, en la realización de inferencias que, a partir de observaciones específicas, den como resultado conceptos generales. Dicho proceso de inferencia no requerirá fundamentarse con las mismas proposiciones generales y no necesitará, por tanto, implicar indicadores comunes. Los enunciados de medición resultantes de la inferencia deben preservar su valor de falso/verdadero sin importar el instrumento y el sistema social que impliquen. Los instrumentos que satisfagan esta condición se denominarán equivalentes y proporcionarán resultados que serán válidos en cada uno de los sistemas y serán estables

¹¹ El adjetivo *articulate* refiere a una persona que tiene la habilidad de hablar coherente y fluido. (Esta nota al pie no forma parte del texto original).

intersistémicamente. Sin embargo, la invarianza del valor lógico de los enunciados de medición, entendido en términos probabilísticos, será el elemento constitutivo de la equivalencia en la investigación transistémica, y no lo será en cambio la comparabilidad de las puntuaciones obtenidas. Será criterio suficiente de equivalencia el que los indicadores de todos los sistemas nos permitan inferir, con igual validez, la magnitud de la propiedad general. El criterio de la estricta intercambiabilidad de puntajes particulares podría ser marcadamente riguroso e innecesario para la mayor parte de las situaciones que requieran de las mediciones transistémicas. Las puntuaciones podrán intercambiarse si, y sólo si, contamos con una regla de transformación que permita la expresión del puntaje de un instrumento en términos de la escala de otro. A fin de identificar tal regla de transformación, será necesario conocer las propiedades de cada escala. Para transformar los puntajes, por ejemplo, de una escala de intervalos diseñada para brasileños a una escala elaborada para mexicanos, deberán conocerse dos propiedades de la escala: 1) la magnitud de los intervalos, y 2) su origen. Debido a que esta condición es altamente restrictiva, habitualmente se la evade en la investigación comparativa, mediante la suposición de que los intervalos son iguales en todos los sistemas. Cuando se desconoce el origen de las escalas, también se presume que es común a cada uno de los sistemas¹².

Es importante recordar que el análisis comparativo de las relaciones en el interior de los sistemas no requiere del conocimiento de un origen común o de la magnitud de los intervalos, ya que ello sólo es necesario para las comparaciones descriptivas. El análisis comparativo de las relaciones internas de los sistemas requiere de instrumentos de medición menos poderosos que en los casos de los elementos de información descriptiva que cotejan parámetros intrasistémicos. Tanto los proyectos teóricos como los costos de la medición deben encaminar la investigación comparativa hacia comparaciones de las relaciones en el interior de los sistemas, más que de las propiedades agregadas de ellos.

La investigación comparativa requiere una estrategia para establecer la equivalencia. Los modelos de medición basados sobre la inferencia proporcionan un medio para deponer el efecto multiplicante de cientos de sistemas sociales, con el cual se pueden medir las propiedades en términos de un sistema particular que se encuentra exento de contexto. La estrategia, que aquí proponemos de hecho, mantiene constante el sistema social ya que acomoda las bases de la inferencia de tal manera que puede suprimirse el impacto del sistema social. Asimismo, proporciona un fundamento para codificar válidamente las características de los sistemas específicos en una jerga común: el lenguaje de la generalidad. Ello constituyó la perspicacia y el impacto de los primeros funcionalistas: la reducción de la diversidad social a un lenguaje general. En ese sentido hablaron de la equivalencia funcional, pero lo hicieron sin la ayuda de un lenguaje de medición explícito. La equivalencia requiere de criterios sistemáticos derivados de un lenguaje científico. Con criterios sistemáticos, la medición se verá liberada del sesgo de los investigadores

¹² Ello se supone, por ejemplo, en R. D. Meade y J. O. Whitaker, "A Cross-Cultural Study of Authoritarianism". *Journal of Social Psychology*, 72, 1967.

específicos y de la visión de las autoridades, para así poder ingresar al dominio público de la indagación científica.