



Conceptuación y medición

Lo que aprenderá en este capítulo

Descubrirá que las palabras del lenguaje cotidiano comunican significados vagos e inespecíficos. En cambio, en la ciencia es esencial especificar exactamente lo que entendemos (y lo que no entendemos) con los términos que usamos.

En este capítulo...

Introducción

La medición de todo lo que existe

¿Cómo se sabe?

Ideas y conceptos

Conceptuación

Indicadores y dimensiones

Carácter intercambiable de los indicadores

La confusión sobre definiciones y realidad

Creación de un orden conceptual

Ejemplo de conceptuación

Definiciones y propósitos de investigación

Cráteros de las mediciones de calidad

Confiabilidad

Validez

¿Quién decide qué es válido?

Tensión entre confiabilidad y validez

Puntos principales

Preguntas y ejercicios de repaso

Proyecto de continuidad

Lecturas adicionales

Introducción

Este capítulo es el primero de tres que se ocupan del proceso de pasar de las ideas vagas sobre lo que quiere estudiar a reconocerlo y medirlo en el mundo real. En este capítulo tratamos el tema general de la *conceptuación*, que dispone los cimientos para estudiar la *operacionalización* en el capítulo 6. En el capítulo 7, que examina las formas más complejas de medición, concluiremos los temas surgidos en los capítulos 5 y 6.

Quiero comenzar este capítulo con un ataque frontal a la preocupación que esconden algunas personas sobre la posibilidad de medir la sustancia de la vida: amor, odio, prejuicios, radicalismo, enajenación y asuntos por el estilo. La respuesta es sí, pero me tomará algunas páginas explicarme. Cuando usted vea que puede medir todo lo que existe, nos ocuparemos de las etapas necesarias para hacerlo.

sintió insatisfecho con la forma en que los investigadores midieron lo que fuera que estudiaran. Aca-so pensó que eran demasiado superficiales, que omitieron los aspectos que importan más. Quizá midieron la *religiosidad* como las veces que una persona iba a la iglesia o midieron el *liberalismo* de acuerdo con los votos de los ciudadanos en una sola elección. De seguro su insatisfacción habría aumentado si se descubriera mal clasificado por el sistema de medición. La gente suele tener esa experiencia.

O buscó en el diccionario la definición de una palabra como *compasivo*, la encontró deficiente y murmuró para sí mismo: "es más que eso". De hecho, cada vez que busca la definición de algo que ya entiende bien, es probable que imagine formas en que la gente confunda el término si sólo se basa en esa definición.

Ya hemos dicho que uno de los dos pilares de la ciencia es la *observación*. Como esta palabra puede sugerir una actividad más bien casual y pasiva, a menudo los científicos prefieren usar *medición*, con el significado de observaciones cuidadosas y deliberadas del mundo real, cuyo propósito es describir objetos y acontecimientos de acuerdo con los atributos que componen una variable. Si la variable en estudio fuera la *afiliación a un partido político*, consultaríamos el padrón de electores registrados para saber si las personas que estudiamos están registradas como demócratas o republicanas. De esta manera habríamos medido su afiliación partidista. O en una entrevista que indague a fondo al-

La medición de todo lo que existe

Me parece totalmente posible que tenga algunas reservas sobre la capacidad de las ciencias para medir los aspectos verdaderamente importantes de la existencia social humana. Habrá leído informes de investigación que se ocupan de algo como el liberalismo o la religión o los prejuicios, y tal vez se

gunos temas políticos, simplemente les preguntariamos el partido con el que se identifican y tomaríamos su respuesta como nuestra medición. En general, existe más de una manera de medir una variable. Si existe, podemos medirla.

¿Cómo se sabe?

Para demostrarle que los científicos sociales pueden medir cualquier cosa que exista, quisiera que imagine que estamos debatiendo el asunto. Yo escribiré el guión, pero siéntase libre de modificar su parte del diálogo según le acomode.

YO: Los científicos sociales pueden medir todo lo que existe.

USTED: ¡Jal Apuesto que no.

YO: Dígame algo que exista y le diré cómo medirlo.

USTED: Bueno, veamos su medición de los prejuicios.

YO: Buena elección. Pero no estoy dispuesto a perder nuestro tiempo tratando de medir algo que no existe, así que dígame si existen los prejuicios.

USTED: Claro que existen. Todos saben que los prejuicios existen, todos. Si usted fuera tan listo, también lo sabría. Hasta los tontos lo saben.

YO: Todos solían "saber" también que el mundo era plano. Quiero saber cómo sabe usted que los prejuicios existen en realidad.

USTED: Bueno, bueno. Como parece muy aferrado a las "observaciones", le diré que he visto los prejuicios.

YO: ¿Qué cosa ha visto usted que pruebe que los prejuicios existen?

USTED: Una vez, un empresario me dijo que nunca contrataría a una mujer para un puesto ejecutivo porque pensaba que todas eran caprichosas e irracionales. ¿Qué le parece?

YO: ¡Excelente! Eso me suena a prejuicios, así que supongo que podemos aceptar que existen. Ahora estoy preparado para medirlos. ¿Listo?

USTED: Listo.

YO: Usted y yo vamos a circular discretamente por las compañías locales para hablar con los empresarios acerca de las contrataciones. Cada vez que alguno nos diga que no contrataría a

una mujer para un puesto ejecutivo porque piensa que todas son caprichosas e irracionales lo contaremos como un caso de prejuicios. Cuando no se nos diga tal, lo anotaremos como un caso de desprejuicio. Al terminar, clasificaremos a los empresarios como prejuiciosos o desprejuiciados.

USTED: ¡Espere un momento! Ésa no es una medición muy buena de los prejuicios. Vamos a dejar fuera muchas clases. Nada más mediremos los prejuicios descarados en la contratación de mujeres. Justamente por eso dudo que pueda medirlos.

YO: Ya veo lo que quiere decir. Pero su comentario también significa que la situación que describió sólo prueba que existen los prejuicios descarados en la contratación de mujeres. Mejor reconsideremos si existen los prejuicios. ¿Existen?

USTED: Desde luego. Yo sólo le daba un ejemplo. Hay cientos de casos.

YO: Mencíname uno que pruebe que los prejuicios existen.

USTED: De acuerdo, pruebe éste. Estaba yo en el campus la otra noche y dos tipos, uno blanco y otro negro, discutían sobre política. El tipo blanco se enojó tanto que comenzó a usar un espantoso lenguaje racista y gritó: "Habría que enviarlos a todos ustedes de regreso a África". ¿Es bastante prejuicioso para usted?

YO: Lo encuentro bien. Parece probar que existen los prejuicios, así que estoy listo para volver a medirlos. Nos dividiremos y pasaremos por el campus todas las noches. Iremos con las orejas bien abiertas hasta escuchar a un blanco que lance epítetos racistas y diga: "habría que enviarlos a todos..."

USTED: ¡Un momento! Ya sé adónde va y no es por ahí tampoco. Si alguien dice eso sería prejuicioso, pero clasificariamos a mucha gente prejuiciosa como desprejuiciada sólo porque no se exalta ni usa epítetos racistas.

YO: Todo lo cual nos regresa a mi primera pregunta. ¿Existen realmente los prejuicios o sólo me está aguijoneando?

USTED: Sí, existen.

YO: Pues yo ya no estoy tan seguro. Usted me persuadió de que existen empresarios que discriminan en el empleo a las mujeres; usted lo dijo y yo le creo. Me persuadió de que hay

gente que insulta a los negros y les dice que regresen a África. Pero no estoy seguro de que existan los prejuicios. Desde luego que me gustaría encontrarlos para demostrarle que puedo medirlos. Pero, para ser honesto, empiezo a dudar de que existan realmente; es decir, ¿alguna vez ha visto prejuicios? ¿De qué color son? ¿Cuánto pesan? ¿Dónde se localizan?

USTED: ¿De qué está hablando? ¿Hola? Tierra llamando a Babbie. E.B., phone home.

El punto de este diálogo, como habrá conjeturado, es demostrar que los *prejuicios no existen*. No sabemos qué parecen, de qué tamaño o de qué color son. Nadie los ha tocado jamás ni ha cabalgado en ellos. Pero hablamos mucho de prejuicios. Veamos cómo sucedió.

Al recorrer el camino de la vida, usted y yo observamos muchas cosas y supimos que eran reales a partir de nuestras observaciones. Oímos muchas otras cosas que los demás nos dijeron que observaron y también parecían existir. Alguien dijo que vio un linchamiento y lo describió con lujo de detalles.

Conforme adquirimos más experiencia, advertimos algo más. Es muy probable que quienes participan en linchamientos también insulten a los negros. Más aún, muchos de ellos desean que las mujeres "se queden en su lugar". Al cabo, empezamos a comprender que había cierta clase de personas con esas tendencias. Cuando hablábamos de la gente que conocíamos, a veces resultaba conveniente identificar a alguien de acuerdo con dichas tendencias. Acostumbrábamos decir que fulano era "uno de esos que participan en linchamientos, insultan a los negros y no contratarían a una mujer para un puesto ejecutivo". Sin embargo, al poco rato se volvía engorroso decir todo eso y usted tuvo una idea brillante: "Vamos a usar la palabra *prejuicios* como abreviatura para referirnos a esas personas. Podemos usar el término aunque no hagan todas esas cosas, mientras se parezcan mucho".

Como soy complaciente y me interesa la eficiencia, accedí a emplear el sistema. De ahí salieron los *prejuicios*. En realidad nunca existieron, nunca los vimos. Sólo fue un atajo para hablar a espaldas de los demás. En última instancia, "prejuicios" no es más que un término que aceptamos usar en la comunicación: un nombre con el que representa-

mos todo un conjunto de fenómenos aparentemente relacionados que hemos observado en el curso de la vida. Cada uno de nosotros nos formamos una imagen mental de lo que representan en general estos fenómenos reales que observamos, y de lo que tienen en común.

Cuando empleo la palabra *prejuicios*, sé que se activa una imagen mental en su mente, así como en la mía. Es como si tuviéramos unos archiveros mentales llenos de miles de hojas con un rótulo en la esquina superior derecha. Una de las hojas en cada uno de sus archiveros tiene escrita la palabra *prejuicios*. Ahí se encuentra todo lo que le han dicho sobre los prejuicios y lo que usted ha observado que los ejemplifica. Mi hoja tiene lo que me han dicho más las situaciones que he visto que me parece que los ejemplifica.

Ideas y conceptos

El término técnico para estas imágenes mentales, estas hojas en nuestros archiveros mentales, es *idea*. Así, cada hoja es una idea. Ahora bien, no podemos comunicar directamente estas imágenes mentales. No hay modo de que pueda revelarle en forma directa lo que está escrito en las mías, así que nos valemos de los términos rotulados en la esquina superior derecha como medio de comunicar tanto nuestras ideas como lo que observamos que se les relaciona. Los términos vinculados a las ideas en nuestras mentes separadas nos posibilitan la comunicación y, en última instancia, concordar en lo que queremos decir específicamente con ellos. El proceso de ponerse de acuerdo es la *conceptuación*, y el resultado se llama *concepto*.

Supongamos que voy a conocer a alguien llamada Lola, a quien usted ya conoce; por eso, le pregunto cómo es ella. Digamos que usted la ha visto ayudar a que los niños perdidos encuentren a sus padres y devolver pajaritos a sus nidos. Lola influyó en el hecho de que usted regalara pavos a las familias pobres el día de acción de gracias y en el de que visitara un hospital infantil en Navidad. La ha sorprendido llorando con una película sobre una madre que supera las adversidades para salvar y proteger a su hijo. Mientras repasa su archivero mental, encuentra que todos o casi todos estos fenómenos están anotados en una úni-

ca hoja, etiquetada *compasiva*. Usted inspecciona otras entradas de la hoja y le parece que dan una descripción adecuada de Lola. Por tanto, me dice: "Lola es compasiva".

Ahora yo reviso mi archivero mental hasta que localizo una hoja titulada *compasiva*. Entonces, veo lo que está escrito y digo "ah, qué bien". Ahora creo que sé cómo es Lola, pero mis expectativas al respecto reflejan las entradas de mi archivero, no las del de usted. Después, cuando conozco a Lola, resulta que mis impresiones corresponden a las entradas de mi hoja *compasiva* y le digo que tenía razón. Pero también comento que mis observaciones de Lola contradicen cosas que tengo en mi hoja y que no creo que sea muy compasiva. Comenzamos entonces a comparar notas.

Usted me dice que "una vez vi llorar a Lola con una película sobre una madre que vence la adversidad para salvar y proteger a su hijo". Veo en mi hoja *compasiva* y no encuentro nada así. Buscando en otra parte del archivo, localizo esa clase de fenómeno en una hoja titulada *sentimental*. Replico: "Eso no es compasión; es sentimentalismo".

Para fortalecer mi argumentación, le digo que vi que Lola se negó a dar dinero para una organización dedicada a salvar de la extinción a las ballenas. "Eso es una falta de compasión", arguyo. Usted revisa su archivo y explica que tiene salvar a las ballenas en dos hojas: *activismo ambiental* y *encuentros entre especies*. Al cabo, nos ponemos a comparar las entradas de nuestras respectivas hojas tituladas *compasiva*. Descubrimos que las representamos con imágenes mentales muy diferentes.

En total, el lenguaje y la comunicación funcionan sólo porque las entradas de las hojas de nuestros respectivos archiveros se superponen considerablemente. Las similitudes de nuestras hojas representan los acuerdos de nuestra sociedad. De niños, nos dicen aproximadamente lo mismo cuando conocemos un término particular. Los diccionarios formalizan estos acuerdos que tiene nuestra sociedad sobre los términos. Así, cada uno de nosotros moldea sus imágenes mentales para que correspondan a dichos acuerdos, pero como tenemos experiencias y observaciones distintas, no hay dos personas que terminen exactamente con el mismo grupo de entradas en ninguna hoja de sus archiveros. Para regresar a la afirmación del principio del capítulo, podemos me-

dir todo lo que es real. Por ejemplo, podemos medir si Lola de verdad regresa pajaritos a sus niños, si visita el hospital en Navidad, si llora con las películas o si se niega a salvar a las ballenas. Todos estos fenómenos existen, así que podemos medirlos. ¿Pero en realidad es compasiva Lola? No podemos responder esta pregunta; en este sentido no podemos medir la compasión, porque no existe del mismo modo en que existen las cosas que acabamos de describir.

La *compasión* como *término* no existe. Podemos contar las letras y estar de acuerdo en que son nueve. Aceptamos que tiene tres sílabas y que comienza con C. En suma, podemos medir los aspectos que son reales.

Algunos aspectos de nuestras ideas también son reales. Tener o no una imagen mental del término *compasión* es real. Cuando un maestro de escuela primaria le pregunta a sus alumnos cuántos conocen el significado de *compasión*, es posible contar a los que levanten la mano. La presencia de determinadas entradas en las hojas rotuladas también es real y podemos medirla. Podemos medir cuántas personas relacionan el acto de dar dinero para salvar a las ballenas con su idea de la compasión. Lo único que no podemos medir es lo que sea realmente la compasión, porque no es real. La compasión existe sólo en la forma en que hemos acordado usar el término al comunicarnos cosas que son reales.

Si recuerda nuestro estudio del posmodernismo, imaginará que algunos objetarían el grado de "realidad" que confiero en los comentarios anteriores. En este capítulo no vamos a ser radicalmente posmodernos; creo que aceptará la importancia de una concepción intelectual dura de lo que es real y lo que no (cuando el intelectual se vuelve duro, se convierte en científico social).

En este contexto, Abraham Kaplan (1964) distingue tres clases de objetos que miden los científicos. La primera clase son los *observables directos*, aquello que observamos en forma simple y directa, como el color de una manzana o la marca en un cuestionario. Los *observables indirectos* requieren "observaciones relativamente más cuidadosas, complejas o indirectas" (1964:55). En un cuestionario, si observamos una marca en el recuadro *mujer*, hemos observado indirectamente el sexo de esta persona. Los libros de historia o las minutas de las juntas de los directivos de las em-

presas ofrecen observaciones indirectas de acciones sociales pasadas. Por último, los *constructos* son creaciones teóricas basadas en observaciones, que no pueden observarse en forma directa ni indirecta. El CI es un buen ejemplo. Es un constructo matemático de las observaciones de las respuestas a muchas preguntas en una prueba de inteligencia (en el capítulo 7 se encuentran otros ejemplos de constructos).

Kaplan (1964:49) define *concepto* como "una familia de ideas". Los conceptos —anota Kaplan— son constructos. Así, el concepto de la *compasión* es un constructo creado a partir de su idea de la compasión, de mi idea y de la de todos los que hayan empleado ese término. No se puede observar directa ni indirectamente porque no existe. Nosotros la construimos.

Conceptuación

La comunicación diaria se establece mediante un sistema de acuerdos vagos y generales sobre el uso de los términos. Los demás no entienden exactamente lo que queremos comunicar, sino que captan el sentido general. Aunque usted y yo no estemos de acuerdo por completo en el uso del término *compasiva*, creo que puedo dar por hecho que Lola no le quitaría las alas a las moscas. El precio que pagamos por nuestra imprecisión es una gama amplia de malos entendidos y conflictos —entre personas y entre naciones—, pero de algún modo nos las arreglamos. En cambio, la ciencia aspira a algo más que salir del paso; no funciona en el contexto de tal imprecisión.

Catherine Marshall y Gretchen Rossman (1995:18) hablan de un "embudo conceptual" por el que los intereses de los investigadores se concentran cada vez más. Así, un interés general en el activismo social se restringiría a "los individuos dedicados al cambio social y a dar el poder a la gente", y aún más en descubrir "cuáles son las experiencias que originaron el desarrollo de activistas sociales completamente comprometidos". Este proceso de concentración se vincula en forma inevitable al lenguaje que utilizamos.

Como hemos visto, la *conceptuación* es el proceso por el que especificamos lo que queremos comunicar cuando usamos determinados términos. Por ejemplo, supongamos que queremos averiguar si

las mujeres son más compasivas que los hombres. Sospecho que muchas personas dan por hecho que así es, pero sería interesante saber si así es en verdad. No podemos estudiar con provecho la cuestión, para no hablar de coincidir con el resultado, sin algún acuerdo de trabajo sobre el significado de *compasión*. Hablamos de acuerdos *de trabajo* en el sentido de que nos permiten trabajar en el tema. No tenemos que concordar —o pretender que concordamos— en que esa especificación sea en definitiva la mejor.

Indicadores y dimensiones

El producto de este proceso de conceptuación es la especificación de uno o más *indicadores* de lo que tenemos en mente y que señalan la presencia o la ausencia del concepto que estudiamos. Veamos un ejemplo bastante sencillo.

Estáramos de acuerdo en que visitar hospitales infantiles en Navidad y en Hanukkah es un indicador de compasión; devolver los pajaritos a sus nidos sería otro indicador, etc. Si la unidad de análisis de la investigación fuera el individuo, observaríamos entonces la presencia o la ausencia de cada indicador en cada persona de nuestro estudio. Aparte, sumaríamos los indicadores de compasión que observáramos en cada individuo. Por ejemplo, podríamos ponernos de acuerdo sobre 10 indicadores y encontrar seis en nuestro estudio de Lola, tres en Juan, nueve en María, etcétera.

Para regresar a nuestra pregunta original, digamos que calculamos que las mujeres de nuestro estudio tienen en promedio 6.5 indicadores de compasión mientras que los hombres tienen 3.2. Por tanto, concluiríamos sobre la base de nuestro análisis cuantitativo de las diferencias grupales que las mujeres son, en conjunto, más compasivas que los hombres. Sin embargo, por lo regular no es tan sencillo.

Imagine que a usted le interesa comprender un pequeño culto fundamentalista cristiano, en particular sus ásperas opiniones sobre varios grupos: homosexuales, ateos, feministas y otros. De hecho, sus miembros afirman que cualquiera que se niegue a unirse a su grupo arderá en el infierno. En el contexto de su antiguo interés en la compasión, no parecen tener mucha. No obstante, los escritos del grupo suelen hablar de su compasión por los demás.

Para explorar este interés de investigación, podría decidir alternar con los miembros del culto para conocerlos y aprender más sobre sus puntos de vista. Les diría que es un investigador social interesado en saber del grupo, o quizá sólo expresaría su interés sin decir por qué.

En el curso de sus conversaciones con los miembros y quizá en sus ceremonias religiosas, usted estaría en la posibilidad de llegar a entender lo que significa *compasión* para ellos. Por ejemplo, aprendería que están tan preocupados por los pecadores que arden en el infierno que están dispuestos a ser agresivos, incluso violentos, para conseguir que la gente cambie sus costumbres pecaminosas. Por tanto, con este paradigma, consideran que es un acto de compasión golpear homosexuales, prostitutas y médicos que practican abortos.

Los científicos sociales suelen centrar su atención en los significados que las personas que son objeto de estudio dan a las palabras y los actos. Esto puede aclararles sus conductas, pero casi siempre complica los conceptos que les interesan.

Cuando analizamos con seriedad nuestros conceptos y especificamos lo que entendemos por ellos, descubrimos desacuerdos e inconsistencias. No sólo disintimos usted y yo, sino que además es probable que encontremos bastante confusas nuestras imágenes mentales. Si dedica un momento a considerar lo que usted entiende por compasión, verá que su imagen contiene varias clases de compasión. Las entradas en la hoja de su archivo se pueden componer en grupos y subgrupos, y hasta encontrará varios métodos para hacer combinaciones. Por ejemplo, podría agrupar las entradas en sentimientos y actos.

El término técnico para estos agrupamientos es **dimensión**: un aspecto o faceta especificable de un concepto. Así, hablaríamos de la "dimensión de sentimientos" de compasión y de la "dimensión de actos" de compasión. En otro agrupamiento, distinguiríamos "compasión por los humanos" de "compasión por los animales". O podríamos pensar que la compasión consiste en ayudar a las personas a tener lo que nosotros queremos para ellas en lugar de lo que ellas quieren para sí mismas. Y todavía cabe distinguir la "compasión como perdón" de la "compasión como piedad".

Así, podemos dividir *compasión* en varias dimensiones. La concepción consiste tanto en es-

pecificar dimensiones como en identificar los indicadores de cada una.

A menudo, especificar las dimensiones de un concepto prepara el terreno para una comprensión más profunda de lo que estudiamos. Por ejemplo, podemos observar que las mujeres son más compasivas en términos de sentimientos y los hombres en términos de actos (o viceversa); pero, como quiera que sea, no podríamos decir si las mujeres son más compasivas que los hombres. De hecho, nuestra investigación habría mostrado que no hay una sola respuesta.

Carácter intercambiable de los indicadores

Supongamos por un momento que usted y yo hemos compilado una lista de 100 indicadores de *compasión* con sus dimensiones. Digamos también que estamos muy en desacuerdo sobre los indicadores que prueban más claramente la presencia o la falta de compasión. Si estamos de acuerdo en algunos indicadores, podríamos centrar en ellos nuestra atención y es probable que coincidiéramos en la respuesta que dan. Pero supongamos que no concordamos en ningún indicador posible. Todavía podemos llegar a un acuerdo sobre si las mujeres o los hombres son más compasivos.

Si disintimos completamente en el valor de los indicadores, una solución sería estudiarlos todos. Supongamos que resulta que las mujeres son más compasivas que los hombres en los 100 indicadores (en todos los que usted favorece y en todos los míos). Entonces, podríamos convenir en que las mujeres son más compasivas aunque sigamos en desacuerdo sobre lo que signifique la compasión en general.

El **carácter intercambiable de los indicadores** significa que si varios indicadores representan en algún grado el mismo concepto, todos funcionarán de la misma manera que lo haría el concepto si fuera real y observable. Así, si en general las mujeres son más compasivas que los hombres, deberíamos observar esta diferencia utilizando cualquier medición razonable de la compasión.

Ya conoce la lógica fundamental de la concepción y la medición. Los análisis que siguen en este capítulo y el siguiente son ante todo perfeccionamientos y extensiones de lo que acabamos de

presentar. Sin embargo, antes de pasar a las elaboraciones más técnicas del marco teórico principal, quiero tocar unos cuantos temas generales útiles.

Primero, sé que los párrafos anteriores quizá no se ajusten exactamente a su conocimiento del significado de términos como *prejuicios* y *compasión*. En la vida diaria nos conducimos como si tales términos tuvieran significados reales y definitivos. Por eso, en la siguiente subsección quiero comentar brevemente la manera en que llegamos a ese conocimiento.

Segundo, para que todo este análisis no produzca una imagen de anarquía en el significado de las palabras, vamos a describir algunas de las formas en que los científicos sociales han organizado la confusión para normar y dar consistencia y generalidad al significado de los términos. Deberá terminar esta lectura con un nuevo sentido del orden, uno basado en la comprensión consciente más que en una aceptación casual del uso común.

La confusión sobre definiciones y realidad

Para empezar, un repaso breve. Los conceptos proceden de imágenes mentales (ideas) que resumen grupos de observaciones y experiencias que parecen estar relacionadas. Estas experiencias y observaciones son reales, por lo menos subjetivamente, en tanto que los conceptos son meras creaciones mentales. Los términos asociados con los conceptos no son más que ardidres creados con la finalidad de clasificar y comunicar. Por ejemplo, el término *prejuicios* es sólo un grupo de letras y no tiene significado intrínseco.

Ahora bien, solemos caer en la trampa de creer que los términos tienen significado real. El peligro aumenta cuando comenzamos a tomar en serio los términos y a tratar de usarlos con precisión. Más aún, el peligro es mayor en presencia de expertos que parecen saber más que uno sobre el significado real de los términos; en tal situación, es muy fácil ceder a la autoridad.

Una vez que hemos asumido que los términos tienen significados reales, iniciamos la tortuosa tarea de descubrir cuáles son los significados reales y qué constituye una medición genuina de

ellos. La figura 5.1 ilustra la historia de este proceso. Fabricamos resúmenes conceptuales de observaciones reales porque son convenientes; pero esta conveniencia es tal que damos en pensar que son reales. Se llama **cosificación** al acto de considerar real aquello que no lo es. En la vida diaria, la cosificación de conceptos es bastante común.

El diseño y la realización de investigaciones sociales requiere que aclaremos la confusión entre conceptos y realidad. Para este fin, lógicos y científicos han encontrado útil distinguir tres clases de definiciones: *reales*, *nominales* y *operacionales*. La primera refleja la cosificación de los términos. Carl Hempel advierte:

De acuerdo con la lógica tradicional, una definición "real" no es una estipulación del significado de alguna expresión, sino una afirmación de la "naturaleza esencial" o los "atributos esenciales" de una entidad. Sin embargo, la noción de naturaleza esencial es tan vaga que esta caracterización es inservible para los propósitos de una investigación rigurosa.

(1952:6)

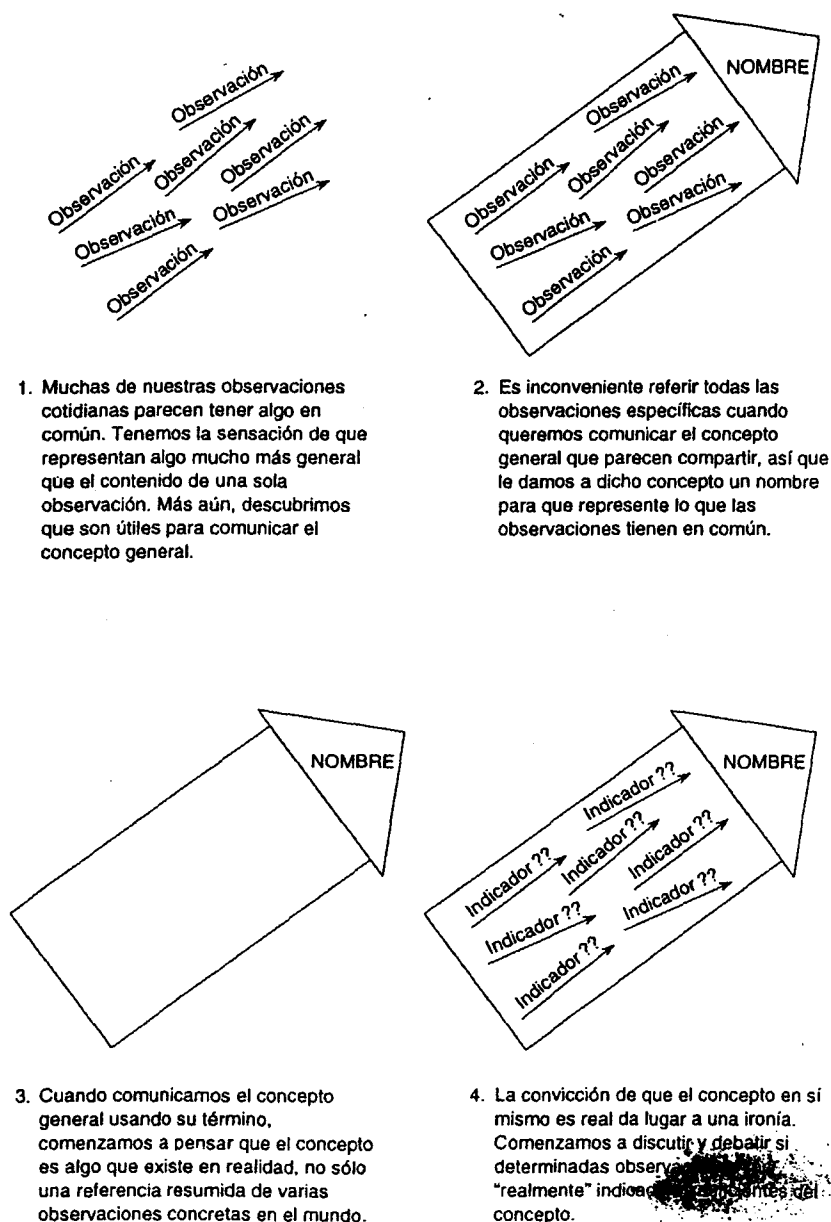
La especificación de conceptos en la investigación científica depende de definiciones nominales y operacionales. Las definiciones *nominales* son las que se asignan a los términos. En medio de los desacuerdos y la confusión sobre lo que significa un término, podemos especificar una definición de trabajo para los propósitos de una investigación. Por ejemplo, si deseamos examinar en un estudio el nivel socioeconómico (NSE), podemos especificar simplemente que vamos a tratarlo como una combinación de ingresos y nivel de estudios. Con esta decisión, descartamos otros aspectos del NSE: ocupación, dinero en el banco, propiedades, linaje, estilo de vida, etcétera.

Creación de un orden conceptual

La clarificación de conceptos es un proceso continuo en la investigación social. En algunas formas de investigación cualitativa, es un elemento clave en la recopilación de datos. Supongamos que realizamos entrevistas y observaciones en un grupo político radical dedicado a combatir la opresión en la sociedad. Imagine los cambios que sufrirá el sig-

Figura 5.1

El proceso de la confusión conceptual



nificado de *opresión* a medida que profundice en las experiencias y la visión del mundo de los miembros.

En el análisis de textos, los investigadores sociales hablan del "círculo hermenéutico", un proceso cíclico de comprensión cada vez más profundo.

La comprensión de un texto se da mediante un proceso en que el significado de las partes está determinado por el sentido general del texto tal como fue anticipado. En última instancia, la determinación más precisa de las partes puede cambiar el sentido anticipado del total, que de nuevo influirá en el sentido de las partes, y así sucesivamente.

(KVALE, 1996:47)

El perfeccionamiento continuo de los conceptos ocurre en todos los métodos de investigación social. Usted verá que afina los significados incluso cuando escriba su informe final.

Aunque la concepción es un proceso continuo, es vital que uno la emprenda específicamente desde el comienzo de cualquier diseño de estudio, en particular en los que tienen una estructura rigurosa, como las encuestas y los experimentos. Por ejemplo, en una encuesta la operacionalización ocasiona una limitación al conjunto específico de preguntas del cuestionario que representan los conceptos que se estudian. Sin esta limitación, el estudio no avanzaría.

Sin embargo, también en los métodos de investigación menos estructurados se debe comenzar con un conjunto de significados anticipados que se puedan perfeccionar durante la recopilación e interpretación de los datos. Nadie cree de verdad que podamos observar la vida sin nociones preconcebidas; así, el observador científico debe estar consciente y ser explícito sobre estos puntos iniciales.

Exploremos la concepción inicial tal como se aplica a las investigaciones estructuradas, como las encuestas y los experimentos. El hecho de especificar las definiciones nominales concentra nuestra estrategia de observación, pero no nos permite observar. En la siguiente etapa debemos especificar exactamente qué es lo que vamos a observar, cómo lo haremos y qué interpretaciones asignaremos a las observaciones posibles. Todas estas especificaciones nuevas conforman la denominada **definición operacional** del concepto:

una definición que explica con precisión la manera en que será medido. En estricto sentido, una definición operacional es una descripción de las "operaciones" que emprenderemos para medir el concepto.

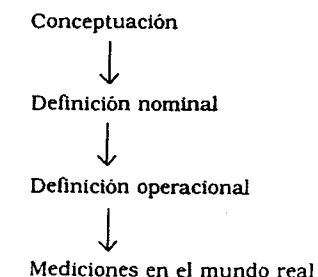
Para continuar con el caso del NSE, podríamos formular a los entrevistados dos preguntas:

1. ¿Cuál fue el ingreso total de su familia durante los últimos 12 meses?
2. ¿Cuál es el mayor grado de estudios que cursó?

Aquí, tal vez quisiéramos especificar un sistema para clasificar las respuestas. En cuanto a los ingresos, estableceríamos categorías como "menos de 5 000 dólares" o "de 5 000 a 10 000 dólares". También agruparíamos en categorías el nivel escolar. Por último, especificaríamos cómo combinaríamos las respuestas de las dos categorías para tener una medida del NSE (en el capítulo 7 presentaremos algunos métodos para lograrlo).

Al final, crearíamos una definición de trabajo viable del NSE. Aunque otros investigadores no estén de acuerdo con nuestra concepción y operacionalización, la definición tendría una virtud científica esencial: sería por completo específica e inequívoca. Aun en el caso de que alguien no concordara con nuestra definición, tendría una buena idea sobre cómo interpretar los resultados de nuestra investigación porque sería claro lo que entendemos por NSE (lo cual quedaría manifestado en nuestros análisis y conclusiones).

Veamos un diagrama de la progresión de las etapas de medición desde nuestra idea vaga de lo que significa el término hasta las mediciones específicas en una investigación científica completamente estructurada:



Ejemplo de conceptuación

Quiero reunir los párrafos anteriores en una breve historia de un concepto social científico. Los investigadores que estudian las riñas callejeras suelen interesarse en el papel que juegan los sentimientos de impotencia. En este contexto, a veces emplean la palabra *anomia*. El gran sociólogo francés Emile Durkheim introdujo este término a las ciencias sociales en su clásico estudio de 1897: *El suicidio*.

A partir sólo de las publicaciones gubernamentales sobre los índices de suicidio en varias regiones y países, Durkheim produjo una obra de genio analítico. Para determinar los efectos de la religión en el suicidio, comparó los índices de los países predominantemente protestantes con los de las naciones predominantemente católicas, las regiones protestantes de los países católicos con las regiones católicas de los países protestantes, etc. Para determinar los posibles efectos del clima, comparó los índices de naciones y regiones septentrionales y australes, y los examinó según los meses y las estaciones del año. Así, llegó a conclusiones sobre un acto supremamente individualista y personal sin tener dato alguno sobre las personas que lo cometieron.

En un nivel más general, Durkheim postuló que el suicidio también refleja el grado al que los acuerdos de la sociedad son claros y estables. Al observar que las épocas de agitaciones y cambios sociales enfrentan al individuo con graves incertidumbres sobre lo que se espera de él, Durkheim indicó que éstas le producen confusión, ansiedad e incluso lo llevan a destruirse. Para describir la condición social de falta de normas, eligió el término *anomia*. Vale la pena anotar que Durkheim no se sacó la palabra de la manga. Usada tanto en alemán como en francés, significa literalmente "sin ley", y el término en inglés, *anomia*, se emplea desde hace siglos con la significación peculiar de "desacato de la ley divina". Pero Durkheim hizo de *anomia* un concepto propio de las ciencias sociales.

En los años que siguieron a la publicación de *El suicidio*, los científicos sociales encontraron útil el término *anomia* como concepto y muchos ampliaron el significado que le asignó Durkheim. Robert Merton, en un artículo clásico titulado "Social Structure and Anomie" (1938), concluyó que la anomia es el resultado de una disparidad entre las

metas y los medios prescritos por la sociedad. Por ejemplo, el éxito económico es una meta que muchos individuos comparten en nuestra sociedad, pero no todos cuentan con los recursos para alcanzarla por medios aceptables. Según Merton, resaltar la meta en sí provoca una falta de normas, porque los que rechazan las vías tradicionales para hacerse ricos proceden por medios ilegítimos para lograrlo. Por tanto, el estudio de Merton puede considerarse una conceptuación más profunda del concepto de anomia.

Aunque Durkheim tomaba el concepto de *anomia* como una característica de las sociedades, lo mismo que Merton después de él, otros científicos sociales lo han referido a los individuos (para aclarar la distinción, algunos estudiosos prefieren usar el término en su sentido social original y usar *anómico* para referirse a la característica del individuo). Así, en cualquier sociedad algunos individuos sufren de anomia y otros no. En un texto 20 años posterior al de Morton, Elwin Powell ofreció la siguiente conceptuación de anomia como característica de los individuos:

Cuando los fines de la acción se vuelven contradictorios, inaccesibles o insignificantes, aparece una condición de anomia. Caracterizada por una pérdida general de orientación y acompañada de sentimientos de "vacío" y apatía, la anomia puede concebirse simplemente como una falta de significación.

(1957:132)

Powell indicó en seguida que hay dos clases de anomia y examinó cómo surgen en diferentes experiencias laborales y desembocan, a veces, en el suicidio. Sin embargo, Powell no midió la anomia *per se*, sino que estudió la relación entre suicidio y ocupación, e hizo inferencias sobre ambas clases de anomia. Así, el estudio no provee una definición operacional de anomia, sino sólo una nueva conceptuación.

Muchos investigadores han ofrecido definiciones operacionales, pero un nombre sobresale entre todos. Dos años antes de que apareciera el artículo de Powell, Leo Srole (1956) publicó un conjunto de reactivos para cuestionario que, según él, brindaba una buena medición de la anomia que sufren los individuos. Se trata de cinco enunciados con los que los sujetos deben o no estar de acuerdo:

1. A pesar de lo que digan algunos, la suerte del hombre común está empeorando.
2. No es justo traer niños al mundo dado cómo se ve el futuro.
3. En nuestros días, uno tiene que vivir para el hoy y dejar que el mañana se cuide solo.
4. En la actualidad, uno no sabe realmente con quién puede contar.
5. No tiene caso escribir a los funcionarios públicos porque no están interesados verdaderamente en el hombre común.

(1956:713)

En las décadas que siguieron a la publicación, la escala de Srole se convirtió en un elemento básico de la investigación para los científicos sociales. Descubrirá que esta operacionalización de anomia se emplea en muchos proyectos de investigación publicados en revistas académicas. De esto trata Srole en el recuadro, "Los orígenes del constructo 'anómico'", que preparó para este libro antes de su muerte.

En caso de que le interese en el concepto de *anómico*, he aquí unos datos de una encuesta realizada en 1994 en Estados Unidos por el Centro Nacional de Investigación de la Opinión de la Universidad de Chicago, como parte de la Encuesta General Social permanente (véase la página en Internet de la ESG en <http://www.icpsr.umich.edu/gss/subject/anomia.htm>). En cada una de las afirmaciones siguientes anoto el porcentaje de quienes están de acuerdo entre todos los que expresaron una opinión.

	Porcentaje de acuerdo
A pesar de lo que digan algunos, la suerte del hombre común no mejora, sino que empeora.	69%
No es justo traer niños al mundo dado cómo se ve el futuro.	45%
La mayoría de los funcionarios públicos no están interesados verdaderamente en los problemas del hombre común.	76%

Presenté esta historia muy abreviada de los conceptos sociales científicos de anomia y anómico por varias razones. Primera, porque ejemplifica el proceso mediante el cual los conceptos generales se convierten en mediciones operacio-

nalizadas, aunque no quiero que piense que la distinción entre anomia y anómico está resuelta de una vez para siempre. Es seguro que los estudiosos los conceptuarán y operacionalizarán de nuevo en los próximos años, en busca de mediciones más útiles.

Por ahora, puse fin a la historia con la escala de Srole porque ilustra otro punto importante. El hecho de dejar abiertas la conceptuación y la operacionalización no produce necesariamente anarquía y caos, como acaso usted imaginaria. Con frecuencia surge un orden en el que figuran varios elementos. Primero, aunque uno *puede* definir anomia como quiera —digamos, la talla de zapatos—, lo más probable es que lo haga en formas que no difieran mucho de las imágenes mentales de los demás. Si usted adoptara una definición verdaderamente excéntrica, la gente lo ignoraría.

Segundo, cuando los investigadores descubren la *utilidad* de una conceptuación y operacionalización determinadas de un concepto, las adoptan, lo que origina definiciones uniformes de los conceptos. Además de la escala de Srole, citemos como ejemplos las pruebas de inteligencia y un caudal de mediciones demográficas y económicas que elaboran las oficinas de censos. El uso de mediciones establecidas tiene dos ventajas: han sido probadas y depuradas a fondo y es posible comparar los estudios que las utilizan. Si usted y yo realizamos sendos estudios sobre grupos distintos y aplicamos la escala de Srole, podemos comparar nuestros grupos sobre la base de la anomia.

Por tanto, los científicos sociales pueden medir todo lo real, e incluso hacen un buen trabajo en la medición de lo que no lo es. Admitiendo que a fin de cuentas conceptos como el nivel socioeconómico, los prejuicios, la compasión y la anomia no son reales, vemos que los científicos sociales pueden crear un orden al manejarlos, un orden que no se basa en la veracidad, sino en la utilidad.

Dedicaremos el resto del capítulo a algunas consideraciones y alternativas para formular definiciones y mediciones útiles. Primero, veremos las relaciones entre las definiciones y los propósitos de las investigaciones. Después, concluiremos el capítulo con el examen de algunos criterios para determinar la *calidad* de las mediciones que concebimos.

Los orígenes del constructo "atómico"

por Leo Srole
Centro de Geriatria y Gerontología, Universidad de Columbia

Mi larga fijación profesional en la anomia comenzó cuando leí *Le Suicide*, de Durkheim, mientras estudiaba en Harvard. Después, como alumno de posgrado en Chicago, estudié con dos antropólogos de la corriente de Durkheim: William Lloyd Warner y Alfred Radcliffe-Brown. Este último había mantenido una nutrida correspondencia con Durkheim, lo que me convierte en "descendiente" colateral del gran sociólogo francés.

Para mí, el primer impacto de la obra de Durkheim sobre el suicidio fue contradictorio pero permanente. Por un lado, tenía serias reservas sobre sus esfuerzos vigorosos, ingeniosos y a veces torpes por violentar los crudos registros burocráticos sobre tasas de suicidio para que casaran con su determinismo sociológico unidireccional. Por el otro, me conmovió su preocupación tenaz por la fuerza moral de los

lazos personales que nos vinculan a nuestro tiempo, lugar y pasado, y también su perspicacia al contemplar las consecuencias mortales que pueden seguir a la contracción y la descomposición de dichos lazos.

Mi interés en la anomia recibió una sacudida de testigo ocular al final de la Segunda Guerra Mundial, cuando trabajé para la Administración de Socorro y Rehabilitación de las Naciones Unidas (ASRNU) en la reconstrucción de una Europa despedazada por la guerra. En el campo de concentración de los nazis en Dachau vi de primera mano los abismos de deshumanización que las fuerzas macrosociales, como las que ocupaban a Durkheim, producían en individuos como Hitler, Eichmann y otros que obedecían sus dictados en todos los niveles de las fábricas nazis de la muerte.

Al volver de mi puesto en la ASRNU, sentía con toda urgencia que se acababa el tiempo para comprender la dinámica de la desintegración de los vínculos sociales. Necesitábamos trabajar en

de otro término: *fuerza laboral*. Si es del todo absurdo considerar desempleado a un niño de tres años, es porque no lo consideramos un miembro de la fuerza laboral. Así, en concordancia con una convención general de la oficina de censos, podemos excluir a los menores de 14 años de la fuerza laboral.

Sin embargo, esta sola convención no nos da una definición satisfactoria, porque contaría como desempleados a los estudiantes de preparatoria, los jubilados, los incapacitados y las amas de casa. Podemos ampliar la convención y definir como miembros de la fuerza de labor a "todas las personas mayores de 14 años que trabajan, buscan trabajo o esperan ser reinstaladas en un puesto del que han sido despedidas o después de un permiso". Entonces, los desempleados serían todos los miembros de la fuerza laboral que no tienen empleo. Si un estudiante, una ama de casa o un jubilado no buscan trabajo, no debemos incluirlos en la fuerza laboral.

¿Pero qué significa "buscar trabajo"? ¿Debe

forma expedita, aminorar la proliferación de las teorías más generales en favor del encuentro exploratorio directo con los individuos mediante la recién creada metodología moderna de la investigación con encuestas. También pensaba que una investigación así debía concentrarse en un espectro más amplio de patologías sociales que sólo el suicidio.

Mis primeras investigaciones fueron tentativas y diversas. Por ejemplo, en 1950 entrevisté a una muestra de 451 pasajeros de autobús en Springfield, Mass. Cuatro años después, el estudio de salud mental del centro de Manhattan me permitió llegar a una población mucho más grande. Estos proyectos de campo, entre otros, me proporcionaron la competencia para que expandiera y perfeccionara mis mediciones de la cualidad individual que reflejaba la cualidad macrosocial que Durkheim llamó *anomia*.

Al principio utilicé el término de Durkheim en mi propio trabajo, pero pronto decidí que era necesario limitar su uso a su significado

macrosocial y separarlo de tajo de sus manifestaciones individuales. Para este propósito, convenía adoptar la forma adjetiva del viejo término griego: *anómico* que se proponía solo.

El constructo del individuo *anómico* vio la luz en un artículo de 1956 en la *American Sociological Review*.^{*} Ahí expliqué las formas de operacionalizarlo y presenté los resultados de su primera aplicación en la investigación de campo. Para 1982, el índice de Referencias Científicas y el índice de Referencias de las Ciencias Sociales anotaba unas 400 publicaciones de ciencias políticas, psicología, trabajo social y sociología de todo el mundo que se habían valido de los instrumentos y descubrimientos de aquel artículo, lo que le valió que el Instituto Estadounidense de Información Científica lo designara "referencia clásica".

^{*}Leo Srole, "Social Integration and Certain Corollaries: An Exploratory Study", en *American Sociological Review* 21 (1956): 709-16.

Definiciones y propósitos de investigación

Recuerde que en el capítulo 4 vimos que los dos propósitos generales de las investigaciones son la *descripción* y la *explicación*. La distinción entre ambas tiene implicaciones importantes para las definiciones y las mediciones. Si usted tiene la idea de que la descripción es más simple que la explicación, se sorprenderá de saber que las definiciones son más problemáticas en la investigación descriptiva que en la explicativa. Es necesario que sepa por qué sucede así antes de que nos ocupemos de otros aspectos de la medición, si bien en la parte 4 estudiaremos más a fondo el punto.

Es evidente la importancia de las definiciones para la investigación descriptiva. Si queremos describir e informar el índice de desempleo en una ciudad, nuestra definición de *estar desempleado* es crucial. Además, dependerá de nuestra definición

estar registrado en una oficina oficial de colocaciones o debe ir de puerta en puerta pidiendo empleo? ¿Basta querer trabajar o estar disponible para una oferta de empleo? Convencionalmente, la definición operacional de "buscar trabajo" es responder sí al entrevistador que pregunta "¿ha buscado trabajo durante los últimos siete días?" (siete días es el periodo más especificado, pero para algunos propósitos de investigación podría tener más sentido reducirlo o alargarlo).

He explicado con algún detalle estas consideraciones para que usted vea que la conclusión de un estudio descriptivo sobre, por ejemplo, el índice de desempleo depende directamente de cómo se resuelve cada cuestión. Incrementar el periodo que cuenta para decir que alguien busca trabajo tiene el efecto de añadir más desempleados a la fuerza laboral (tal como la definimos), y por tanto aumentar el índice de desempleo que se informa. Si seguimos otra convención y hablamos de fuerza laboral civil e índice de desempleo civil, excluiríamos a los

militares, lo que también aumenta el índice de desempleo resultante, puesto que los militares estarían empleados *por definición*.

Así, la afirmación descriptiva de que el índice de desempleo en cierta ciudad es de tres por ciento o de nueve o de lo que sea depende en forma directa de las definiciones operacionales que se utilizaron. Este ejemplo es relativamente claro porque hay varias convenciones aceptadas en cuanto a la fuerza laboral y el desempleo; pero piense en lo difícil que sería estar de acuerdo en las definiciones que necesitaríamos para decir que "45 por ciento de los estudiantes tienen posturas políticas conservadoras". Este porcentaje, como el índice de desempleo, depende directamente de la definición de lo que se mide. Otra definición propiciaría quizá la conclusión de que "cinco por ciento del estudiantado tiene posturas políticas conservadoras".

Irónicamente, las definiciones son menos problemáticas en el caso de la investigación explicativa. Supongamos que nos interesa explicar el con-

servadurismo político. ¿Por qué unas personas son conservadoras y otras no? Más en concreto, digamos que nos interesa saber si los ancianos son en general más conservadores que los jóvenes. ¿Qué ocurre si usted y yo tenemos 25 definiciones operacionales de *conservador* y no podemos coincidir en cuál sea la mejor? Como ya vimos, éste no es un obstáculo necesariamente insalvable. Por ejemplo, supongamos que descubrimos que los ancianos son más conservadores que los jóvenes según las 25 definiciones (recuerde nuestro análisis de la compasión en hombres y mujeres). Supongamos, pues, que los ancianos son más conservadores que los jóvenes según toda definición razonable de conservadurismo que se nos ocurra. No importaría cuál fuera la definición; concluiríamos que los ancianos son en general más conservadores que los jóvenes aunque no nos hayamos puesto de acuerdo en lo que es el conservadurismo.

En la práctica, la investigación explicativa rara vez es tan inequívoca como en este ejemplo; no obstante, la pauta general es bastante común en la investigación real. Existen relaciones constantes en la vida social humana que dan por resultado hallazgos congruentes. Pero esta congruencia no aparece en una situación descriptiva. El hecho de cambiar las definiciones lleva casi inevitablemente a conclusiones descriptivas diferentes.

Criterios de las mediciones de calidad

Hemos avanzado un poco en el capítulo. Comenzamos con la afirmación directa de que los científicos sociales pueden medir todo lo que existe. Entonces, descubrimos que casi todas las cosas que quisiéramos medir y estudiar en realidad no existen. Después aprendimos que de todos modos es posible medirlas. Quiero concluir el capítulo con un análisis de algunas de las varas con las que juzgamos nuestro relativo éxito o fracaso en la medición de las cosas, incluso de las que no existen.

Para comenzar, las mediciones se pueden realizar con varios grados de *precisión*, que representan la finura de las distinciones hechas entre los atributos que componen una variable. La descripción "una mujer de 43 años" es más precisa que "en sus cuarenta". Decir que una pandilla callejera se for-

mó en el verano de 1996 es más preciso que decir "en la década de 1990".

Como regla general, las mediciones precisas son superiores a las imprecisas, como dicta el sentido común. No hay condiciones en las que las mediciones imprecisas sean intrínsecamente superiores a las precisas. Con todo, la precisión no siempre es necesaria o deseable. Si saber que una mujer está en sus cuarenta satisface los requisitos de su investigación, cualquier otro esfuerzo invertido en precisar su edad es un desperdicio. Así, la operacionalización de conceptos debe guiarse en parte por un conocimiento del grado de precisión que se requiere. Si sus necesidades no están claras, sea más preciso y no menos.

Ahora bien, no confunda precisión con *exactitud*. Describir a alguien como "nacido en Stowe, Vermont" es más preciso que decir "nacido en Nueva Inglaterra"; pero supongamos que esta persona nació realmente en Boston. En este caso, la descripción menos precisa es un reflejo mejor y más exacto de la realidad.

Es evidente que la precisión y la exactitud son cualidades importantes en la medición científica y de hecho no requieren mayores explicaciones. Sin embargo, digamos aún que cuando los científicos sociales disponen y evalúan sus mediciones, prestan atención a dos consideraciones técnicas: **confiabilidad** y **validez**.

Confiabilidad

En términos abstractos, la *confiabilidad* se refiere al hecho de que determinada técnica, aplicada repetidamente al mismo objeto, arroje cada vez el mismo resultado. Supongamos que usted quiere saber cuánto peso yo (quién sabe por qué quiere saberlo). Una técnica consistiría en pedirle a dos personas que estimen mi peso. Si una calcula 80 kilogramos y la otra 160, tendríamos que concluir que la técnica de hacer que la gente calcule mi peso no es muy confiable.

Supongamos como alternativa que usamos una báscula de baño como nuestra técnica de medición. Me subo dos veces a la báscula y obtengo ambos resultados. Es de creer que la báscula arrojará el mismo peso las dos veces, lo que indicaría que es una técnica más confiable para medir el peso que pedirle a la gente que lo calcule.

Sin embargo, la confiabilidad no garantiza la exactitud más que la precisión. Supongamos que

ajusté mi báscula para que me quitara dos kilogramos y medio, sólo para hacerme sentir mejor. Usted referiría el mismo peso las dos veces (confiabilidad), pero en ambas estaría equivocado. En el capítulo 8 estudiamos este nuevo elemento, llamado **sesgo**. Por ahora, basta estar alerta al hecho de que la confiabilidad no garantiza la exactitud.

Supongamos que nos interesa estudiar la moral entre los obreros de dos clases de fábricas. En la primera, los obreros tienen puestos especializados de acuerdo con una división del trabajo muy acentuada. Cada obrero contribuye con una diminuta parte a la realización del proceso general en una línea de montaje larga. En la otra clase de fábricas, cada obrero desempeña muchas tareas y todo el proceso está a cargo de equipos de trabajo pequeños.

¿Cómo vamos a medir la *moral*? Si seguimos una estrategia, observaríamos a los obreros de cada fábrica para anotar si se hacen bromas, sonríen mucho, etc. Podríamos preguntarles si les gusta su trabajo e incluso si preferirían el sistema actual o el otro del estudio. Al comparar lo que observamos en las fábricas, llegaríamos a una conclusión sobre qué proceso de montaje eleva más la moral.

Ahora consideremos algunos problemas de confiabilidad inherentes a este método. Primero, nuestros sentimientos al realizar las observaciones de seguro matizarán lo que veamos. Tal vez interpretemos mal lo que veamos. Quizá los trabajadores bromeaban y creímos que estaban discutiendo o estaban en un día malo. Si durante varios días observáramos al mismo grupo de trabajadores, podríamos llegar a evaluaciones diferentes cada día. Por otro lado, si varios observadores evalúan la misma conducta, también podrían llegar a conclusiones distintas sobre la moral de los obreros.

Tomemos otra estrategia para evaluar la moral. Supongamos que verificamos los expedientes de la compañía para saber cuántas quejas se plantearon en el sindicato durante un periodo determinado. Es de creer que esto nos daría un indicador de la moral: entre más quejas, menos moral. Esta estrategia de medición parecería más confiable: si contamos las quejas una y otra vez, llegaríamos al mismo resultado.

Si al considerar la segunda estrategia usted de pronto piensa "¡un momento!", está preocupado por la *validez*, no por la confiabilidad. Terminemos nuestro análisis de la confiabilidad para poder ocuparnos de la validez.

En la investigación social, los problemas de confiabilidad aparecen de muchas formas. La confiabilidad nos preocupa cada vez que un solo observador es la fuente de los datos, porque no tenemos ninguna salvaguarda contra el efecto de su subjetividad. No podemos estar seguros sobre qué parte del informe proviene de la situación observada y cuál del observador. Sin embargo, no es sólo un problema de los observadores únicos.

Quienes encuestan en sus investigaciones saben desde hace mucho que los entrevistadores, a causa de sus actitudes y conductas, reciben respuestas distintas de los entrevistados. Asimismo, si realizáramos un estudio de las posturas editoriales sobre algún asunto público, podríamos formar un equipo de codificadores que asumiera el trabajo de leer cientos de editoriales y clasificarlos de acuerdo con su opinión sobre el tema; pero codificarían de modo diverso el mismo editorial. También, si quisiéramos clasificar unos cientos de ocupaciones según un sistema fijo de codificación, digamos las categorías establecidas por el gobierno o el censo, usted y yo no colocaríamos las ocupaciones en las mismas categorías.

Todos estos ejemplos ilustran problemas de confiabilidad. Surgen problemas similares cada vez que le pedimos a las personas que nos den información sobre ellas mismas. Algunas veces formulamos preguntas que no saben cómo contestar: ¿cuántas veces ha ido a la iglesia? Otras, les preguntamos cosas que les parecen totalmente irrelevantes: ¿está satisfecho de las relaciones actuales entre China y Albania? Y otras más exploramos temas tan complicados que quienquiera que tenga una opinión clara al respecto podría interpretar de otra manera la pregunta si la formuláramos de nuevo.

¿Cómo elaborar mediciones confiables? Hay varias técnicas. Primera, al solicitarle información a la gente —si eso pide su diseño de investigación—, tenga cuidado de preguntar sólo lo que es probable que los entrevistados sepan contestar. Pregúnteles cosas importantes para ellos y hágalo en forma clara. En estos casos, el peligro es que la gente le *dará* respuestas, confiables o no. La gente le dirá lo que opina de las relaciones de China con Albania aunque no tenga ni la más vaga idea de cuáles sean.

Por fortuna, los investigadores sociales han concebido varias técnicas para enfrentar el problema básico de la confiabilidad.

Método de test-retest A veces es apropiado hacer la misma medición dos veces. Si usted no contempla que la información cambie, entonces debe recibir la misma respuesta las dos ocasiones. Si las respuestas varían, el método de medición puede, en la medición de la variación, no ser confiable. Veamos un ejemplo.

En su investigación de la evaluación de riesgos para la salud (ERS), una parte de la medicina preventiva, Jeffrey Sacks, W. Mark Krushat y Jeffrey Newman (1980) querían determinar el peligro de varios factores antecedentes y de estilo de vida, para que los médicos aconsejaran apropiadamente a sus pacientes. Si conocieran la vida de sus pacientes, los médicos podrían aconsejarlos sobre sus posibilidades de sobrevivencia y cómo mejorarlas. Desde luego, el objetivo dependía en buena medida de la exactitud de la información reunida sobre cada sujeto del estudio.

Para probar la confiabilidad de su información, Sacks y sus colaboradores pidieron a 207 sujetos que llenaran un cuestionario inicial sobre sus características y conducta. Tres meses después, los mismos sujetos respondieron un cuestionario de seguimiento sobre la misma información y se compararon los resultados de las dos encuestas. En general, apenas 15 por ciento de los sujetos dio la misma información en los dos estudios.

Sacks y sus colaboradores (1980:730) informan lo siguiente:

Casi 10 por ciento de los sujetos señaló una estatura diferente en el segundo cuestionario. La edad de los padres cambió en uno de cada tres sujetos. Un padre envejeció 20 años en tres meses. Uno de cada cinco ex fumadores y ex alcohólicos tenía dificultades para recordar confiablemente su consumo anterior.

Algunos sujetos borraron toda traza de soplos cardíacos, diabetes, enfisema, antecedentes penales e ideas de suicidio que habían revelado antes. La madre de un sujeto, muerta en el primer cuestionario, al parecer estaba viva y coleando en la época del segundo. A una de las encuestadas le faltaba un ovario en el primer estudio pero no en el segundo, mientras que otra a la que no le faltaba en el primer estudio ya no lo tenía en el segundo (ni en los 10 años anteriores). Un sujeto dijo que tenía 55 años en el primer estudio y 50 en el segundo, tres meses después. Uno se pregunta si los conse-

jeros médicos llegarían a tener en sus pacientes siquiera parte del efecto que tenían sus memorias. Así, el método de recopilación de datos no era especialmente confiable.

Método de las mitades Como regla general, siempre es bueno efectuar más de una medición de cualquier concepto social sutil o complicado, como los prejuicios, la enajenación o la clase social. Este proceder dispone los cimientos para otra verificación de la confiabilidad. Digamos que ha elaborado un cuestionario que contiene 10 preguntas que en su opinión miden los prejuicios contra las mujeres. Con la técnica de las mitades, usted dividiría al azar (véase el capítulo 8) las 10 preguntas en dos conjuntos de cinco. Como vimos en nuestra exposición del "carácter intercambiable de los indicadores" de Lazarsfeld, cada conjunto debe proporcionar una buena medición de los prejuicios contra las mujeres y ambos deben coincidir en la manera de clasificar a los sujetos del estudio. Si los dos conjuntos de preguntas miden a la gente en forma distinta, entonces aquí también hay un problema en la confiabilidad de la forma de medir la variable.

Uso de mediciones establecidas Otra forma de garantizar la confiabilidad al obtener información de la gente consiste en usar mediciones que han probado su confiabilidad en investigaciones anteriores. Por ejemplo, si quiere medir la anomia, tal vez convenga seguir los pasos de Srole.

Sin embargo, el uso frecuente de mediciones no garantiza su confiabilidad. Por ejemplo, las pruebas de aptitudes escolares y el inventario multifásico de personalidad de Minnesota (IMPM) han sido aceptados como normas establecidas en sus respectivos campos durante décadas. No obstante, en años recientes les ha hecho falta una revisión detallada para que reflejen los cambios de la sociedad.

Confiabilidad de los asistentes También es posible que la falta de confiabilidad proceda de los asistentes de la investigación, por ejemplo, entrevistadores y codificadores. Hay varias soluciones. Para precaverse de la falta de confiabilidad del entrevistador, es una costumbre común en las encuestas hacer que un supervisor llame por teléfono a una muestra de los entrevistados y verifique partes selectas de la información.

La repetición funciona también en otras situaciones. Si a usted le preocupa que las ocupaciones o los editoriales de los periódicos no estén bien clasificados, ¿por qué no hace que los codifiquen varios codificadores? Los que generen más desacuerdos deben ser evaluados con más cuidado y resolverse en consecuencia.

Por último, la claridad, la especificidad, la capacitación y la práctica evitarán muchos problemas de falta de confiabilidad y fracasos. Si usted y yo dedicamos un tiempo a ponernos de acuerdo sobre cómo evaluar las posturas editoriales sobre cierta materia —si analizamos varias posturas y las leemos juntos con detenimiento—, es más probable que hagamos un buen trabajo al clasificarlos del mismo modo por nuestra parte.

La confiabilidad de las mediciones es un tema fundamental de la investigación social, y regresaremos a él más de una vez en los capítulos siguientes. Por ahora, recordemos que incluso la total confiabilidad no garantiza que nuestras mediciones midan lo que pensamos que miden. Estamos listos ya para sumergirnos en la cuestión de la validez.

Validez

En el uso convencional, el término *validez* se refiere al grado en que una medición empírica refleja adecuadamente el *significado real* del concepto que consideramos. ¡Vaya! Ya nos habíamos adherido a la opinión de que los conceptos no tienen significado real. ¿Cómo, entonces, pretendemos llegar a saber si una medición refleja adecuadamente el significado del concepto? Desde luego, en última instancia no podemos. Al mismo tiempo, como vimos, toda la vida social, incluyendo la investigación social, opera según consensos sobre los términos que empleamos y los conceptos que representan. Hay varios criterios para determinar nuestro éxito en la medición adecuada de estos consensos.

Primero, existe algo llamado **validez patente**. Las mediciones empíricas pueden concordar o no con nuestros consensos y nuestras imágenes mentales respecto a cierto concepto. Usted y yo podríamos referir por la suficiencia de medir la moral de los obreros contando las quejas presentadas en el sindicato, pero desde luego que coincidimos en que el número de quejas tiene que ver con la moral. Si propusiera que para medir la moral averiguáramos cuántos libros pidieron en la biblioteca los trabaja-

dores en su tiempo libre, usted me haría una objeción más seria: esa medición no tendría mucha validez patente.

Segundo, ya señalé muchos de los consensos más formales y establecidos que definen algunos conceptos. Por ejemplo, el censo ha preparado definiciones operacionales de conceptos como familia, hogar y ocupación que tienen una validez práctica en la mayor parte de los estudios que los utilizan.

Edward Carmines y Richard Zeller (1979) analizan otras tres formas de validez: *validez de criterio*, *de constructo* y *de contenido*.

La **validez de criterio**, a veces llamada *validez predictiva*, se basa en algún criterio externo. Por ejemplo, la validez de la junta escolar se muestra en su capacidad de predecir el éxito universitario de los estudiantes. En este sentido, la validez de una prueba escrita de manejo está determinada por la relación entre las calificaciones de la gente en la prueba y su historial posterior de conducción. En estos ejemplos, el éxito en la universidad y la capacidad de conducir son los *criterios*.

Para que ponga a prueba su comprensión de este concepto vea si se le ocurren conductas que sirvan para validar las siguientes actitudes:

Es muy religioso

Respalda por igual a hombres que a mujeres

Apoya a grupos paramilitares de extrema derecha

Está preocupado por el ambiente

A veces es difícil encontrar criterios conductuales para validar las mediciones en forma tan directa como en estos ejemplos. Sin embargo, en tales casos, podemos aproximar los criterios considerando cómo debería relacionarse teóricamente la variable en cuestión con otras variables. La **validez de constructo** se basa en las relaciones lógicas entre las variables.

Por ejemplo, supongamos que usted quiere estudiar la satisfacción matrimonial, sus causas y consecuencias. Como parte del estudio concibe una medición de esta satisfacción y quiere evaluar su validez.

Además de idear su medición, también se formó ciertas expectativas teóricas sobre la relación de la variable *satisfacción matrimonial* con otras variables. Por ejemplo, podría concluir razonablemente que los cónyuges satisfechos tienden menos a la infidelidad que los insatisfechos. Si su medición re-

laciona la fidelidad matrimonial de la manera esperada, constituye una prueba de la validez de constructo de su medición. Sin embargo, si las parejas casadas satisfechas tendieran a ser infieles tanto como las insatisfechas, se pondría en tela de juicio la validez de su medición.

Así, el examen de la validez de constructo confiere un *valor de prueba* de que su medición toca o no la cualidad que quiere medir, sin que sea una prueba definitiva. Si bien afirmo que las pruebas de la validez de constructo son menos sólidas que las de validez de criterio, advierta que hay lugar para los desacuerdos en cuanto a la clase de prueba que representa determinada variable en cierta situación. Menos importante que distinguir ambos tipos es entender la lógica de la validación que comparten: si hemos logrado medir alguna variable, entonces nuestras mediciones deben relacionarse de alguna forma lógica con otras mediciones.

Por último, la **validez de contenido** se refiere al grado en que la medición abarca la gama de significados que comprende el concepto. Por ejemplo, Carmínes y Zeller (1979) señalan que un examen de las capacidades aritméticas no puede limitarse a la suma, sino que también debe cubrir la resta, la multiplicación, la división, etc. Asimismo, si decimos que medimos los prejuicios *en general*, ¿reflejan nuestras mediciones los prejuicios contra los grupos raciales y étnicos, las minorías religiosas, las mujeres, los ancianos y demás?

¿Quién decide qué es válido?

Al iniciar los comentarios anteriores sobre la validez recordé que dependemos de los consensos para determinar lo que es real, y acabamos de ver algunas de las formas en que los científicos sociales se ponen de acuerdo en que han hecho mediciones válidas. Hay todavía otra forma de considerar la validez.

A veces, los investigadores sociales se objetan unos a otros que asumen implícitamente que son superiores de alguna manera a quienes estudian. En efecto, a menudo buscamos motivaciones ocultas de las que no están al tanto los propios actores sociales. *Ustedes creen* que compraron ese Burpo-Blaster por su gran desempeño y buen aspecto, pero *nosotros sabemos* que en realidad lo que quieren es darse una posición social más elevada.

Este sentimiento implícito de superioridad cabe cómodamente en un planteamiento positivista a ul-

tranza (el biólogo se siente superior a la rana de su mesa del laboratorio), pero choca con la postura más humanista y cualitativa que adoptan muchos científicos sociales. Así, por ejemplo, David Silverman (1993:94-95) dice lo siguiente sobre la validez en el contexto de las entrevistas a fondo:

Si vamos a ver a los entrevistados como sujetos que construyen activamente las peculiaridades de su mundo cognoscitivo, entonces uno ha de lograr una intersubjetividad honda en ambos lados para llegar a una profunda comprensión mutua.

Al tratar de comprender la forma en que la gente común conceptúa e impone un sentido a su mundo, los etnometodólogos han pedido a todos los científicos sociales que respeten más estos procesos sociales naturales. Por lo menos, la conducta que acaso parezca irracional según el paradigma del científico puede verse lógica contemplada desde el paradigma del actor.

Como usted sabe, dediqué varios años al estudio de los trances mediumnímicos: el canalizador entra en trance y habla con una voz que dice que pertenece a otra persona, el espíritu de un muerto, un ser de otro planeta, etc. Desde el comienzo, vi con claridad que nunca le "encontraría el sentido" a lo que observara mientras exigiera que se acataran las reglas de la visión del mundo con que crecí. Así, por ejemplo, cuando sondeé el tema de "los buenos y los malos espíritus" quería descubrir los consensos entre los propios espíritus y entre quienes los canalizan.

En última instancia, los investigadores sociales deben considerar tanto a sus colegas como a sus sujetos fuentes de consenso sobre los significados y las mediciones más útiles de los conceptos que estudian. Algunas veces una fuente es más provechosa; otras, una distinta, pero ninguna debe quedar descartada.

Tensión entre confiabilidad y validez

En una nota al pie de estas páginas, quiero señalar que suele haber una tensión entre los criterios de confiabilidad y validez, un estira y afloja. Si recuerda un momento el ejemplo de la medición de la moral en varias fábricas, creo que verá que la estrategia de sumergirse todos los días en la rutina de la línea de montaje, observar qué pasa y hablar con los obreros proporciona una medición más vá-

lida de la moral que contar las quejas. Parece obvio que uno se formaría una idea más clara de cuán alta o baja se encuentra la moral.

Sin embargo, como ya dije, la estrategia de contar sería más confiable. Esta situación manifiesta una tensión más general en las mediciones científicas. Casi todos los conceptos realmente interesantes que queremos estudiar tienen muchos matices sutiles, y es difícil especificar con precisión lo que queremos que signifiquen. Los investigadores suelen decir que estos conceptos poseen una "riqueza de significados". Se han escrito muchísimos libros y artículos sobre el tema de la anomia, pero aún no lo han agotado.

Por tanto, con mucha frecuencia la especificación de mediciones y definiciones operacionales confiables parece robar a los conceptos su riqueza de significados. Después de todo, la moral es mucho más que la falta de quejas presentadas en el sindicato; la anomia es mucho más que los cinco enunciados de Leo Srole. Pero entre más toleremos variaciones y riqueza de significados en un concepto, más posibilidades hay de que no estemos de acuerdo sobre la manera en que se aplica en determinada situación, lo que reduce su confiabilidad.

En cierta medida, este dilema explica la persistencia de dos metodologías muy diferentes en la investigación social: técnicas cuantitativas, nomotéticas y estructuradas, como las encuestas y los cuestionarios, o los métodos cualitativos idiográficos, como las investigaciones de campo y los estudios históricos.

Para estar prevenido, debe guardarse de este dilema duradero e inevitable. Prepárese y enfrentelo. Si no hay un acuerdo claro sobre la forma de medir un concepto, mézalo de varias maneras. Si el concepto tiene varias dimensiones, mézalas todas. Por encima de todo, sepa que el concepto no tiene ningún otro significado aparte del que usted y yo le damos. La utilidad es la única justificación para asignarle cierto sentido a un concepto. Mida los conceptos en la forma que nos sirva para comprender el mundo que nos rodea.

Puntos principales

- Las ideas son imágenes mentales personales que usamos como resúmenes para conjuntar experiencias y observaciones que parecen

compartir algo. Empleamos términos o nombres para referirnos a esas ideas.

- Los conceptos son los significados aceptados que asignamos a los términos para facilitar la comunicación, la medición y la investigación.
- Nuestros conceptos no existen en el mundo real, así que no podemos medirlos directamente.
- Es posible medir lo que resumen nuestros conceptos.
- La conceptualización es el proceso de especificar las vagas imágenes mentales de nuestros conceptos y elegir las clases de observaciones y mediciones que serán apropiadas para nuestra investigación.
- El carácter intercambiable de los indicadores nos permite estudiar y sacar conclusiones sobre los conceptos aunque no estemos de acuerdo en sus definiciones.
- La precisión se refiere a la exactitud de la medición de una observación o la descripción de un atributo. Por ejemplo, describir a una persona como de "1.83 metros de estatura" es más preciso que decir "como de 1.80 metros".
- La confiabilidad se refiere a la probabilidad de que determinado procedimiento de medición arroje la misma descripción de cierto fenómeno si se repitiera. Por ejemplo, calcular la edad de una persona interrogando a sus amigos sería menos confiable que preguntar a la propia persona o consultar su acta de nacimiento.
- La validez es el grado al que cierta medición proporciona datos que se relacionen con los significados aceptados de un concepto. Existen varios medios de determinar la validez: validez patente, de criterio, de contenido y de constructo.
- La creación de mediciones específicas y confiables parece disminuir a veces la riqueza de significados que poseen nuestros conceptos generales. Se trata de un problema inevitable. La mejor solución es utilizar varias mediciones que exploren los diferentes aspectos del concepto.

Preguntas y ejercicios de repaso

1. Tome un concepto de las ciencias sociales, como liberalismo o enajenación, y especifíquelo

de modo que pueda estudiarse en un proyecto de investigación. No olvide especificar las dimensiones que quiere incluir y las que quiere excluir de su conceptualización.

- Busque en un periódico o revista un caso de medición sin validez o confiabilidad. Justifique su elección.
- Vaya a la página en internet de Estudios del Holocausto: Prejuicios (<http://www.socialstudies.com:80/12/126item.s.html>) y eche un vistazo a los materiales que se describen. Haga una lista de las dimensiones que se abordan de los prejuicios.
- Examine las formas en que la Encuesta Social General ha explorado las opiniones sobre la pena de muerte (<http://www.icpsr.umich.edu/gss/subject/cup.pun.htm>). ¿Qué aspectos del tema ha estudiado la ESG?

Proyecto de continuidad

Hay muchas dimensiones en el concepto de igualdad y desigualdad sexual. Redacte una lista de por lo menos cinco dimensiones y proponga una forma de medirlas. Está bien si utiliza diferentes técnicas de investigación para cada una.

Lecturas adicionales

Bohrnstedt, George W., "Measurement", en Peter H. Rossi, James D. Wright y Andy B. Anderson (comps.), *Handbook of Survey Research*, Nueva York, Academic Press, 1983, pp. 70-121. Este ensayo ofrece las bases lógicas y estadísticas de la confiabilidad y la validez en las mediciones.

Carmines, Edward G., y Richard A. Zeller, *Reliability and Validity Assessment*, Beverly Hills, Sage, 1979. En este capítulo examinamos la lógica básica de la confiabilidad y la validez en las mediciones de las ciencias sociales. Carmines y Zeller exploran estos temas con más detalle y estudian algunas formas de calcular matemáticamente la confiabilidad.

Gould, Julius, y William Kolb, *A Dictionary of the Social Sciences*, Nueva York, Free Press, 1964. Referencia fundamental de los acuerdos sociales científicos sobre varios conceptos. Aunque los términos que utilizan los científicos sociales no tienen en última instancia significados "verdaderos", este libro despliega los significados que tienen en mente cuando los emplean.

Grimes, Michael D., *Class in Twentieth-Century American Sociology: An Analysis of Theories and Measurement Strategies*, Nueva York, Praeger, 1991. Este libro ofrece un panorama excelente y de largo plazo sobre la conceptualización mientras el autor examina una variedad de posturas teóricas sobre las clases sociales y las técnicas de medición adecuadas.

Lazarfeld, Paul F., y Morris Rosenberg (comps.), *The Language of Social Research*, Nueva York, Free Press of Glencoe, 1955, sec. I. Excelente y variada colección clásica de descripciones de mediciones específicas en las investigación social del pasado. Estos 14 artículos son relaciones útiles y legibles de operaciones de medición reales efectuadas por científicos sociales, así como análisis más conceptuales de la medición en general.

Silverman, David, *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*, Newbury Park, Cal., Sage, 1993, cap. 7. Este capítulo aborda los temas de confiabilidad y validez específicamente en cuanto a la investigación cualitativa.

U.S. Department of Health and Human Services, *Survey Measurement of Drug Use*, Washington, D.C., Government Printing Office, 1992. Extenso repaso de las técnicas diseñadas y usadas para medir diversas formas de drogadicción.

Wallace, Walter, *The Logic of Science in Sociology*, Chicago, Aldine-Atherton, 1971, cap. 3. Presentación breve y lúcida de la formación de conceptos en el contexto de otras etapas de la investigación. Este análisis relaciona la conceptualización con la observación, por un lado, y con la generalización, por el otro.



Operacionalización

Lo que aprenderá en este capítulo

Ahora pasaremos de la conceptualización a la siguiente etapa en la medición: ver de qué modo los científicos sociales encuentran los conceptos reflejados en el mundo real. En particular, consideraremos algunas destrezas que se requieren para formular preguntas.