

capítulo 9

Recolección de datos cuantitativos

El momento de aplicar los instrumentos de medición y recolectar los datos representa la oportunidad para el investigador de confrontar el trabajo conceptual y de planeación con los hechos.

Roberto Hernández-Sampieri

Proceso de investigación cuantitativa

Paso 8 Recolectar los datos

- Definir la forma idónea de recolectar los datos de acuerdo con el planteamiento del problema y las etapas previas de la investigación.
- Seleccionar o elaborar uno o varios instrumentos o métodos para recolectar los datos requeridos.
- Aplicar los instrumentos o métodos.
- Obtener los datos.
- Codificar los datos.
- Archivar los datos y prepararlos para su análisis por computadora.



Objetivos de aprendizaje

Al terminar este capítulo, el alumno será capaz de:

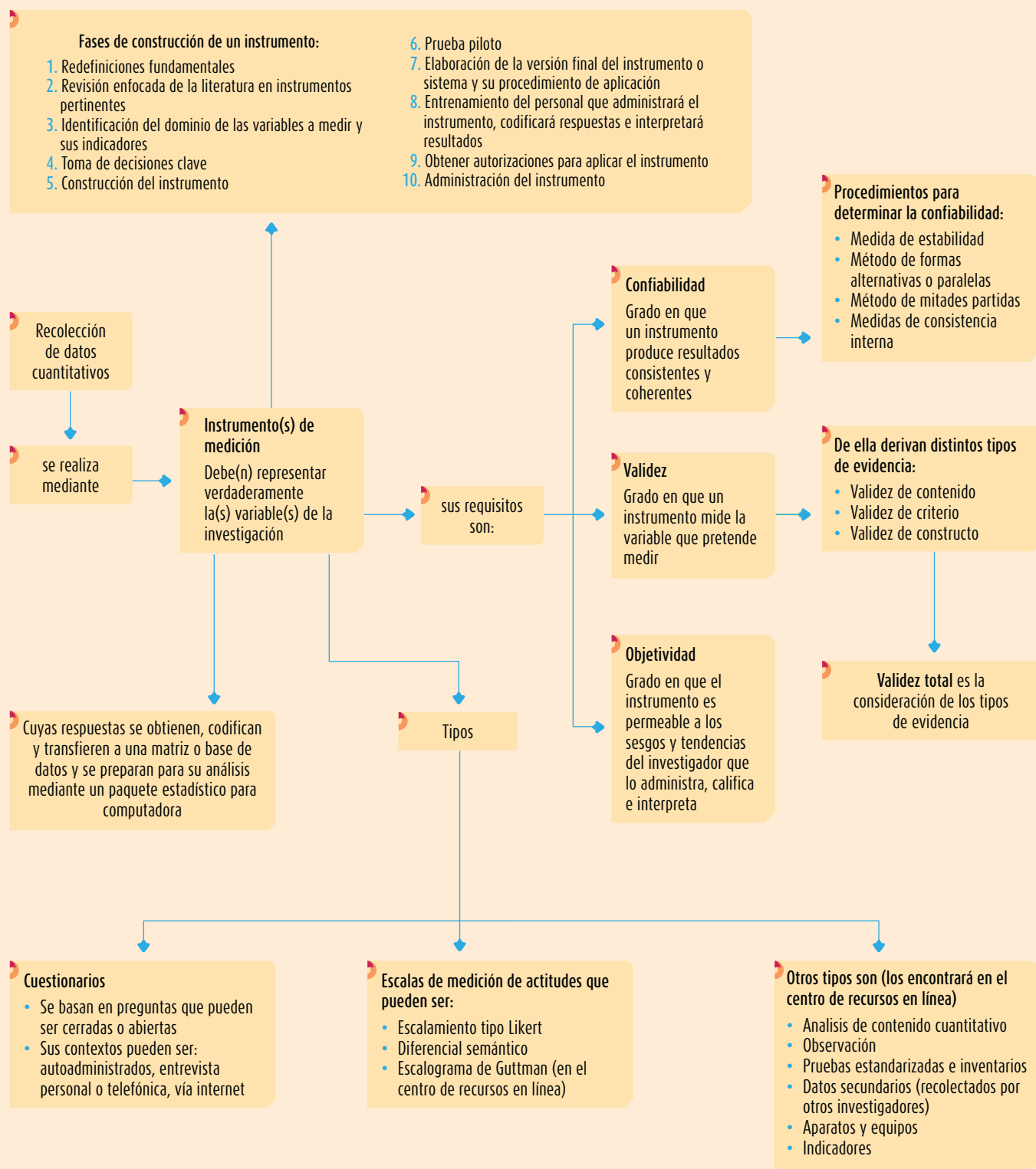
1. Visualizar diferentes métodos para recolectar datos cuantitativos.
2. Entender el significado de “medir” y su importancia en el proceso cuantitativo.
3. Comprender los requisitos que toda recolección de datos debe incluir.
4. Conocer los principales instrumentos para recolectar datos cuantitativos.
5. Elaborar y aplicar los diferentes instrumentos de recolección de datos cuantitativos.
6. Preparar los datos para su análisis cuantitativo.

Síntesis

En el capítulo se analizan los requisitos que un instrumento de medición debe cubrir para recolectar apropiadamente datos cuantitativos: confiabilidad, validez y objetividad. Asimismo, se define el concepto de medición y los errores que pueden cometerse al recolectar datos.

También se explica el proceso para elaborar un instrumento de medición y las principales alternativas para recolectar datos: cuestionarios y escalas de actitudes. Por último, se examina el procedimiento de codificación de datos cuantitativos y la forma de prepararlos para su análisis. Otros temas relacionados con la recolección de los datos se incluyen en el capítulo 5 del centro de recursos en línea.





Nota: El capítulo se termina de integrar con otro que puede descargarse del centro de recursos en línea en: Material complementario → Capítulos → Capítulo 7, “Recolección de los datos cuantitativos: segunda parte”, que contiene alternativas de instrumentos para recolectar datos como el análisis de contenido y los sistemas de observación, además de pruebas e inventarios, escalograma de Guttman (escala de actitudes), datos secundarios e indicadores.

¿Qué implica la etapa de recolección de datos?

-  **1** Una vez que seleccionamos el diseño de investigación apropiado y la muestra adecuada de acuerdo con nuestro problema de estudio e hipótesis (si es que se establecieron), la siguiente etapa consiste en recolectar los datos pertinentes sobre los atributos, conceptos o variables de las unidades de muestreo/ análisis o casos (participantes, grupos, fenómenos, procesos, organizaciones, etcétera).

Recolectar los datos implica elaborar un plan detallado de procedimientos que nos conduzcan a reunir datos con un propósito específico. Este plan incluye determinar:

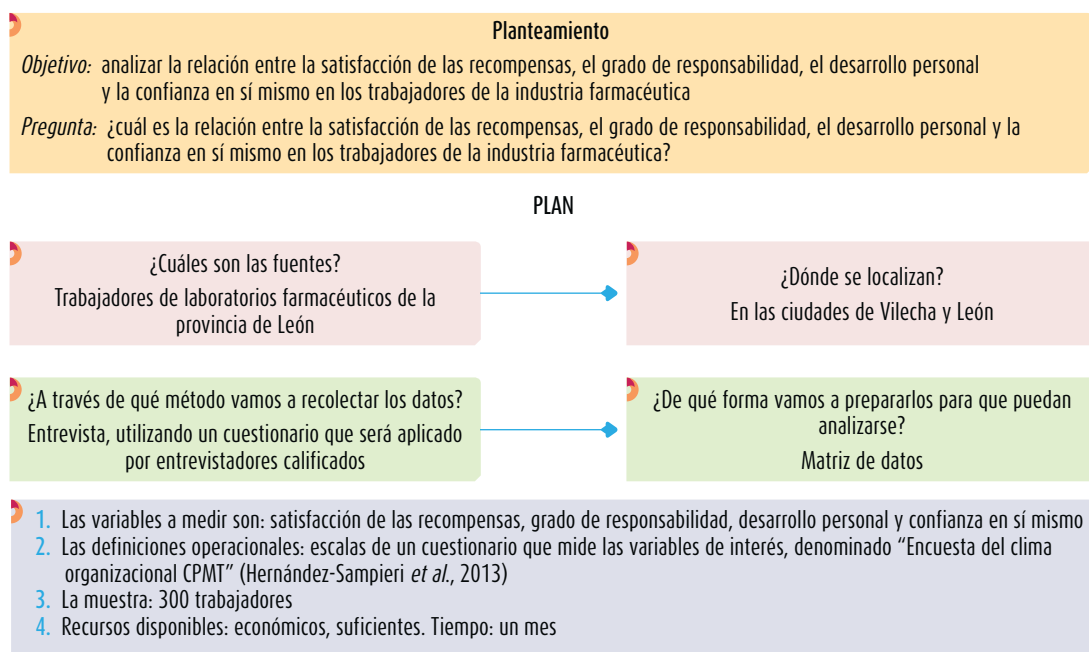
- ¿Cuáles son las fuentes de las que se obtendrán los datos? Es decir, los datos van a ser proporcionados por personas, se producirán de observaciones y registros o se encuentran en documentos, archivos, bases de datos, etcétera.
- ¿En dónde se localizan tales fuentes? Regularmente en la muestra seleccionada, pero es indispensable definir con precisión.
- ¿A través de qué medio o método vamos a recolectar los datos? Esta fase implica elegir uno o varios medios y definir los procedimientos que utilizaremos en la recolección de los datos. El método o métodos deben ser confiables, válidos y “objetivos”.
- Una vez recolectados, ¿de qué forma vamos a prepararlos para que puedan analizarse y respondamos al planteamiento del problema?

El plan se nutre de diversos elementos:

- Las *variables, conceptos o atributos* a medir (contenidos en el planteamiento e hipótesis o directrices del estudio).
- Las *definiciones operacionales*. La manera como hemos operacionalizado las variables es crucial para determinar el método para medirlas, lo cual a su vez, resulta fundamental para realizar las inferencias de los datos.
- La *muestra*.
- Los *recursos disponibles* (de tiempo, apoyo institucional, económicos, etcétera).

El plan se implementa para obtener los datos requeridos, no olvidemos que todos los atributos, cualidades y variables deben ser medibles. En la figura 9.1 se encuentra un ejemplo de un plan de este tipo.

 **Figura 9.1** Ejemplo de plan para la obtención de datos.



Con la finalidad de recolectar datos disponemos de una gran variedad de instrumentos o técnicas, tanto cuantitativas como cualitativas, es por ello que en un mismo estudio podemos utilizar ambos tipos. Incluso, hay instrumentos como la Prueba de Propósito Vital (PIL) (que evalúa el propósito de vida de una persona) de Crumbaugh y Maholick (1969) que contiene una parte cuantitativa y una cualitativa (Thoits, 2012; Marsh, Smith, Piek y Saunders, 2003; y Brown, Ashcroft y Miller, 1998). Esto se revisará en el capítulo 17: “Los métodos mixtos”.

Por ahora, nos concentraremos en la recolección de los datos cuantitativos y sus conceptos esenciales.

¿Qué significa medir?

En la vida diaria medimos constantemente. Por ejemplo, al levantarnos por las mañanas, miramos el reloj despertador y “medimos” la hora; al bañarnos, *ajustamos* la temperatura del agua en la tina o la regadera, *calculamos* la cantidad de café que habremos de servir en la cafetera; nos asomamos por la ventana y estimamos cómo será el día para decidir la ropa que nos pondremos. Al ver el tráfico evaluamos e inferimos a qué hora llegaremos a la universidad o al trabajo, así como la velocidad a la que transitamos (o bien observamos el velocímetro); en ocasiones contamos cuántos anuncios espectaculares observamos en el trayecto u otras cuestiones, incluso deducimos, a partir de ciertos signos, acerca del operador del autobús y de otros conductores: ¿qué tan alegres o enojados están? Como esas, hacemos muchas otras actividades. Medir es parte de nuestras vidas (Bostwick y Kyte, 2005).

En concordancia con la definición clásica del término, ampliamente difundida, **medir** significa “asignar números, símbolos o valores a las propiedades de objetos o eventos de acuerdo con reglas” (Stevens, 1951). Desde luego, no se asignan a los objetos, sino a sus propiedades (Bostwick y Kyte, 2005). Sin embargo, como señalan Carmines y Zeller (1991), esta definición es más apropiada para las ciencias físicas que para las ciencias sociales, ya que varios de los fenómenos que son medidos en éstas no pueden caracterizarse como objetos o eventos, son demasiado abstractos para ello. La disonancia cognitiva, la pareja ideal, el clima organizacional, la cultura fiscal y la credibilidad son conceptos tan abstractos que no deben ser considerados “cosas que pueden verse o tocarse” (definición de objeto) ni solamente como “resultado, consecuencia o producto” (definición de evento) (Carmines y Zeller, 1991). Este razonamiento nos lleva a proponer que es más adecuado definir la **medición** como “el proceso de vincular conceptos abstractos con indicadores empíricos”, el cual se realiza mediante un plan explícito y organizado para clasificar (y con frecuencia cuantificar) los datos disponibles (los indicadores), en términos del concepto que el investigador tiene en mente (Carmines y Zeller, 1991). En este proceso, el instrumento de medición o de recolección de datos tiene un papel central. Sin él, no hay observaciones clasificadas.

La definición sugerida incluye dos consideraciones: la primera es desde el punto de vista empírico y se resume en que el centro de atención es la respuesta observable (sea una alternativa de respuesta marcada en un cuestionario, una conducta registrada mediante observación, un valor de un instrumento y su interpretación o una respuesta dada a un entrevistador). La segunda es desde una perspectiva teórica y se refiere a que el interés se sitúa en el concepto subyacente no observable que se representa por medio de la respuesta. Así, los registros del instrumento de medición representan valores visibles de conceptos abstractos. Un **instrumento de medición** adecuado es aquel que registra datos observables que representan verdaderamente los conceptos o las variables que el investigador tiene en mente (Grinnell, Williams y Unrau, 2009). En términos cuantitativos: capturo verdaderamente la “realidad” que deseo capturar. Bostwick y Kyte (2005) lo señalan de la siguiente forma: la función de la medición es establecer una correspondencia entre el “mundo real” y el “mundo conceptual”. El primero provee evidencia empírica, el segundo proporciona modelos teóricos para encontrar sentido a ese segmento del mundo real que estamos tratando de describir.

En toda investigación cuantitativa aplicamos un instrumento para medir las variables contenidas en las hipótesis (y cuando no hay hipótesis simplemente para medir las variables de interés). Esa



Medición Proceso que vincula conceptos abstractos con indicadores empíricos.

Instrumento de medición Recurso que utiliza el investigador para registrar información o datos sobre las variables que tiene en mente.

medición es eficaz cuando el instrumento de recolección de datos en realidad representa las variables que tenemos en mente. Si no es así, nuestra medición es deficiente; por tanto, la investigación no es digna de tomarse en cuenta. Desde luego, no hay medición perfecta. Es casi imposible que representemos con fidelidad variables como la inteligencia emocional, la motivación, el nivel socioeconómico, el liderazgo democrático, el abuso sexual infantil y otras más; pero es un hecho que debemos acercarnos lo más posible a la representación fiel de las variables que se van a observar mediante el instrumento de medición que desarrollemos. Se trata de un precepto básico del enfoque cuantitativo. Al medir estandarizamos y cuantificamos los datos (Babbie, 2014 y Bostwick y Kyte, 2005).

¿Qué requisitos debe cubrir un instrumento de medición?

-  3
- Toda medición o instrumento de recolección de datos debe reunir tres requisitos esenciales: *confiabilidad*, *validez* y *objetividad*.

Confiabilidad

Confiabilidad Grado en que un instrumento produce resultados consistentes y coherentes.

La **confiabilidad** de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo individuo u objeto produce resultados iguales (Hernández-Sampieri *et al.*, 2013; Kellstedt y Whitten, 2013; y Ward y Street, 2009). Por ejemplo, si se midiera en este momento la temperatura ambiental usando un termómetro y éste indicara que hay 22°C, y un minuto más tarde se consultara otra vez y señalara 5°C, tres minutos después se observara nuevamente y éste indicara 40°C, dicho termómetro no sería confiable, ya que su aplicación repetida produce resultados distintos. Asimismo, si una prueba de inteligencia (*Intelligence Quotient*, IQ) se aplica hoy a un grupo de personas y da ciertos valores de inteligencia, se aplica un mes después y proporciona valores diferentes, al igual que en subsecuentes mediciones, tal prueba no sería confiable (analice los valores de la tabla 9.1, suponiendo que los coeficientes de inteligencia oscilaran entre 100 y 135). Los resultados no son coherentes, pues no se puede “confiar” en ellos.

● **Tabla 9.1** Ejemplo de resultados proporcionados por un instrumento de medición sin confiabilidad

Primera aplicación		Segunda aplicación		Tercera aplicación	
Mariana	135	Sergio	131	Guadalupe	127
Viridiana	125	Laura	130	Guadalupe	127
Sergio	118	Chester	125	Mariana	118
Laura	110	Guadalupe	112	Laura	115
Guadalupe	108	Mariana	110	Chester	112
Chester	106	Viridiana	105	Viridiana	108
Agustín	100	Agustín	101	Sergio	105

La confiabilidad de un instrumento de medición se determina mediante diversas técnicas, las cuales se comentarán brevemente después de revisar los conceptos de validez y objetividad.

Validez

Validez Grado en que un instrumento en verdad mide la variable que se busca medir.

La **validez**, en términos generales, se refiere al grado en que un instrumento mide realmente la variable que pretende medir. Por ejemplo, un instrumento válido para medir la inteligencia debe medir la inteligencia y no la memoria. Un método para medir el rendimiento bursátil tiene que medir precisamente esto y no la imagen de una empresa. Un ejemplo —aunque muy obvio— de completa invalidez sería intentar medir el peso de los objetos con una cinta métrica en lugar de con una báscula.

En apariencia es sencillo lograr la validez. Después de todo, como dijo un estudiante: “pensamos en la variable y vemos cómo hacer preguntas o imaginar indicadores sobre esa variable”. Esto resultaría factible en unos cuantos casos (como lo sería el género al que pertenece una persona). Sin embargo, la situación no es tan simple cuando se trata de variables como la motivación, la calidad del servicio a los clientes, la actitud hacia un candidato político y menos aún con sentimientos y emociones, así como de otras variables con las que trabajamos en todas las ciencias. La validez es una cuestión más compleja que debe alcanzarse en todo instrumento de medición que se aplica. Kerlinger (1979, p. 138) plantea la siguiente pregunta respecto de la validez: ¿está midiendo lo que cree que está midiendo? Si es así, su medida es válida; si no, evidentemente carece de validez.

La validez es un concepto del cual pueden tenerse diferentes tipos de evidencia (Babbie, 2014; Hays, 2013; Kellstedt y Whitten, 2013; The SAGE Glossary of the Social and Behavioral Sciences, 2009a; Streiner y Norman, 2008; Wiersma y Jurs, 2008; Gallestey, 2007; Rupp y Pant, 2006; Carmines y Woods, 2003a; y Gronlund, 1990): 1) *evidencia relacionada con el contenido*, 2) *evidencia relacionada con el criterio* y 3) *evidencia relacionada con el constructo*. A continuación analizaremos cada una de ellas.

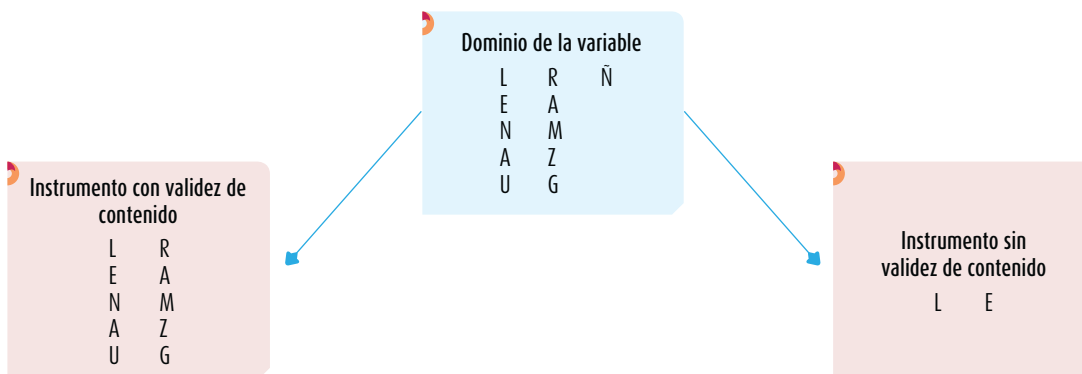
1. Evidencia relacionada con el contenido

La **validez de contenido** se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide. Es el grado en el que la medición representa al concepto o variable medida (The SAGE Glossary of the Social and Behavioral Sciences, 2009b y Bohrnstedt, 1976). Por ejemplo, una prueba de operaciones aritméticas no tendrá validez de contenido si incluyera sólo problemas de resta y excluyera problemas de suma, multiplicación o división. O bien, una prueba de conocimientos sobre las canciones de Los Beatles no deberá basarse solamente en sus álbumes *Let it Be* y *Abbey Road*, sino que debe incluir canciones de todos sus discos. O una prueba de conocimientos de líderes históricos de América Latina que omita a Simón Bolívar, Salvador Allende, Túpac Amaru II, Ernesto el Che Guevara, José de San Martín o Benito Juárez.

Un instrumento de medición requiere tener representados a todos o la mayoría de los componentes del dominio de contenido de las variables que se van a medir. Este hecho se ilustra en la figura 9.2.

Validez de contenido Grado en que un instrumento refleja un dominio específico de contenido de lo que se mide.

● **Figura 9.2** Ejemplo de un instrumento de medición con validez de contenido frente a otro que carece de ella.



El dominio de contenido de una variable normalmente está definido o establecido por la literatura (teoría y trabajos antecedentes). En indagaciones exploratorias en las que las fuentes previas son escasas, el investigador comienza a adentrarse en el problema de estudio y a proponer cómo puede estar constituido tal dominio. De cualquier manera, en cada investigación uno debe probar que el instrumento utilizado es válido. Un ejemplo del intento por establecer el dominio de contenido de una variable es el siguiente:

Ejemplo

Hernández-Sampieri (2005), para establecer el dominio de la variable clima organizacional, revisó 20 estudios clásicos sobre el concepto, comprendidos entre 1964 y 1977, así como más de 100 investigaciones publicadas en revistas científicas entre 1975 y 2005. Por otro lado, consideró diversos libros sobre el tema, tres metaanálisis y otras tantas revisiones del estado del conocimiento sobre dicho clima. También evaluó 15 estudios efectuados en el contexto en el que habría de llevar a cabo su propia investigación. Encontró que en la literatura se han considerado decenas de dimensiones o componentes del clima organizacional, por lo que realizó un análisis para determinar cuáles habían sido los más frecuentes, éstos fueron: 1) moral, 2) apoyo de la dirección, 3) innovación, 4) identificación con la empresa, 5) comunicación, 6) percepción del desempeño, 7) motivación intrínseca, 8) autonomía, 9) satisfacción general, 10) liderazgo, 11) visión y 12) recompensas o retribución. Dejó a un lado otros, como confianza en sí mismo y estándares de excelencia o conformidad. Con lo anterior generó su instrumento de medición.

Si el dominio de un instrumento es demasiado estrecho con respecto al dominio de la variable, el primero no la representará. La pregunta que se responde con la **validez de contenido** es: ¿el instrumento mide adecuadamente las principales dimensiones de la variable en cuestión? En un cuestionario, por ejemplo, cabría interrogar: ¿qué tan bien representan las preguntas a todas las que pudieran hacerse?

Validez de criterio Validez que se establece al correlacionar las puntuaciones resultantes de aplicar el instrumento con las puntuaciones obtenidas de otro criterio externo que pretende medir lo mismo.

2. Evidencia relacionada con el criterio

La **validez de criterio** de un instrumento de medición se establece al comparar sus resultados con los de algún criterio externo que pretende medir lo mismo. Supongamos que Fernando trata de “medir” el grado en que es aceptado por Laura. Entonces decide que va a tomarla de la mano y observará su reacción. Supuestamente, si ella no retira la mano, esto indicaría cierta aceptación. Pero para asegurarse de que su medición es válida, decide utilizar otra forma de medición adicional, por ejemplo, mirarla fijamente sin apartar la vista de sus ojos. En apariencia, si Laura le sostiene la mirada, esto sería otro indicador de aceptación. Así, su medición de aceptación se valida mediante dos métodos al comparar dos criterios. El ejemplo tal vez sea simple, pero describe la esencia de la validez relativa al criterio.

Este criterio es un estándar con el que se juzga la validez del instrumento (Jackson, 2011 y The SAGE Glossary of the Social and Behavioral Sciences, 2009c). Cuanto más se relacionen los resultados del instrumento de medición con los del criterio, la validez será mayor. Por ejemplo, un investigador valida un examen sobre manejo de aviones al mostrar la exactitud con la que el examen predice qué tan bien un grupo de pilotos es capaz de operar un aeroplano.

Si el criterio se fija en el presente de manera paralela, se habla de **validez concurrente** (Kaplan y Saccuzzo, 2013) (los resultados del instrumento se correlacionan con el criterio en el mismo momento o punto de tiempo). Por ejemplo, Núñez (2001) desarrolló una herramienta para medir el sentido de vida de acuerdo con las ideas de Viktor Frankl, el test Celaya. Para aportar evidencia de validez en relación con su instrumento, lo aplicó y a su vez administró otros instrumentos que miden conceptos muy similares, tal como la PIL (Prueba de Propósito Vital) de Crumbaugh y Maholick (1969) y el Logo Test de Lukas (1984). Luego comparó las puntuaciones de los participantes en las tres pruebas, demostró que las correlaciones entre las puntuaciones eran significativamente elevadas. De esta manera fue como aportó validez concurrente para su instrumento.

Si el criterio se fija en el futuro, se habla de **validez predictiva** (The SAGE Glossary of the Social and Behavioral Sciences, 2009d). Por ejemplo, una prueba para determinar la capacidad gerencial de candidatos a ocupar altos puestos ejecutivos se validaría comparando sus resultados con el desempeño posterior de los ejecutivos en su trabajo regular. Un cuestionario para detectar las preferencias del electorado por los distintos partidos contendientes y por sus candidatos en la época de las campañas, puede validarse comparando sus resultados con los resultados finales y definitivos de la elección.

El principio de la validez de criterio es sencillo: si diferentes instrumentos o criterios miden el mismo concepto o variable, deben arrojar resultados similares. Bostwick y Kyte (2005) lo expresan de la siguiente forma:

Si hay validez de criterio, las puntuaciones obtenidas por ciertos casos en un instrumento deben estar correlacionadas y predecir las puntuaciones de estos mismos casos logradas en otro criterio.

La pregunta que se responde con la validez de criterio es: ¿en qué grado el instrumento comparado con otros criterios externos mide lo mismo?

3. Evidencia relacionada con el constructo

La **validez de constructo** es probablemente la más importante, sobre todo desde una perspectiva científica, y se refiere a qué tan bien un instrumento representa y mide un concepto teórico (Babbie, 2014; Grinnell, Williams y Unrau, 2009; The SAGE Glossary of the Social and Behavioral Sciences, 2009d; y Sawilowsky, 2006). A esta validez le concierne en particular el significado del instrumento, esto es, qué está midiendo y cómo opera para medirlo. Integra la evidencia que soporta la interpretación del sentido que poseen las puntuaciones del instrumento (Messick, 1995).

Parte del grado en el que las mediciones del concepto proporcionadas por el instrumento se relacionan de manera consistente con mediciones de otros conceptos o variables vinculadas empírica y teóricamente (por la teoría, modelos e hipótesis previas o investigaciones antecedentes). A tales conceptos se les denomina “constructos”. Un **constructo** es una variable medida y tiene lugar dentro de una hipótesis, teoría o modelo teórico. Es un atributo que no existe aislado sino en relación con otros y debe ser inferido de la evidencia que tenemos en nuestras manos y que proviene de las puntuaciones del instrumento aplicado.

La validez de constructo incluye tres etapas (Carmines y Zeller, 1991):

1. Sobre la base de la revisión de la literatura, se establece y especifica la relación entre el concepto o variable medida por el instrumento y los demás conceptos incluidos en la teoría, modelo teórico o hipótesis.
2. Se asocian estadísticamente los conceptos y se analizan cuidadosamente las correlaciones.
3. Se interpreta la evidencia empírica de acuerdo con el nivel en el que se clarifica la validez de constructo de una medición en particular.

El proceso de *validación de un constructo* está vinculado con la teoría. No es conveniente llevar a cabo tal validación, a menos que exista un marco teórico que soporte la variable en relación con otras variables. Desde luego, no es necesaria una teoría muy desarrollada, pero sí investigaciones que hayan demostrado que los conceptos se asocian. Cuanto más elaborada y comprobada se encuentre la teoría que apoya la hipótesis, la validación del constructo arrojará mayor luz sobre la validez general de un instrumento de medición. Tenemos más confianza en la validez de constructo de una medición cuando sus resultados se correlacionan significativamente con un mayor número de mediciones de variables que, en teoría y de acuerdo con estudios antecedentes, están relacionadas. Veamos la validez de constructo con el ejemplo ya comentado sobre el clima organizacional.

Ejemplo

En el caso de Hernández-Sampieri (2005), que como recordamos, desarrolló un instrumento para evaluar al clima organizacional considerando 12 variables: moral, apoyo de la dirección, innovación, etc., la pregunta obvia es: ¿tal instrumento realmente mide el clima organizacional? ¿Verdaderamente lo representa? En cuanto a contenido se demostró que sí reflejaba las principales dimensiones del clima organizacional. Pero esto no es suficiente, necesita demostrar que su instrumento concuerda con la teoría. Ésta, basada en diversos estudios, indica que tales dimensiones se encuentran fuertemente vinculadas y que se unen o “funden” entre sí para formar un constructo multidimensional denominado clima organizacional, y que además se asocian con el involucramiento en el trabajo y el compromiso organizacional. Entonces, para aportar validez de constructo, se correlacionaron todas las dimensiones entre sí y luego la escala de clima con dicho involucramiento y compromiso. Tales vínculos se encontraron mediante análisis estadístico y los resultados coincidieron con la teoría y se obtuvo evidencia sobre la validez de constructo del instrumento.

Validez de constructo Debe explicar cómo las mediciones del concepto o variable se vinculan de manera congruente con las mediciones de otros conceptos correlacionados teóricamente.

Constructo o construcción Variable medida que tiene lugar dentro de una hipótesis, teoría o modelo teórico.

Las preguntas que se responden con la validez de constructo son: ¿el concepto teórico está realmente reflejado en el instrumento? ¿Qué significan las puntuaciones del instrumento? ¿El instrumento mide el constructo y sus dimensiones? ¿Por qué? ¿Cómo opera el instrumento?

Validez de expertos Grado en que un instrumento realmente mide la variable de interés, de acuerdo con expertos en el tema.

Otro tipo de validez que algunos autores consideran es la **validez de expertos** o *face validity*, la cual se refiere al grado en que aparentemente un instrumento mide la variable en cuestión, de acuerdo con “voces calificadas”. Se encuentra vinculada a la validez de contenido y, de hecho, se consideró por muchos años como parte de ésta. Hoy se concibe como un tipo adicional de evidencia (Gravetter y Forzano, 2011; Streiner y Norman, 2008; y Mostert, 2006). Regularmente se establece mediante la evaluación del instrumento ante expertos. Por ejemplo, Hernández-Sampieri (2005) sometió el instrumento a revisión por parte de asesores en desarrollo organizacional, académicos y gerentes de recursos humanos. Asimismo, más recientemente se ha hablado de la *validez consecuente*, que se refiere a las secuelas sociales del uso e interpretación de una prueba (Mertens, 2010).

Validez total

La validez de un instrumento de medición se evalúa sobre la base de todos los tipos de evidencia. Cuanta mayor evidencia de validez de contenido, de validez de criterio y de validez de constructo tenga un instrumento de medición, éste se acercará más a representar las variables que pretende medir.

Validez total = validez de contenido + validez de criterio + validez de constructo

Relación entre la confiabilidad y la validez

Un instrumento de medición puede ser confiable, pero no necesariamente válido (un aparato, por ejemplo, quizá sea consistente en los resultados que produce, pero puede no medir lo que pretende).

Por ello es requisito que el instrumento de medición demuestre ser **confiable y válido**. De no ser así, los resultados de la investigación no deben tomarse en serio.

Para ampliar este comentario, recurriremos a la analogía de Bostwick y Kyte (2005, pp. 108-109). Supongamos que vamos a probar un arma con tres tiradores. Cada uno debe realizar cinco disparos, entonces:

Tirador 1 Sus disparos no impactan en el centro del blanco y se encuentran diseminados por todo el blanco.

Tirador 2 Tampoco impacta en el centro del blanco, aunque sus disparos se encuentran cercanos entre sí, fue consistente, mantuvo un patrón.

Tirador 3 Los disparos se encuentran cercanos entre sí e impactaron en el centro del blanco.

Sus resultados podrían visualizarse como en la figura 9.3, en la cual se vinculan la confiabilidad y la validez.

● **Figura 9.3** Representación de la confiabilidad y la validez.



Factores que pueden afectar la confiabilidad y la validez

Hay diversos factores que llegan a afectar la confiabilidad y la validez de los instrumentos de recolección de los datos e introducen errores en la medición.¹ A continuación se mencionarán los más comunes.

La improvisación. Algunas personas creen que elegir un instrumento de medición o desarrollarlo es algo que puede tomarse a la ligera. Incluso, ciertos profesores piden a los alumnos que construyan instrumentos de medición de un día para otro o, lo que es igual, de una semana a otra, lo cual habla del poco o nulo conocimiento del proceso de elaboración de instrumentos de recolección de los datos. Esta improvisación genera casi siempre instrumentos poco válidos o confiables.

Aun a los investigadores experimentados les toma tiempo desarrollar un instrumento de medición. Además, para construirlo se requiere conocer muy bien la variable que se pretende medir, así como la teoría y la práctica que la sustentan.

Cuando las mediciones se refieren a seres humanos hay otros factores:

- a) *Utilizar instrumentos desarrollados en el extranjero que no han sido validados en nuestro contexto:* cultura y tiempo. Traducir un instrumento, aun cuando adaptemos los términos a nuestro lenguaje y los contextualicemos, no es ni remotamente una validación. Constituye un primer y necesario paso, aunque sólo es el principio. En el caso de traducciones, es importante verificar que los términos centrales tengan referentes con el mismo significado —o alguno muy parecido— en la cultura en la que se va a utilizar dicho instrumento (vincular términos entre la cultura de origen y la cultura destinataria). A veces se traduce, se obtiene una versión y ésta, a su vez, se vuelve a traducir al idioma original (traducción inversa).

Por otra parte, existen instrumentos que fueron validados en nuestro contexto, pero hace mucho tiempo. Hay instrumentos en los que hasta el lenguaje “nos suena anticuado”. Las culturas, los grupos y las personas cambian; y esto debemos tomarlo en cuenta al elegir o desarrollar un instrumento de medición.

- b) *Usar instrumentos inadecuados para las personas a quienes se les aplica:* no son empáticos. Utilizar un lenguaje muy elevado para los sujetos respondientes, no tomar en cuenta diferencias de género, edad, conocimientos, memoria, nivel ocupacional y educativo, motivación para contestar, capacidades de conceptualización y otras diferencias en los participantes, son errores que llegan a afectar la validez y la confiabilidad. Este error ocurre a menudo cuando los instrumentos deben administrarse a niños. Asimismo, hay grupos de la población que requieren instrumentos apropiados para ellos, tal es el caso de las personas con capacidades distintas. En la actualidad se han desarrollado diversas pruebas que las toman en cuenta (por ejemplo, pruebas en sistema Braille para individuos con capacidades visuales distintas o pruebas orales para personas que no pueden escribir). Otro ejemplo son los indígenas o inmigrantes de otras culturas, pues en ocasiones se les aplican instrumentos en un idioma o contexto que no es el suyo.

Quien realiza una investigación debe adaptarse siempre a los participantes y no al revés, ya que es necesario brindarles todo tipo de facilidades. Si éste es el caso, se sugiere consultar a Mertens y McLaughlen (2004), en cuyo libro hay un capítulo dedicado a la recolección de información de personas con capacidades diferentes o de culturas especiales, y a Eckhardt y Anastas (2007). Asimismo, es recomendable revisar la página web de alguna asociación internacional como la American Psychological Association.

- c) *Cuestiones vinculadas con los estilos personales de los participantes* (Bostwick y Kyte, 2005) como: deseabilidad social (tratar de dar una impresión muy favorable a través de las respuestas), tendencia a asentir con respecto a todo lo que se pregunta, dar respuestas inusuales o contestar siempre negativamente.

¹ Se ha omitido intencionalmente la exposición de los errores sistemáticos y no sistemáticos que afectan la confiabilidad y la validez, con objeto de simplificar al lector las explicaciones. Un comentario se incluye en el centro de recursos en línea en: Material complementario → Capítulos → Capítulo 7 “Recolección de los datos cuantitativos: segunda parte”.

- d) *Condiciones en las que se aplica el instrumento de medición.* El ruido, la inadecuada iluminación, el frío (por ejemplo, en una encuesta de casa en casa), un instrumento demasiado largo o tedioso, una encuesta telefónica después de que algunas compañías han utilizado el mercadeo telefónico en exceso y a destiempo (promocionar servicios a las 7 a.m. de un domingo o después de las 11 p.m. entre semana) son cuestiones que llegan a afectar negativamente la validez y la confiabilidad, al igual que si el tiempo que se brinda para responder al instrumento es inapropiado. Por lo común en los experimentos se cuenta con instrumentos de medición más largos y complejos que en los diseños no experimentales. Por ejemplo, en una encuesta pública sería muy difícil aplicar una prueba larga o compleja.

Otro factor que se encuentra en todo tipo de instrumentos en cualquier campo de conocimiento es la *falta de estandarización*: que las instrucciones no sean las mismas para todos los participantes, que el orden de las preguntas sea distinto para algunos individuos, que los instrumentos de observación no resulten equivalentes, que el procedimiento para administrar la medición no sea exactamente el mismo para todos los casos (por ejemplo, al medir la presión arterial, que a unos pacientes se les tome después de un tiempo de relajación y a otros, no; o bien, que al pesar ciertas piezas no se utilice la misma báscula). Este elemento también se vincula con la objetividad.

Aspectos mecánicos como que si el instrumento es escrito, que no sean legibles las instrucciones, falten páginas, no haya espacio adecuado para contestar o no se comprendan las instrucciones, también influyen de manera desfavorable. Lo mismo pasaría con un aparato mal calibrado.

Con respecto a la validez de constructo dos factores pueden afectarla significativamente:

- a) La estrechez del contenido, es decir, que se excluyan dimensiones importantes de la variable o las variables medidas.
- b) La amplitud exagerada, donde el riesgo es que el instrumento contenga excesiva intrusión de otros constructos similares.

Muchos de los errores se pueden evitar mediante una adecuada revisión de la literatura, que nos permite seleccionar las dimensiones apropiadas de las variables del estudio, otros criterios para comparar los resultados de nuestro instrumento, teorías de respaldo, instrumentos de dónde elegir, etcétera.

Objetividad

Es difícil lograr la objetividad, particularmente en el caso de las ciencias sociales. En ciertas ocasiones se alcanza mediante el consenso o mediciones múltiples. Al tratarse de cuestiones físicas las percepciones suelen compartirse (por ejemplo, la mayoría de las personas estarían de acuerdo en que el agua de mar contiene sal o los rayos del Sol queman), pero en temas que tienen que ver con la conducta humana como los valores, las atribuciones y las emociones, el consenso es más complejo. Imaginemos que 10 observadores deben ver una película y calificarla como “muy violenta”, “violenta”, “neutral”, “poco violenta” y “nada violenta”. Tres personas indican que es muy violenta, tres que es violenta y cuatro la evalúan como neutral; es difícil responder qué tan violenta es la película. O bien, ¿quién fue mejor compositor: Mozart, Beethoven o Bach? Todo es relativo. Sin embargo, la objetividad aumenta al reducirse la incertidumbre (Unrau, Grinnell y Williams, 2011). Desde luego, la certidumbre total no existe en ninguna ciencia; el conocimiento es aceptado como verdadero, hasta que nueva evidencia demuestra lo contrario.

En un instrumento de medición, la **objetividad** se refiere al grado en que éste es o no permeable a la influencia de los sesgos y tendencias del investigador o investigadores que lo administran, califican e interpretan (Mertens, 2010). Investigadores racistas o machistas quizás influyan negativamente por su sesgo contra un grupo étnico o el género femenino. Lo mismo podría suceder con las tendencias ideológicas,

políticas, religiosas o la orientación sexual. En este sentido, los aparatos y sistemas calibrados (por ejemplo, una pistola láser para medir la velocidad de un automóvil) son más objetivos que otros sistemas que requieren cierta interpretación (como un detector de mentiras) y éstos, a su vez, más objetivos que las pruebas estandarizadas, las cuales son menos subjetivas que las pruebas proyectivas.

Objetividad del instrumento Grado en que el instrumento es o no permeable a la influencia de los sesgos y tendencias de los investigadores que lo administran, califican e interpretan.

La objetividad se refuerza mediante la estandarización en la aplicación del instrumento (mismas instrucciones y condiciones para todos los participantes) y en la evaluación de los resultados; así como al emplear personal capacitado y experimentado en el instrumento. Por ejemplo, si se utilizan observadores, su proceder en todos los casos debe ser lo más similar posible y su entrenamiento tendrá que ser profundo y adecuado.

Los estudios cuantitativos buscan que la influencia de las características y las tendencias del investigador se reduzca al mínimo posible, lo que insistimos es un ideal, pues la investigación siempre es realizada por seres humanos.

La validez, la confiabilidad y la objetividad no deben tratarse de forma separada. Sin alguna de las tres, el instrumento no es útil para llevar a cabo un estudio.

¿Cómo se sabe si un instrumento de medición es confiable y válido?

En la práctica es casi imposible que una medición sea perfecta y generalmente se tiene un grado de error. Desde luego, se trata de que este error sea el mínimo posible, por lo cual la medición de cualquier fenómeno se conceptualiza con la siguiente fórmula básica:

$$X = t + e$$

Donde X representa los valores observados (resultados disponibles); t , los valores verdaderos; y e , el grado de error en la medición. Si no hay error de medición (e es igual a cero), el valor observado y el verdadero son equivalentes. Esto puede verse claramente así:

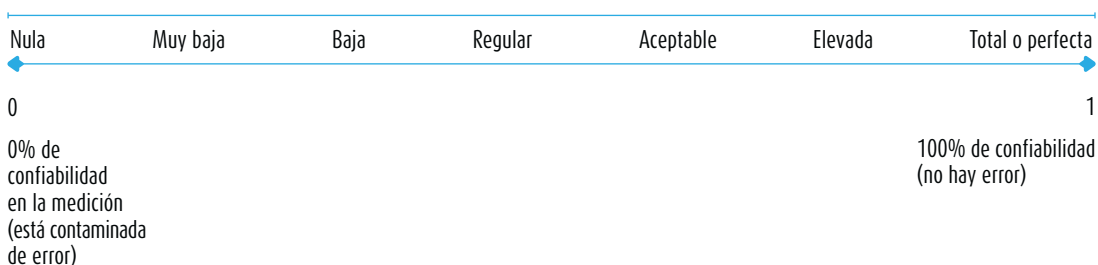
$$\begin{aligned} X &= t + 0 \\ X &= t \end{aligned}$$

Esta situación representa el ideal de la medición. Cuanto mayor sea el error al medir, el valor que observamos (en el cual nos basamos) se alejará más del valor real o verdadero. Por ejemplo, si medimos la motivación de un individuo y la medición está contaminada por un grado de error considerable, la motivación registrada por el instrumento será bastante diferente de la motivación real de esa persona. Por ello, es importante que el error se reduzca lo más posible. Pero, ¿cómo sabemos el grado de error que tenemos en una medición? Al calcular la confiabilidad y la validez.

Cálculo de la confiabilidad o fiabilidad

Hay diversos procedimientos para calcular la confiabilidad de un instrumento de medición. Todos utilizan procedimientos y fórmulas que producen coeficientes de fiabilidad. La mayoría oscilan entre cero y uno, donde un coeficiente de cero significa nula confiabilidad y uno representa un máximo de confiabilidad (fiabilidad total, perfecta). Cuanto más se acerque el coeficiente a cero, mayor error habrá en la medición. Esto se ilustra en la figura 9.4.

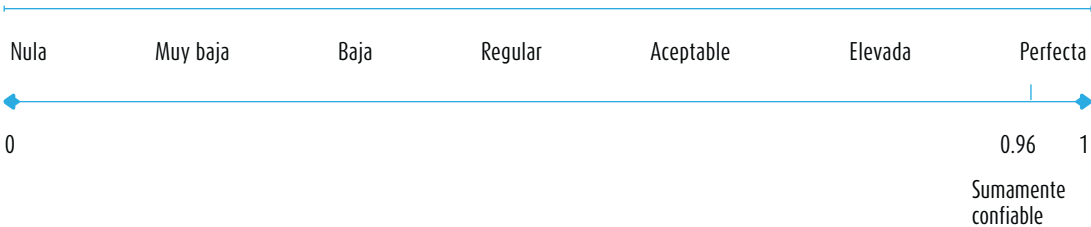
● **Figura 9.4** Interpretación de un coeficiente de confiabilidad.



Los procedimientos más utilizados para determinar la confiabilidad mediante un coeficiente son: 1) *medida de estabilidad* (confiabilidad por *test-retest*), 2) *método de formas alternativas o paralelas*, 3) *método de mitades partidas* (*split-halves*) y 4) *medidas de consistencia interna*. Estos procedimientos no se detallan en esta sección, sino que se explican en el capítulo 10, “Análisis de los datos cuantitativos”, debido a que requieren conocimientos de ciertos conceptos estadísticos.² Simplemente se comentará su interpretación con la medida de congruencia interna denominada “coeficiente alfa Cronbach”, que tal vez es la más utilizada.

En el caso de mediciones sociales, supongamos que una investigadora desarrolló un instrumento para medir el grado de “amor romántico” entre parejas de jóvenes universitarios, el cual se fundamentó en cuatro de las herramientas tal vez más conocidas para ello: la medida de Rubin sobre el amar y el vincularse con los demás, la escala sobre actitudes hacia el amor, la medida sobre el amor apasionado y la escala del amor triangular (Graham y Christiansen, 2009). Para estimar la confiabilidad de su instrumento lo debe aplicar a su muestra y sobre la base de los resultados calcular tal coeficiente. Imaginemos que obtiene un valor alfa Cronbach de 0.96, que es muy elevado, lo que significa que su medida del “amor romántico” es sumamente confiable, esto se representa en la figura 9.5.

● **Figura 9.5** Interpretación de un coeficiente de confiabilidad sobre un instrumento que mide el “amor romántico”.



La confiabilidad varía de acuerdo con el número de indicadores específicos o ítems³ que incluya el instrumento de medición. Cuantos más ítems haya, mayor tenderá a ser ésta, lo cual resulta lógico. Veámoslo con un ejemplo cotidiano: si se desea probar qué tan confiable o constante es la lealtad de un amigo hacia nuestra persona, cuantas más pruebas le pongamos, su fiabilidad será mayor. Claro está que demasiados ítems provocarán cansancio en los participantes.

Cada vez que se administra un instrumento de medición debe calcularse la confiabilidad, al igual que evaluarse la evidencia sobre la validez.

Cálculo de la validez

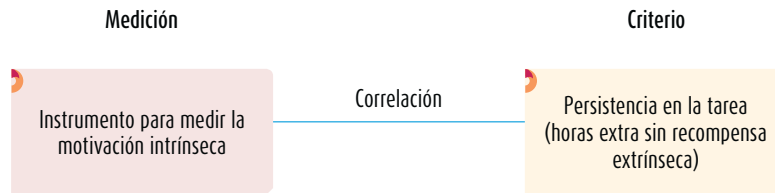
Con respecto a la validez de contenido, primero es necesario revisar cómo han medido la variable otros investigadores. Y, con base en dicha revisión, elaborar un universo de ítems o reactivos posibles para medir la variable y sus dimensiones (el universo debe ser lo más exhaustivo posible). Después, se consulta a investigadores familiarizados con la variable para ver si el universo es verdaderamente exhaustivo. Se seleccionan los ítems bajo una cuidadosa evaluación, uno por uno. Y si la variable está compuesta por diversas dimensiones o facetas, se extrae una muestra probabilística de reactivos, ya sea al azar o estratificada (cada dimensión constituiría un estrato). Se administran los ítems, se correlacionan las puntuaciones de éstos entre sí (tiene que haber correlaciones altas, en especial entre ítems que miden una misma dimensión, pero teniendo cuidado que sean capaces de discriminar entre participantes o casos) (Punch, 2009); y se hacen estimaciones estadísticas para ver si la muestra es representativa. Para calcular la validez de contenido son necesarios varios coeficientes. Éste sería un procedimiento

² Tales procedimientos se aplican a instrumentos o sistemas constituidos por escalas con varios reactivos, indicadores o mediciones (mediciones compuestas).
³ Un ítem es la unidad mínima que compone una medición; es un reactivo que estimula una respuesta en un sujeto (por ejemplo, una pregunta, una frase, una lámina, una fotografía o un objeto de descripción).

ideal. Pero, como veremos más adelante, a veces no se calculan estos coeficientes, sino que se seleccionan los ítems mediante un proceso que asegura la representatividad (no de manera estadística sino conceptual).

La *validez de criterio* se estima al correlacionar la medición con el criterio externo (puntuaciones del instrumento frente a las puntuaciones en el criterio), y este coeficiente se toma como coeficiente de validez (Bohrnstedt, 1976). Que podría representarse con el ejemplo de la figura 9.6.⁴

● **Figura 9.6** Ejemplo para la estimación de la validez de criterio.



La validez de constructo suele determinarse mediante procedimientos de análisis estadístico multivariado (“análisis de factores”, “análisis discriminante”, “regresiones múltiples”, etc.), los cuales podrá bajar del centro de recursos en línea en: Material complementario → Capítulos → Capítulo 8: “Análisis estadístico: segunda parte”.

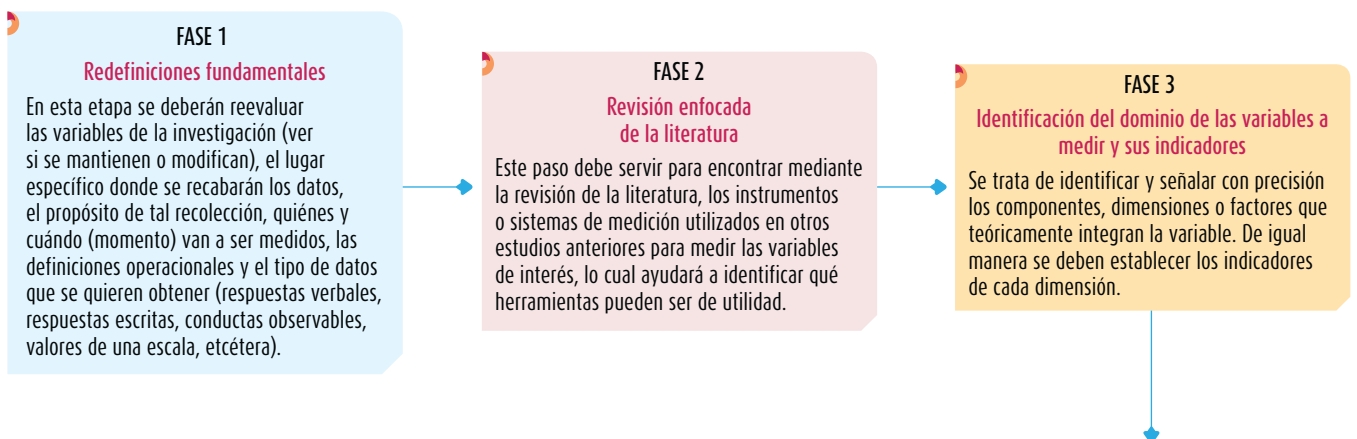


¿Qué procedimiento se sigue para construir un instrumento de medición?

Existen diversos tipos de instrumentos de medición, cada uno con características diferentes. Sin embargo, el procedimiento general para construirlos y aplicarlos es semejante. Éste se resume mediante etapas en el diagrama de la figura 9.7 y corresponde a la parte del plan de recolección que contesta a la pregunta: ¿a través de qué método vamos a recolectar los datos? Y cabe señalar que cada etapa o fase no se detalla en este capítulo, sino en el centro de recursos en línea en: Material complementario → Capítulos → Capítulo 7: “Recolección de los datos cuantitativos: segunda parte”.



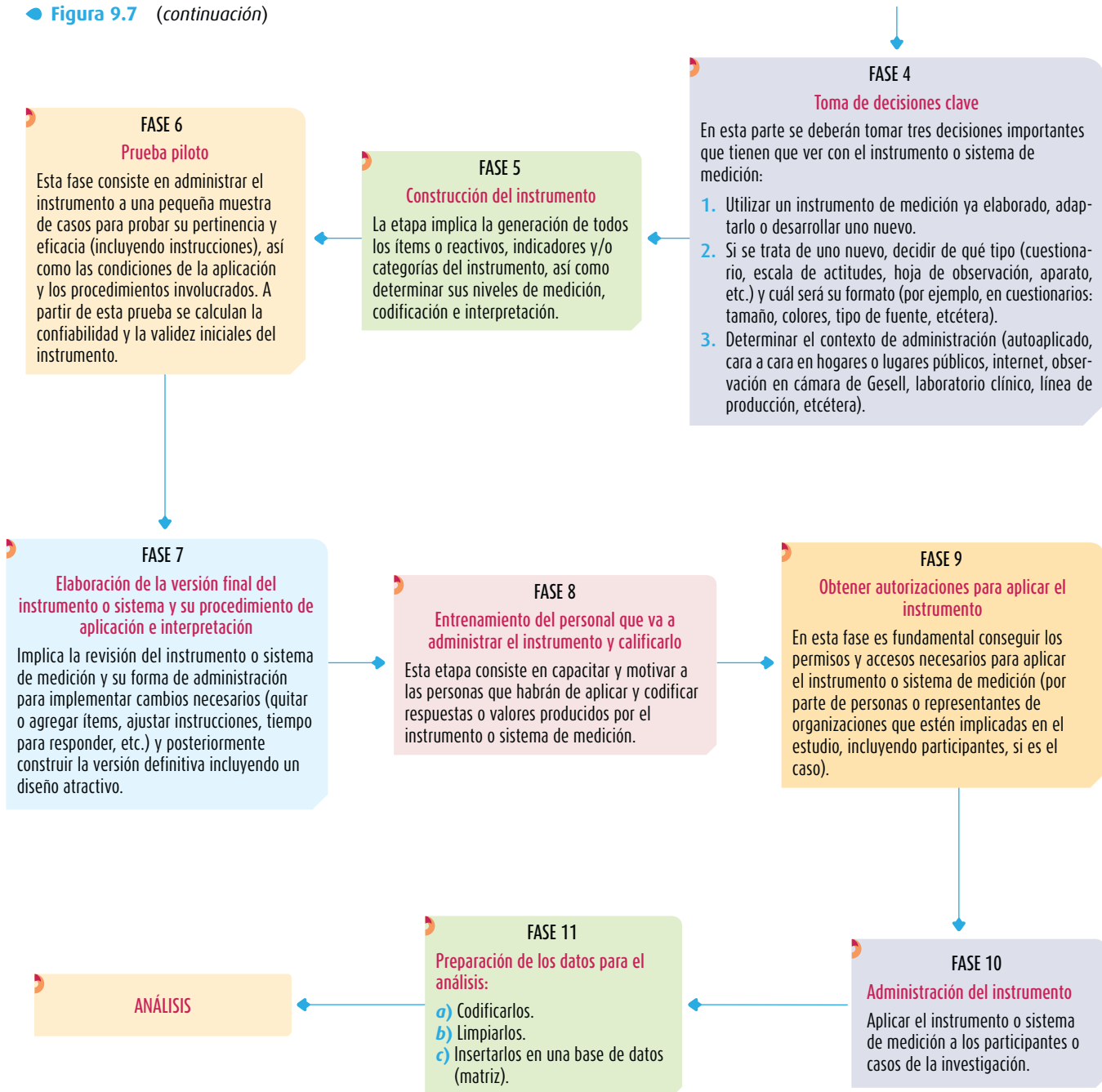
● **Figura 9.7** Proceso para construir un instrumento de medición.



(continúa)

⁴ Las pruebas de correlación se presentan en el siguiente capítulo: “Análisis de los datos cuantitativos”.

Figura 9.7 (continuación)



Las fases 1 a 7 del diagrama se refieren propiamente al desarrollo del instrumento o sistema de medición, mientras que las etapas 8 a 11 representan su administración y la preparación de los datos para su análisis.

Tres cuestiones fundamentales para un instrumento o sistema de medición

Existen tres cuestiones básicas respecto a las variables que deben considerarse al momento de construir un instrumento: *a)* la operacionalización, *b)* la codificación y *c)* establecer los niveles de medición.

La operacionalización: el tránsito de la variable al ítem o valor

El paso de una variable teórica a indicadores empíricos verificables y medibles e ítems o equivalentes se le denomina *operacionalización* (Solís, 2013). La operacionalización se fundamenta en la definición conceptual y operacional de la variable. Su proceso se amplía en el capítulo 7 adicional que se puede descargar del centro de recursos en línea. Por ahora, se comentará que cuando se construye un instrumento, el proceso más lógico para hacerlo es transitar de la variable a sus dimensiones o componentes, luego a los indicadores y finalmente a los ítems o reactivos y sus categorías. En la tabla 9.2 se dan ejemplos de este tránsito.

► **Tabla 9.2** Ejemplo de desarrollo de ítems

Estudio de las preferencias de los jóvenes para divertirse (ejemplo sencillo) ⁵			
Variable	Dimensión	Indicadores	Ítems
Preferencia de actividad para salir con alguien del género opuesto	Actividad nocturna preferida entre semana	Jerarquía de preferencias de actividades de lunes a jueves (aunque algunos comienzan el fin de semana desde el jueves).	De lunes a jueves, ¿cuál sería tu actividad preferida nocturna para salir con el chico o chica que más te gusta? (marcar la que más te agrade). 1. Salir a cenar a un restaurante. 2. Ir al cine. 3. Ir a un bar, “antro”, <i>grill</i> , taberna, <i>pub</i> o cervecería. 4. Acudir a una cafetería, heladería o a algún establecimiento en el que sirvan bebidas no alcohólicas. 5. Ir a bailar a una discoteca, disco o “antro”. 6. Ir a una fiesta privada. 7. Acudir al teatro. 8. Acudir a un concierto. 9. Pasear por un parque, jardín o avenida. 10. Otra (especificar).
	Actividad nocturna preferida en fin de semana	Jerarquía de preferencias de actividades en viernes y sábado.	Mismas categorías u opciones de respuesta.
	Actividad nocturna preferida en domingo	Jerarquía de preferencias de actividades en domingo	Mismas categorías u opciones de respuesta.
Clima organizacional	Moral	Grado en que los miembros de una organización o departamento perciben que colaboran y cooperan entre sí, se apoyan mutuamente y mantienen relaciones de amistad y compañerismo.	• Mis compañeros de trabajo son mis amigos. 5. Totalmente de acuerdo. 4. De acuerdo. 3. Ni de acuerdo ni en desacuerdo. 2. En desacuerdo. 1. Totalmente en desacuerdo. • En mi trabajo hay mucho compañerismo (mismas opciones de respuesta que el ítem anterior). • Siempre que lo necesito mis compañeros de trabajo me brindan apoyo (mismas opciones de respuesta).

(continúa)

⁵ En los ejemplos de la tabla sólo se incluyen unos cuantos ítems o reactivos por cuestiones de espacio, son ejemplos muy resumidos.

● Tabla 9.2 (continuación)

Investigación sobre el clima organizacional			
Variable	Dimensión	Indicadores	Ítems
	Autonomía	Grado de libertad percibida para tomar decisiones y realizar el trabajo.	• En el departamento donde trabajo nos mantenemos unidos (mismas opciones de respuesta). • La mayoría de las veces en mi departamento compartimos la información más que guardarla para nosotros (mismas opciones de respuesta). • ¿Qué tanto apoyo le brindan sus compañeros cuando usted lo necesita? 5. Total. 4. Bastante. 3. Aceptablemente. 2. Poco. 1. Ninguno.
	Atribución del desempeño	Grado de conciencia compartida por desempeñarse con calidad en las tareas laborales, sobre la base de la cooperación.	• En esta empresa tengo libertad para tomar decisiones que tienen que ver con mi trabajo. • En esta empresa todos tratamos de hacer bien nuestro trabajo (mismas opciones de respuesta que el ítem anterior.) • En esta empresa todos queremos dar lo mejor de nosotros en el trabajo (mismas opciones.) 5. Totalmente de acuerdo. 4. De acuerdo. 3. Ni de acuerdo ni en desacuerdo. 2. En desacuerdo. 1. Totalmente en desacuerdo.

En ocasiones tenemos menos elementos, por ejemplo, sólo variable, dimensiones e ítems; incluso, únicamente variable y categorías de un único ítem o escala. Tal sería el caso del género —variable—, integrada por un ítem solitario —pregunta u observación— y dos categorías: masculino y femenino; o bien, temperatura —variable— y la lectura de un valor o categoría. Algunas variables son sencillas (un ítem) y otras compuestas (varias dimensiones, indicadores, ítems y categorías). En el siguiente capítulo “Análisis de los datos cuantitativos” (en el apartado “Apunte 1”), se discuten las implicaciones de lo anterior para el análisis estadístico. En el ejemplo de Lee y Guerin (2009) tratado en esta obra, la variable “satisfacción respecto a la calidad del diseño ambiental del interior del área de trabajo u oficina”, incluyó la dimensión “satisfacción con respecto al diseño de la oficina”, compuesta por tres indicadores: *a)* satisfacción sobre el espacio aprovechable para el trabajo individual y para guardar o almacenar, *b)* satisfacción en torno al nivel de privacidad visual y *c)* satisfacción en relación con la facilidad para interactuar con compañeros de trabajo; cada uno con uno o varios ítems. Ésta y otras variables fueron correlacionadas con la variable tipo de oficina, integrada por un ítem con las siguientes categorías:

1. Despacho u oficina privada cerrada.
2. Oficina cerrada compartida con otras personas.
3. Cubículo con paneles altos de metro y medio o más.
4. Cubículo con paneles bajos de menos de metro y medio.
5. Espacio de trabajo en una oficina abierta sin divisiones.
6. Otros.

Codificación

Codificar los datos significa asignarles un valor numérico o símbolo que los represente. Es decir, a las categorías (opciones de respuesta o valores) de cada ítem o variable se les asignan valores numéricos o signos que tienen un significado. Por ejemplo, si tuviéramos la variable “género” con sus respectivas categorías, masculino y femenino, a cada categoría le determinaríamos un valor. Esto podría ser:

Categoría	Codificación (valor asignado)
• Masculino	1
• Femenino	2

Por tanto, Paola Yáñez en la variable género sería 2, Luis Gerardo Vera y Ramón Calderón serían 1, Mabel González Bencosme 2 y así sucesivamente.

Otro ejemplo sería la variable “horas de exposición diaria a la televisión”, que podría codificarse como se muestra en la tabla 9.3.

► **Tabla 9.3** Ejemplo de codificación

Categoría	Codificación (valor asignado)
— No ve televisión	0
— Menos de una hora	1
— Una hora	2
— Más de una hora, pero menos de dos	3
— Dos horas	4
— Más de dos horas, pero menos de tres	5
— Tres horas	6
— Más de tres horas, pero menos de cuatro	7
— Cuatro horas	8
— Más de cuatro horas	9

En el primer ejemplo de la tabla 9.2, la respuesta a la pregunta: *De lunes a jueves, ¿cuál sería tu actividad nocturna preferida para salir con el chico o chica que más te gusta?*, la codificación era con números (1 = salir a cenar a un restaurante; 2 = ir al cine; 3 = ir a un bar, “antro”, grill, taberna, pub o cervecería; 4 = acudir a una cafetería, heladería o establecimiento en el que sirvan bebidas no alcohólicas; 5 = ir a bailar a una discoteca, disco o “antro”; 6 = ir a una fiesta privada; 7 = acudir al teatro; 8 = acudir a un concierto; 9 = pasear por un parque, jardín, avenida, y 10 = otra).

Mientras que en el ítem: “En esta empresa tengo libertad para tomar decisiones que tienen que ver con mi trabajo”, la codificación era:

5. Totalmente de acuerdo
4. De acuerdo
3. Ni de acuerdo ni en desacuerdo
2. En desacuerdo
1. Totalmente en desacuerdo

Es necesario insistir que cada ítem o variable sencilla deberá tener una codificación (códigos numéricos o simbólicos) para sus categorías, a esto se le conoce como “precodificación”. Desde luego, hay veces que un ítem no puede ser codificado *a priori* (precodificado), porque es muy difícil conocer cuáles serán sus categorías. Por ejemplo, si en una investigación preguntáramos: ¿qué opina del programa económico que recientemente aplicó el Gobierno? Las categorías podrían ser muchas más de

Codificación Significa asignar a los datos un valor numérico o símbolo que los represente, ya que es necesario para analizarlos cuantitativamente.

las que nos imaginemos y resultaría difícil predecir con precisión cuántas y cuáles serían. En tales situaciones, la codificación se lleva a cabo una vez que se aplica el ítem (*a posteriori*). Éste es el caso de algunos ítems que por ahora denominaremos “abiertos”.

La codificación es necesaria para analizar cuantitativamente los datos (aplicar análisis estadístico). A veces se utilizan letras o símbolos en lugar de números (*, A, Z). La codificación puede o no incluirse en el instrumento de medición, veámoslo con un ejemplo de pregunta:

Pregunta precodificada
¿Tiene usted novia?

☐ 1 Sí
☐ 0 No

Pregunta no precodificada
¿Tiene usted novia?

☐ Sí
☐ No

Asimismo, es muy importante indicar el nivel de medición de cada ítem y, por ende, el de las variables, porque es parte de la codificación y dependiendo de dicho nivel se selecciona uno u otro tipo de análisis estadístico (por ejemplo, la prueba estadística para correlacionar dos variables de intervalo es muy distinta de la prueba para correlacionar dos variables ordinales). De esta forma, es necesario hacer una relación de variables, ítems y niveles de medición.

Niveles de medición

Existen cuatro niveles de medición ampliamente conocidos.⁶

1. *Nivel de medición nominal.* En este nivel hay dos o más categorías del ítem o la variable. Las categorías no tienen orden ni jerarquía. Lo que se mide (objeto, persona, etc.) se coloca en una u otra categoría, lo cual indica tan sólo diferencias respecto de una o más características. Por ejemplo, la variable “género” de la persona posee sólo dos categorías: masculino y femenino. Ninguna de las categorías implica mayor jerarquía que la otra. Las categorías únicamente reflejan diferencias en la variable. No hay orden de mayor a menor.



Si les asignamos una etiqueta o un símbolo a cada categoría, esto identificará exclusivamente a la categoría. Por ejemplo:

* = Masculino
z = Femenino

Si usamos numerales, es lo mismo:

1 = Masculino 2 = Masculino
 es igual a
2 = Femenino 1 = Femenino

Los números utilizados en este nivel de medición tienen una función puramente de clasificación y *no* se pueden manipular de manera aritmética. Por ejemplo, la afiliación religiosa es una variable nominal; si pretendiéramos operarla de forma aritmética se presentarían situaciones tan ridículas como ésta:

⁶ Para ampliar este tema se sugiere Issel (2014), Wolf (2008), Stoutenborough (2008), Dykema, Blixt y Stevenson (2008); Scanlan (2008), Gershkoff (2008) y Rasmussen (2006).

1 = Católico	
2 = Judío	1 + 2 = 3
3 = Protestante	
4 = Musulmán	¿Un católico + un judío = un protestante?
5 = Otros	(No tiene sentido)

Las variables nominales pueden incluir dos categorías (dicotómicas), o bien, tres o más categorías (categóricas). Ejemplos de variables nominales dicotómicas serían el género, el veredicto de un jurado (culpable-no culpable) y el tipo de escuela a la que se asiste (privada-pública); y como ejemplos de variables nominales categóricas tendríamos la afiliación política (partido A, partido B, etc.), la licenciatura estudiada, el grupo étnico, el departamento, la provincia o el estado de nacimiento, la clase de material de construcción ("no" su resistencia, que sería otra variable), tipo de medicamento administrado ("no" la dosis, que sería una variable distinta), bloques de mercado (asiático, latinoamericano, comunidad europea, etc.) y el canal de televisión preferido.

2. *Nivel de medición ordinal.* En este nivel hay varias categorías, pero además mantienen un orden de mayor a menor. Las etiquetas o los símbolos de las categorías sí indican jerarquía. Por ejemplo, el prestigio ocupacional en Estados Unidos se ha medido por diversas escalas que reordenan las profesiones de acuerdo con su prestigio, por ejemplo:⁷

Valor en escala	Profesión
90	Ingeniero químico
80	Científico de ciencias naturales (excluyendo la química)
60	Actor común
50	Operador de estaciones eléctricas de potencia
02	Manufacturero de tabaco

Los números (símbolos de categorías) definen posiciones, en el ejemplo: 90 es más que 80, 80 más que 60, 60 más que 50 y así sucesivamente. Sin embargo, las categorías no están ubicadas a intervalos iguales (no hay un intervalo común). No podríamos decir con exactitud que entre un actor (60) y un operador de estaciones eléctricas (50) existe la misma distancia en prestigio que entre un científico de ciencias naturales (80) y un ingeniero químico (90). Al parecer, en ambos casos la distancia es 10, pero no es una distancia real. Otra escala⁸ clasificó el prestigio de dichas profesiones de la siguiente manera:

Valor en escala	Profesión
98	Ingeniero químico
95	Científico de ciencias naturales (excluyendo la química)
84	Actor común
78	Operador de estaciones eléctricas de potencia
13	Manufacturero de tabaco

Aquí la distancia entre un actor (84) y un operador de estaciones (78) es de seis, y la distancia entre un ingeniero químico (98) y un científico de ciencias naturales (95) es de tres. Otro ejemplo sería la variable *posición jerárquica en la empresa*:

Presidente	10	Jefe	5
Vicepresidente	9	Empleado A	4
Director general	8	Empleado B	3
Gerente de área	7	Empleado C	2
Subgerente o superintendente	6	Intendencia	1

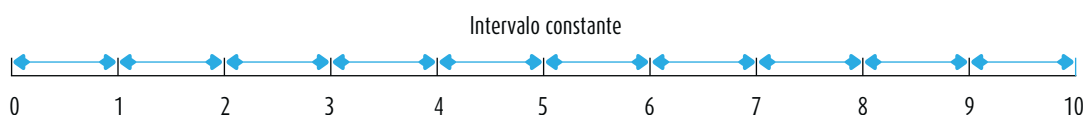
⁷ Duncan (1977). La escala incluye múltiples profesiones, se presentan sólo algunas.

⁸ Nam et al. (1965) y Nam (1983). Para escalas más complejas y actuales, véase Organización Internacional del Trabajo (2012).

Sabemos que el presidente (10) es más que el vicepresidente (9); éste, más que el director general (8); a su vez este último, más que el gerente (7) y así sucesivamente; pero no se precisa en cada caso cuánto más. Tampoco pueden utilizarse las operaciones aritméticas básicas: no podríamos decir que 4 (empleado A) + 5 (jefe) = 9 (vicepresidente), ni que $10 \text{ (presidente)} \div 5 \text{ (jefe)} = 2 \text{ (empleado C)}$. Sería absurdo, no tiene sentido. Otros ejemplos de este nivel serían: la medición por rangos de las preferencias de marcas de bebidas refrescantes con gas (refrescos o sodas), auto-percepción del grado de dolor de cabeza, percepción de sabores respecto a un platillo y jerarquización de valores (en primer lugar, en segundo lugar, en tercero).

3. *Nivel de medición por intervalos.* Además del orden o la jerarquía entre categorías, se establecen intervalos iguales en la medición. Las distancias entre categorías son las mismas a lo largo de toda la escala, por lo que hay un intervalo constante, una unidad de medida.

● **Figura 9.8** Escala con intervalos iguales entre categorías.



Por ejemplo, en una prueba de resolución de problemas matemáticos (30 problemas de igual dificultad). Si Ana Cecilia resolvió 10, Laura resolvió 20 y Abigail, 30. La distancia entre Ana Cecilia y Laura es igual a la distancia entre Laura y Abigail.

Sin embargo, en este nivel de medición el cero (0) es arbitrario, no es real, ya que se asigna arbitrariamente a una categoría el valor de cero y a partir de ésta se construye la escala. Un ejemplo clásico en ciencias naturales es la temperatura, que puede medirse en grados centígrados y Fahrenheit: el cero es arbitrario, pues no implica que en realidad haya cero (ninguna) temperatura (incluso en ambas escalas el cero es diferente).

Cabe agregar que diversas mediciones en el estudio del comportamiento humano no son verdaderamente de intervalo (por ejemplo, escalas de actitudes, pruebas de inteligencia y de otros tipos); pero para algunos autores se acercan a este nivel y se suele tratarlas como si fueran mediciones de intervalo.⁹ Esto se hace porque este nivel de medición permite utilizar las operaciones aritméticas básicas y diversos métodos estadísticos, que de otro modo no se utilizarían. Aunque algunos investigadores no están de acuerdo con suponer tales mediciones como si fueran de intervalo. El producto interno bruto o producto nacional bruto estaría en este estadio.

4. *Nivel de medición de razón.* En este nivel, además de tenerse todas las características del nivel de intervalos (unidad de medida común, intervalos iguales entre las categorías y aplicación de operaciones aritméticas básicas y sus derivaciones), el cero es real y absoluto (no es arbitrario). Cero absoluto implica que hay un punto en la escala donde está ausente o no existe la propiedad medida (véase la figura 9.9).

● **Figura 9.9** Ejemplo de escala para el nivel de medición de razón.



Ejemplos de estas mediciones serían la exposición a la televisión (en minutos), el número de hijos, las ventas de un producto, los metros cuadrados de construcción, escala de temperatura de Kelvin, ingresos en una unidad monetaria (alguien puede tener cero ingresos), consumo de un medicamento específico, etcétera.

⁹ Este comentario nos remite a una polémica que se comentará más adelante en el tema de la escala de Likert.

¿De qué tipos de instrumentos de medición o recolección de datos cuantitativos disponemos en la investigación?

En la investigación disponemos de múltiples tipos de instrumentos para medir las variables de interés y en algunos casos llegan a combinarse varias técnicas de recolección de los datos. A continuación describimos brevemente las más usuales.

Los instrumentos que serán revisados en este capítulo son: cuestionarios¹⁰ y escalas de actitudes, mencionando a otros. En el capítulo adicional 7, “Recolección de los datos cuantitativos, segunda parte”, que se puede descargar en el centro de recursos en línea, se comentan los siguientes: registros del contenido (análisis de contenido) y observación cuantitativa, pruebas estandarizadas, recolección de información factual e indicadores (análisis de datos secundarios de registros públicos y documentación) y metaanálisis, así como otras clases de mediciones. Los metodólogos de cada ciencia y disciplina han desarrollado una gran diversidad de instrumentos y procedimientos, algunos de los cuales puede buscarlos el lector en los manuales y ejemplos del centro de recursos (<http://www.mhhe.com/he/hmi6e>).

La codificación y la preparación de los datos obtenidos se discutirán después de presentar los principales instrumentos de medición.

Cuestionarios

En fenómenos sociales, tal vez el instrumento más utilizado para recolectar los datos es el cuestionario. Un **cuestionario** consiste en un conjunto de preguntas respecto de una o más variables a medir (Chasteauneuf, 2009). Debe ser congruente con el planteamiento del problema e hipótesis (Brace, 2013).

Los cuestionarios se utilizan en encuestas de todo tipo (por ejemplo, para calificar el desempeño de un gobierno, conocer las necesidades de hábitat de futuros compradores de viviendas y evaluar la percepción ciudadana sobre ciertos problemas como la inseguridad). Pero también, se implementan en otros campos. Por ejemplo, un ingeniero en minas usó un cuestionario como herramienta para que expertos de diversas partes del mundo aportaran opiniones calificadas con el fin de resolver ciertas problemáticas de producción.

Ahora, comentaremos primero sobre las preguntas y luego sobre las características deseables de este tipo de instrumento, así como los contextos en los cuales se pueden administrar los cuestionarios.

¿Qué tipos de preguntas se pueden elaborar?

El contenido de las preguntas de un cuestionario es tan variado como los aspectos que mide. Básicamente se consideran *dos tipos de preguntas: cerradas y abiertas*.

Preguntas cerradas

Las **preguntas cerradas** contienen categorías u opciones de respuesta que han sido previamente delimitadas. Es decir, se presentan las posibilidades de respuesta a los participantes, quienes deben acotarse a éstas. Pueden ser dicotómicas (dos posibilidades de respuesta) o incluir varias opciones de respuesta. Ejemplos de preguntas cerradas dicotómicas serían:

- | | |
|-----------------------------|---|
| ¿Estudia usted actualmente? | ¿Durante la semana pasada vio la final de la Liga de Campeones de Europa? |
| () Sí | () Sí |
| () No | () No |



4



4

Cuestionario Conjunto de preguntas respecto de una o más variables que se van a medir.



5

Preguntas cerradas Son aquellas que contienen opciones de respuesta previamente delimitadas. Resultan más fáciles de codificar y analizar.

¹⁰ Las entrevistas se plantean como un contexto en el cual pueden ser administrados los cuestionarios.

Ejemplos de preguntas cerradas con varias opciones de respuesta serían:

Como usted sabe, todos los países desarrollados reciben inmigrantes. ¿Cree que, en términos generales, la inmigración es más bien positiva o más bien negativa para estos países?

- ☐ Positiva
- ☐ Ni positiva ni negativa
- ☐ Negativa
- ☐ No sabría decir

¿Cuál es el puesto que ocupa usted en su empresa?

- | | |
|--|---|
| ☐ Presidente/Director general | ☐ Vicepresidente/Director corporativo |
| ☐ Subdirector/Director/Gerente | ☐ Subgerente/Superintendente |
| ☐ Coordinador | ☐ Jefe de área |
| ☐ Supervisor | ☐ Empleado |
| ☐ Obrero | ☐ Otro (especificar) _____ |

Si usted tuviera elección, ¿preferiría que su salario fuera de acuerdo con su productividad en el trabajo?

- ☐ Definitivamente sí
- ☐ Probablemente sí
- ☐ No estoy seguro
- ☐ Probablemente no
- ☐ Definitivamente no

Como se observa, en las preguntas cerradas las categorías de respuesta son definidas *a priori* por el investigador y se le muestran al encuestado, quien debe elegir la opción que describa más adecuadamente su respuesta. Gambara (2002) hace notar algo muy lógico pero que en ocasiones se descuida y resulta fundamental: cuando las preguntas presentan varias opciones, éstas deben recoger todas las posibles respuestas.

Ahora bien, hay preguntas cerradas en las que el participante puede seleccionar más de una opción o categoría de respuesta (*posible multirrespuesta*).

Ejemplo

Supongamos que un entrevistador pregunta:

¿Esta familia tiene en el hogar...? (Marque con una cruz o tache todas las opciones que el entrevistado o entrevistada señale que tiene en su hogar):

- | | |
|---|---|
| ☐ Radio | ☐ Televisión |
| ☐ Reproductor de DVD o Blu-ray | ☐ TV de paga (SKY, Cablevisión, DirectTV, otros sistemas locales de cable o TV satelital) |
| ☐ Computadora fija | ☐ Internet |
| ☐ Tableta, laptop o iPad | ☐ Equipo de sonido para CD |
| ☐ Teléfono celular o móvil | ☐ Teléfono (línea telefónica en casa) |
| ☐ iPod u otro dispositivo similar | |

En preguntas como la del ejemplo anterior, los participantes pueden marcar una, dos, tres, cuatro o más opciones de respuesta. Las categorías no son mutuamente excluyentes.

En ocasiones, el encuestado tiene que jerarquizar opciones.

Ejemplo

De las siguientes compañeras de clase, ¿quién te atrae más? ¿Cuál en segundo lugar? ¿Cuál en tercer lugar? ¿Cuál en cuarto lugar? ¿Cuál en último lugar?

- | | |
|------------------------------|-------------------------------|
| ☐ Sandra | ☐ Mariana |
| ☐ Lucía | ☐ Paola |
| ☐ Ana | |

O bien, en otras preguntas se debe designar un puntaje a una o diversas cuestiones.

Ejemplo

A continuación voy a mencionarle algunos de los problemas que suelen preocupar a los habitantes de este municipio y le pediría que en cada caso me dijera: ¿qué tanto le preocupa a usted cada uno de ellos?; donde 10 significa: “me preocupa muchísimo” y 0 quiere decir: “no me preocupa en absoluto”.¹¹

- | | |
|---|---|
| ☐ Desempleo | ☐ Escasez de agua |
| ☐ Pobreza | ☐ Cortes en el suministro de agua |
| ☐ Inseguridad al transitar por la calle o viajar en transporte público | ☐ Falta de vivienda |
| ☐ Empleo mal remunerado/bajos salarios | ☐ Servicios de salud insuficientes |
| ☐ Robos/asaltos en los hogares y viviendas | ☐ Carencia/deficiencia de servicios educativos |
| ☐ Robos de vehículos/autos, motocicletas, bicicletas | ☐ Drenaje inadecuado en las calles |
| ☐ Pandillerismo | ☐ Tránsito/tráfico/vialidad |
| ☐ Venta de drogas-narcomenudeo | ☐ Pavimentación y bacheo mal hechos |
| ☐ Secuestros | ☐ Falta de infraestructura (calles, puentes, etcétera) |
| ☐ Recolección de la basura (no todos los días la recogen) | ☐ Corrupción de funcionarios municipales, policías, agentes de tránsito y vialidad |
| ☐ Horarios inadecuados para la recolección de la basura | ☐ Situación económica familiar |

En otras preguntas, se anota una cifra dentro de un rango predeterminado.

Ejemplo

Aproximadamente, ¿cuántos minutos dedica diariamente a hacer deporte entre semana, es decir, de lunes a viernes?

En algunas más, el encuestado se ubica en una escala. El concepto de escala (aplicado a la medición) puede definirse como: “sucesión ordenada de valores distintos de una misma cualidad” (Real Academia Española, 2001, p. 949). Es un patrón, conjunto, medida o estimación regular de acuerdo con algún estándar o tasa, respecto de una variable (Carmines y Woods, 2003b). Ejemplos: escala de temperatura en grados centígrados, escala de inteligencia, escala de distancia en kilómetros, metros y centímetros; escala de peso en kilogramos, escala de presión arterial en milímetros de mercurio (mmHg), escala musical con octavas, etcétera.

Ejemplo

¿Qué tan enamorada está usted de su novio? (del 0 al 100):

- | | | |
|---|-------------------------------|----------------------|
| ☺ | 100 – Completamente enamorada | . |
| | 99 | . |
| | 98 | . |
| | . | 20 |
| | . | 10 |
| | . | . |
| | 80 | . |
| | 70 | 2 |
| | 60 | 1 |
| ☹ | 50 | ⊗ 0 – Nada enamorada |

¹¹ Esta pregunta fue administrada a personas que tuvieran estudios mínimos de bachillerato o preparatoria, aunque funcionó con personas cuyo nivel era de secundaria. No se incluyeron todos los problemas por cuestiones de espacio, solamente algunos para ilustrar el tipo de pregunta.

Finalmente, en ocasiones se encadenan varias preguntas en una, como en el siguiente ejemplo (los candidatos y candidatas son ficticios).

Ejemplo

Le voy a mencionar algunos nombres de políticos de nuestro municipio y le pediría que en cada caso me dijera si sabe usted quién es y a qué partido pertenece, así como su opinión de esta persona.

Político (Rotar opciones)	P. 8 ¿Sabe quién es?		Cuando sabe quién es P. 9 ¿Sabe a qué partido pertenece? (No leer opciones)		Cuando sabe quién es P. 10 ¿Qué tan favorable o desfavorable es su opinión acerca de...? (Leer opciones)			
	Sí	No (Pasar a p. 16)	Sí identificó	No identificó	Muy favorable	Favorable	Desfavorable	Muy desfavorable
Guadalupe Méndez Peña	1	2	(Partido 1) 1	2	4	3	2	1
Agustín Almanza Mendoza	1	2	(Partido 2) 1	2	4	3	2	1
Sandra Hernández Jiménez	1	2	(Partido 3) 1	2	4	3	2	1
Roberto Yáñez Ruiz	1	2	(Partido 4) 1	2	4	3	2	1

Preguntas abiertas No delimitan las alternativas de respuesta. Son útiles cuando no hay suficiente información sobre las posibles respuestas de las personas.

Preguntas abiertas

En cambio, las **preguntas abiertas** no delimitan de antemano las alternativas de respuesta, por lo cual el número de categorías de respuesta es muy elevado; en teoría, es infinito, y puede variar de población en población.

Ejemplo

¿Por qué asiste a psicoterapia?

¿Qué opina de las medidas de apoyo a la población que adoptó el Gobierno para disminuir el impacto del último terremoto ocurrido el 10 de noviembre?

¿Conviene usar preguntas cerradas o abiertas?

Un cuestionario obedece a diferentes necesidades y a un problema de investigación, lo cual origina que en cada estudio el tipo de preguntas sea distinto. Algunas veces se incluyen tan sólo preguntas cerradas, otras ocasiones únicamente preguntas abiertas, y en ciertos casos ambos tipos de preguntas. Cada clase de pregunta tiene sus ventajas y desventajas, las cuales se mencionan a continuación.

Las *preguntas cerradas* son más fáciles de codificar y preparar para su análisis. Asimismo, estas preguntas requieren un menor esfuerzo por parte de los encuestados, que no tienen que escribir o verbalizar pensamientos, sino únicamente seleccionar la alternativa que sintetice mejor su respuesta. Responder a un cuestionario con preguntas cerradas toma menos tiempo que contestar uno con preguntas abiertas. Cuando el cuestionario se envía por correo, se tiene un mayor grado de respuesta porque es fácil de contestar y completarlo requiere menos tiempo. Otras ventajas son: se reduce la ambigüedad de las respuestas y se favorecen las comparaciones entre las respuestas (Burnett, 2009).

La *principal desventaja de las preguntas cerradas* reside en que limitan las respuestas de la muestra y, en ocasiones, ninguna de las categorías describe con exactitud lo que las personas tienen en mente; no siempre se captura todo “lo que pasa por la cabeza” de los participantes. Su redacción exige mayor laboriosidad (Vinuesa, 2005).

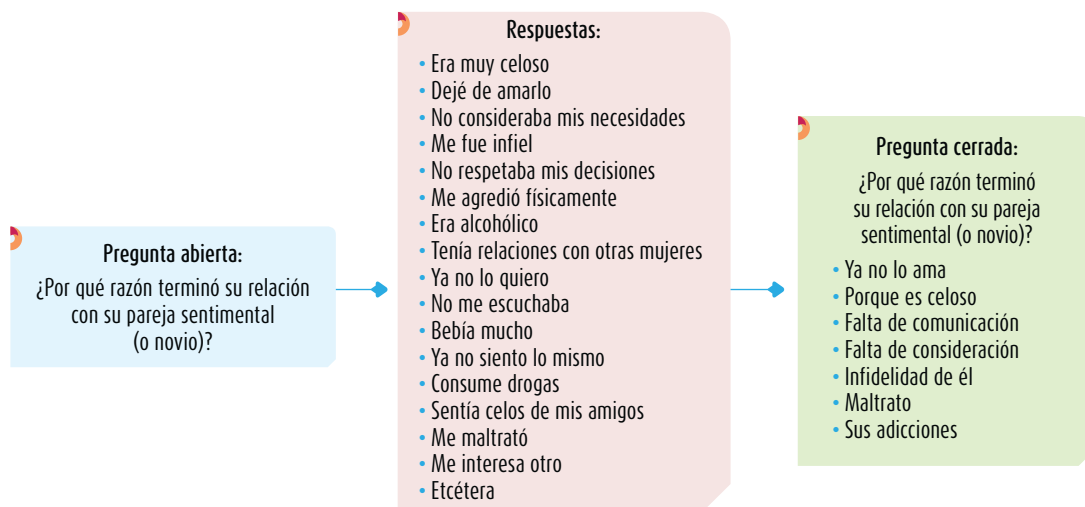
Para *formular preguntas cerradas* es necesario anticipar las posibles alternativas de respuesta. De no ser así, es muy difícil plantearlas. Además, el investigador debe asegurarse de que los participantes a quienes se les administrarán conocen y comprenden las categorías de respuesta. Por ejemplo, si preguntamos qué canal de televisión es el preferido, determinar las opciones de respuesta y que los participantes las comprendan es muy sencillo. Pero si preguntamos sobre las razones que provocan esa preferencia, señalar las alternativas es algo más complejo.

Las *preguntas abiertas* proporcionan una información más amplia y son particularmente útiles cuando no tenemos información sobre las posibles respuestas de las personas o la que tenemos es insuficiente (Phillips, Phillips y Aaron, 2013). También sirven en situaciones donde se desea profundizar una opinión o los motivos de un comportamiento. Su mayor desventaja es que son más difíciles de codificar, clasificar y preparar para el análisis. Además, llegan a presentarse sesgos derivados de distintas fuentes; por ejemplo, quienes enfrentan dificultades para expresarse en forma oral y por escrito quizá no respondan con precisión a lo que en realidad desean, o generen confusión en sus respuestas. El nivel educativo, la capacidad de manejo del lenguaje y otros factores pueden afectar la calidad de las respuestas (Saris y Gallhofer, 2007 y Black y Champion, 1976). Asimismo, responder a preguntas abiertas requiere mayor esfuerzo y más tiempo.

La elección del tipo de preguntas que contenga el cuestionario depende del grado en que se puedan anticipar las posibles respuestas, los tiempos de que se disponga para codificar y si se quiere una respuesta más precisa o profundizar en alguna cuestión. Una recomendación para construir un cuestionario es que se analice, variable por variable, qué tipo de pregunta o preguntas suelen ser más confiables y válidas para medir esa variable, de acuerdo con la situación del estudio (planteamiento del problema, características de la muestra, tipo de análisis a efectuar, etcétera).

Con frecuencia, las preguntas cerradas se construyen con fundamento en preguntas abiertas. Por ejemplo, en la prueba piloto puede elaborarse una pregunta abierta y posteriormente a su aplicación, sobre la base de las respuestas, se genera el ítem cerrado.

● **Figura 9.10** Ejemplo del paso de una pregunta abierta a la elaboración de una pregunta cerrada.



¿Una o varias preguntas para medir una variable?

En ocasiones sólo basta una pregunta para recolectar la información necesaria sobre la variable considerada, no es necesario incluir más. Por ejemplo, para medir el nivel de escolaridad de una muestra basta con preguntar: ¿hasta qué año escolar cursó?, o ¿cuál es su grado máximo de estudios? En otras



ocasiones se requiere elaborar varias preguntas para verificar la consistencia de las respuestas u obtener múltiples indicadores de diversas dimensiones de la variable medida.

Por ejemplo, algunas asociaciones latinoamericanas de investigación de mercados e instituciones educativas miden el nivel socioeconómico tomando en cuenta diversas preguntas o indicadores de la clase social, el estilo de vida, el ingreso y el bienestar:¹²

1. Escolaridad del jefe del hogar.
2. Número de focos en la vivienda.
3. Número de habitaciones en la vivienda, sin incluir baños.
4. Número de baños en general y con regadera.
5. Número de automóviles y otros vehículos en la cochera.
6. Posesión de ciertos aparatos y equipos (tanque de gas estacionario, boiler o calentador, estufa de gas, lavadora automática, aspiradora, horno de microondas, tostadora de pan, etcétera).
7. Posesión de elementos de conectividad y entretenimiento (computadora, televisión, equipo de sonido, equipo de video, línea telefónica, teléfono móvil o celular, internet, servicio de televisión de paga, etcétera).
8. Características de la vivienda (techo estable y seguro en su vivienda —no de cartón ni hule, etc.—, piso firme en su interior —cemento, concreto o piso de mosaico—, agua que llega por tubería a su vivienda, etcétera).
9. Tarjetas de crédito comercial en el hogar.

Con base en estos indicadores se construyen índices, cada uno con un peso o coeficiente, y al final se otorga una puntuación que determina el nivel socioeconómico con mayor precisión.¹³ Sin embargo, esto puede resultar muy complejo para el alumno que comienza con sus primeras investigaciones, por lo cual la alternativa sería preguntar a los miembros de la familia que trabajan aproximadamente cuál es su nivel mensual de ingresos y cuántos focos eléctricos tiene aproximadamente en su casa.¹⁴

Así, es recomendable *hacer solamente las preguntas necesarias* para obtener la información deseada o medir la variable.

Una modalidad de cuestionamientos múltiples lo es la batería de preguntas, la cual sirve para: *a)* ahorrar espacio en el cuestionario, *b)* facilitar la comprensión del mecanismo de respuesta (si se entiende la primera pregunta, se comprenderán las demás) (Corbetta, 2003) y *c)* construir índices que permitan obtener una calificación total.

Ejemplo



Variable a medir: visión departamental

Definición conceptual: percepción de la meta departamental en cuanto a claridad, naturaleza visionaria, grado en que es posible alcanzarla y medida en que puede ser compartida, y que representa una fuerza motivacional para el trabajo (Anderson y West, 1998 y Hernández-Sampieri, 2005).¹⁵

¹² Asociación Mexicana de Agencias de Investigación de Mercados y Opinión Pública, AMAI (2013), López (2009) y Universidad de Celaya (2013). Con el fin de conocer el método por puntos para la ubicación del nivel socioeconómico, se recomienda consultar directamente estas fuentes o consultar a la asociación de empresas de investigación de mercados de su país.

¹³ Desde luego, estos índices evolucionan permanentemente, en el corto plazo se contempla agregar indicadores respecto al ocio y tiempo libre, salud y esperanza de vida, idiomas, información y conocimiento y balance de vida (López, 2011).

¹⁴ En varios estudios se ha demostrado que el nivel de ingresos está relacionado con el número de focos de una casa habitación (residencia, hogar o propiedad). El número de focos se vincula con el número de cuartos, extensión de la casa, tamaño del jardín y otros factores, es decir, con el valor de la propiedad (Universidad de Celaya, 2013). Los rangos podrían ser: 3 focos o menos: estratos muy desfavorecidos; de 4 a 5 focos: estratos desfavorecidos; de 6 a 10 focos: estratos medios característicos; 11 a 15: estratos medios favorecidos; 16 a 20: medios/altos; 21 a 30: altos favorecidos; más de 31: muy altos o completamente favorecidos. Se sugiere excluir candiles. En cada nación cambia la designación de cada estrato, y no se quiere utilizar los términos “bajos”, que son peyorativos.

¹⁵ El ejemplo contiene solamente algunas de las preguntas de la escala original. Asimismo, se conjuntaron con otras escalas y se distribuyeron a lo largo del cuestionario.

Preguntas o ítems:	Completamente (mucho)	Aceptablemente	Regular	Poco	Nada
1. ¿Qué tan claros tiene los objetivos de su departamento?	5	4	3	2	1
2. ¿En qué medida considera usted que los objetivos de su departamento son útiles y apropiados?	5	4	3	2	1
3. ¿Qué tan de acuerdo está usted con estos objetivos de su departamento?	5	4	3	2	1
4. ¿En qué medida piensa usted que los objetivos de su departamento son claros?	5	4	3	2	1
5. ¿En qué medida piensa usted que los objetivos de su departamento son comprendidos por sus compañeros de trabajo del mismo departamento?	5	4	3	2	1
6. ¿En qué medida considera usted que sus compañeros de departamento están de acuerdo con los objetivos?	5	4	3	2	1
7. En qué medida considera que los objetivos del departamento pueden lograrse actualmente?	5	4	3	2	1

¿Las preguntas van precodificadas o no?

Siempre que se pretenda efectuar análisis estadístico, se requiere codificar las respuestas de los participantes a las preguntas del cuestionario, y debemos recordar que esto significa asignarles símbolos o valores numéricos y que cuando se tienen preguntas cerradas es posible codificar *a priori* o *precodificar* las opciones de respuesta, e incluir esta precodificación en el cuestionario (como en el último ejemplo).

Ejemplo

De preguntas precodificadas

¿Tiene usted inversiones en la Bolsa de Valores?

☐ 1 Sí ☐ 2 No

Cuando usted se enfrenta a un problema en su trabajo, para resolverlo recurre generalmente a:

1. Su superior inmediato
2. Su propia experiencia
3. Sus compañeros
4. Los manuales de políticas y procedimientos
5. Otra fuente: _____
(especificar)

En ambas preguntas, las respuestas van acompañadas de su valor numérico correspondiente, es decir, se han precodificado. Obviamente en las preguntas abiertas la codificación se realiza después, una vez que se tienen las respuestas. Las preguntas y opciones de respuesta precodificadas poseen la ventaja de que su codificación y preparación para el análisis son más sencillas y requieren menos tiempo.

¿Qué preguntas son obligatorias?

Regularmente, las preguntas llamadas demográficas o de ubicación del participante encuestado: género, edad, nivel socioeconómico, estado civil, escolaridad (nivel de estudios), religión, afiliación política, colonia, barrio o zona donde vive, pertenencia a ciertas agrupaciones, ocupación (actividad a la que se dedica), años de vivir en el lugar actual de residencia, etc. En empresas: puesto, antigüedad,



área funcional donde trabaja (gerencia, departamento, dirección o equivalente), planta u oficinas donde labora y demás preguntas. En cada investigación debemos analizar cuáles son pertinentes y nos resultarán útiles.

¿Qué características debe tener una pregunta?

Independientemente de que las preguntas sean abiertas o cerradas, y de que sus respuestas estén pre-codificadas o no, hay una serie de características que deben cubrirse al plantearlas:

- a) Las preguntas tienen que ser claras, precisas y comprensibles para los sujetos encuestados. Deben evitarse términos confusos, ambiguos y de doble sentido. Por ejemplo, la pregunta: “¿ve usted televisión?”, es confusa, no delimita cada cuánto. Sería mucho mejor especificar: ¿acostumbra usted ver televisión diariamente?, ¿cuántos días durante la última semana vio televisión?, y después preguntar horarios, canales y contenidos de los programas. Otro ejemplo inconveniente sería: ¿le gusta el deporte? No se sabe si se trata de verlo por televisión o en vivo, si de practicarlo o qué, y en última instancia, ¿cuál deporte? Otro caso que genera confusión son los términos con múltiples significados (Burnett, 2009), por ejemplo: ¿su empleo es estable?, implica un concepto de estabilidad de empleo que no tiene un solo significado. ¿Qué se considera estable?: ¿un contrato por un año, por dos, por cinco...?

Un caso común de confusión son las palabras sobre la temporalidad, resulta nebuloso el cuestionamiento: ¿ha asistido recientemente al cine?, ya que implica otras preguntas: ¿qué significa recientemente?, ¿ayer, la última semana, el último mes? Sería mejor interrogar: durante las últimas dos semanas (o mes), ¿cuántas veces ha ido al cine? De igual forma: ¿ha trabajado desde joven?, habrá de sustituirse por: ¿a partir de qué edad comenzó a trabajar?

- b) Es aconsejable que las preguntas sean lo más breves posible, porque las preguntas largas suelen resultar tediosas, toman más tiempo y pueden distraer al participante; pero como menciona Rojas (2002) no es recomendable sacrificar claridad por concisión. Cuando se trata de asuntos complicados tal vez es mejor una pregunta más larga, debido a que facilita el recuerdo, proporciona al sujeto más tiempo para pensar y favorece una respuesta más articulada (Corbetta, 2003). La directriz es que se incluyan las palabras necesarias para que se comprenda la pregunta, sin ser repetitivos o barrocos.
- c) Deben formularse con un vocabulario simple, directo y familiar para los participantes.
- d) No pueden incomodar a la persona encuestada ni ser percibidas como amenazantes y nunca ésta debe sentir que se le enjuicia. Debemos inquirir de manera sutil. Preguntas como: ¿acostumbra consumir algún tipo de bebida alcohólica?, tienden a provocar rechazo. Es mejor cuestionar: ¿algunos de sus amigos acostumbran consumir cierto tipo de bebida alcohólica?, y después utilizar preguntas tenues que indirectamente nos indiquen si la persona acostumbra consumir esta clase de bebidas (¿cuál es su tipo de bebida favorita?, ¿cada cuánto se reúne con sus amigos?, etc.). Mertens (2010) sugiere sustituir la pregunta: ¿es usted alcohólico? (en extremo amenazante), por la siguiente formulación: “el consumo de bebidas como el ron, tequila, vodka y whisky en esta ciudad es de X botellas de un litro, ¿en qué medida usted estaría por encima o por debajo de esta cantidad?” (alternativas de respuesta: “por encima, igual o por debajo”). Gochros (2005) recomienda cambiar la pregunta: ¿consume drogas?, por: ¿qué opina de las personas que consumen drogas en dosis mínimas? En estos casos de preguntas difíciles es posible usar escalas de actitud en lugar de preguntas o aun otras formas de medición (como se verá en la parte de escalas actitudinales y en otros instrumentos). Hay temáticas en las que a pesar de que se utilicen preguntas sutiles, el encuestado se puede sentir molesto. Tal es el caso del desempleo, la homosexualidad, el sida, la prostitución, la pornografía, los anticonceptivos y las adicciones.
- e) Las preguntas deben referirse preferentemente a un solo aspecto o una relación lógica. Por ejemplo, la pregunta: ¿acostumbra usted ver televisión y escuchar radio diariamente?, expresa dos aspectos y llega a confundir. Es necesario dividirla en dos preguntas, una relacionada con la televisión y otra vinculada con la radio. Otro ejemplo: ¿sus padres eran saludables?, es una pregunta problemática, además del concepto “saludable” (confuso), es imposible de responder en el caso

de que la madre nunca se hubiera enfermado de gravedad y nunca hubiera sido hospitalizada y, en cambio, el padre hubiera padecido severos problemas de salud.

- f) Las preguntas no habrán de inducir las respuestas. Se tienen que evitar preguntas tendenciosas o que dan pie a elegir un tipo de respuesta (directivas). Por ejemplo, ¿considera a nuestro compañero Ricardo Hernández como el mejor candidato para dirigir nuestro sindicato?, es una pregunta tendenciosa, pues induce la respuesta. Lo mismo que la pregunta: ¿los trabajadores argentinos son muy productivos? Se insinúa la respuesta en la pregunta. Resultaría mucho más conveniente interrogar: ¿qué tan productivos considera usted, en general, a los trabajadores argentinos? (y mostrar alternativas).

Ejemplo



¿Qué tan productivos considera usted, en general, a los trabajadores argentinos?

Sumamente productivos	Más bien productivos	Más bien improductivos	Sumamente improductivos
--------------------------	-------------------------	---------------------------	----------------------------

Otro ejemplo inconveniente sería: ¿piensa usted votar por tal partido político en las próximas elecciones? El participante nunca debe sentirse presionado. Un factor importante a considerar es la deseabilidad social (Furr, 2010), a veces las personas utilizan respuestas culturalmente aceptables. Por ejemplo, la pregunta: ¿le gustaría casarse?, podría inducir y forzar a más de una persona a responder de acuerdo con las normas de su comunidad. Resulta mejor cuestionar: ¿qué opina del matrimonio?, y más adelante inquirir sobre sus anhelos y expectativas al respecto. Una interrogante como: ¿acostumbra leer el periódico?, puede llevarnos a respuestas socialmente válidas: “sí, lo leo a diario, yo leo mucho” (cuando no es cierto). Es mejor preguntar: ¿suele tener tiempo para leer el periódico?, ¿con qué frecuencia?

- g) Las preguntas no pueden apoyarse en instituciones, ideas respaldadas socialmente ni en evidencia comprobada. Es también una manera de inducir respuestas. Por ejemplo, la pregunta: la Organización Mundial de la Salud ha realizado diversos estudios y concluyó que el tabaquismo provoca diversos daños al organismo, ¿considera usted que fumar es nocivo para su salud? Esquemas del tipo: “la mayoría de las personas opinan que...”, “la Iglesia considera...”, “los padres de familia piensan que...”, etc., no deben anteceder a las preguntas, ya que influyen y sesgan las respuestas.
- h) Es aconsejable evitar preguntas que nieguen el asunto que se interroga. Por ejemplo: ¿qué niveles de la estructura organizacional *no* apoyan el proceso de calidad? Es mejor preguntar sobre qué niveles *sí* apoyan el proceso. O bien: ¿qué *no* le agrada de este centro comercial?, es preferible cuestionar: ¿qué le desagrada de este centro comercial? Tampoco es conveniente incluir dobles negaciones (son positivas, pero suelen confundir): ¿considera que la mayoría de las mujeres casadas preferiría *no* trabajar si *no* tuviera presión económica? Mejor se redacta de manera positiva.
- i) No deben hacerse preguntas racistas o sexistas ni que ofendan a los participantes. Es obvio, pero no está de más recalcarlo. Se recomienda también sortear las preguntas *con fuerte carga emocional o muy complejas*, que más bien son preguntas para entrevistas cualitativas (por ejemplo: ¿cómo era la relación con su ex marido? —aunque una escala completa puede ser la solución— o ¿qué siente usted sobre la muerte de su hijo?).
- j) En las preguntas con varias categorías de respuesta, y cuando el entrevistado sólo tiene que elegir una, llega a ocurrir que el orden en que se presentan dichas opciones afecta las respuestas de los participantes (por ejemplo, que tiendan a favorecer a la primera o a la última opción de respuesta). Entonces resulta conveniente rotar el orden de lectura de las respuestas a elegir de manera proporcional. Por ejemplo, si preguntamos: ¿cuál de los siguientes cuatro candidatos presidenciales considera usted que logrará disminuir verdaderamente la inflación? Veinticinco por ciento de las veces (o una de cada cuatro ocasiones) que se haga la pregunta se menciona primero al candidato A, 25% se menciona primero al candidato B, 25% al candidato C y el restante 25% al

candidato *D*. Asimismo, cuando las alternativas son demasiadas es más difícil responder, por ello es conveniente limitarlas a las mínimas necesarias.

En la tabla 9.4 se relacionan los problemas que surgen al generar preguntas, adaptada de Creswell (2005).

● **Tabla 9.4** Ejemplos de algunos problemas al elaborar preguntas

Problema	Ejemplo de pregunta problemática	Mejora a la pregunta
Pregunta confusa por la vaguedad de los términos	¿Votará en las próximas elecciones?	Precisar términos: En las próximas elecciones del 10 de noviembre para elegir alcalde de Monterrey, ¿piensa ir a votar?
Dos o más conceptos o dos preguntas en una sola	¿Qué tan satisfecho está usted con el servicio de comedor y el servicio médico que se ofrece en la empresa?	Una pregunta por concepto: ¿Qué tan satisfecho está usted con el servicio de comedor que se ofrece en la empresa? ¿Qué tan satisfecho está usted con el servicio médico que se ofrece en la empresa?
Demasiadas palabras	Como usted sabe, el próximo 10 de noviembre se celebrarán elecciones locales en este municipio de Cortazar para elegir alcalde, en esa fecha: ¿piensa usted acudir a las urnas a emitir su voto por el candidato que considera será el mejor alcalde para el municipio?	Reducir términos: En las próximas elecciones del 10 de noviembre para elegir alcalde de Cortazar, ¿piensa ir a votar?
Pregunta negativa	¿Los estudiantes no deben portar o llevar armas a la escuela?	Cambiarla a neutral: ¿Los estudiantes deben o no portar armas en la escuela?
Contiene “jerga lingüística”	¿Qué tan “chida” o “padre” es la relación con su empresa?	Eliminar dicha jerga: ¿Qué tan orgulloso se encuentra usted de trabajar en esta empresa?
Se traslapan las categorías de respuesta	¿Podría indicarme su edad? __ 18 – 19 __ 19 – 20 __ 20 – 21 __ 21 – 22	Lograr que las categorías sean mutuamente excluyentes: ¿Podría indicarme su edad? __ 18 – 19 __ 20 – 21 __ 22 – 23
Categorías de respuesta sin balance entre las favorables y las desfavorables (positivas y negativas)	¿En qué medida está usted satisfecho con su superior inmediato? ☐ Insatisfecho ☐ Medianamente satisfecho ☐ Satisfecho ☐ Sumamente satisfecho	Proporcionar equilibrio entre opciones favorables y desfavorables: ¿En qué medida está usted satisfecho con su superior inmediato? ☐ Sumamente insatisfecho ☐ Más bien insatisfecho ☐ Ni insatisfecho ni satisfecho ☐ Más bien satisfecho ☐ Sumamente satisfecho
Incongruencia entre la pregunta y las opciones de respuesta	¿En qué medida está usted satisfecho con su superior inmediato? ☐ Muy poco importante ☐ Poco importante	Generar categorías que coincidan con la pregunta: ¿En qué medida está usted satisfecho con su superior inmediato? ☐ Sumamente insatisfecho

(continúa)

► **Tabla 9.4** (continuación)

Problema	Ejemplo de pregunta problemática	Mejora a la pregunta
	☐ Medianamente importante ☐ Importante ☐ Muy importante	☐ Más bien insatisfecho ☐ Ni insatisfecho ni satisfecho ☐ Más bien satisfecho ☐ Sumamente satisfecho
Sólo una parte de los participantes pueden entender la pregunta	¿Cuál es el género y marca de bebida etílica que acostumbra adquirir con un mayor índice de frecuencia en sus compras?	Simplificar términos: ¿Cuál es el tipo de bebida alcohólica y de qué marca acostumbra comprar con mayor frecuencia?
Utilización de términos en otro idioma	¿Qué efectos tuvo en esta empresa el <i>downsizing</i> ?	Traducir términos: ¿Qué efectos tuvo en esta empresa la reducción de empleados?
La pregunta puede ser inadecuada para parte de la población	¿Cómo le afectó el incremento en la tasa impositiva para empleados gubernamentales?	Agregar preguntas que segmenten a la población: ¿Actualmente trabaja? ☐ Sí ☐ No ¿Trabaja usted en... ☐ empresa? ☐ por cuenta propia (independiente)? ☐ gobierno? Entonces, a quienes pertenezcan a la última categoría, se les pregunta: ¿Cómo le afectó el incremento en la tasa impositiva para empleados gubernamentales?

En relación con cada pregunta del cuestionario, León y Montero (2003) sugieren que se reflexione sobre si es necesaria la pregunta. ¿Es lo suficientemente concreta? ¿Responderán los participantes sinceramente?

¿Cómo deben ser las primeras preguntas de un cuestionario?

En algunos casos es conveniente iniciar con preguntas neutrales o fáciles de contestar, para que el participante se adentre en la situación. No se recomienda comenzar con preguntas difíciles o muy directas. Imaginemos un cuestionario diseñado para obtener opiniones en torno al aborto que empiece con una pregunta poco sutil como: ¿está de acuerdo con que se legalice el aborto en este país? Sin lugar a dudas, sería un fracaso. Los primeros cuestionamientos deben resultar interesantes para los participantes. A veces incluso, pueden ser divertidos (por ejemplo, en la investigación de la moda y la mujer mexicana que se verá en la cuarta parte del libro sobre modelos mixtos, al comenzar a inquirir sobre los tipos de prendas que compraban las participantes, la primera pregunta fue: ¿sueles ponerte una pijama para dormir?, cuestionamiento que resultó sumamente divertido y provocó hilaridad, logrando relajar a las encuestadas. Desde luego, la pregunta la hicieron mujeres entrevistadoras jóvenes).

A veces los cuestionarios comienzan con las preguntas demográficas ya mencionadas, pero en otras ocasiones es mucho mejor hacer este tipo de preguntas al final del cuestionario, particularmente en casos donde los participantes puedan sentir que se comprometen si responden el cuestionario.

Cuando construimos un cuestionario, es indispensable que pensemos en cuáles son las preguntas ideales para iniciar. Éstas deberán lograr que el sujeto se concentre en el cuestionario. Al respecto, se



sugiere el procedimiento de “embudo” en la presentación de las preguntas: ir de las más generales a las más específicas. Una característica fundamental de un cuestionario es que las preguntas importantes nunca deben ir al final.

¿De qué está formado un cuestionario?

Además de las preguntas y categorías de respuestas, un cuestionario está formado básicamente por: portada, introducción, instrucciones insertas a lo largo del contenido y agradecimiento final.

Portada

Esta incluye la carátula; en general, debe ser atractiva gráficamente para favorecer las respuestas. Debe incluir el nombre del cuestionario y el logotipo de la institución que lo patrocina. En ocasiones se agrega un logotipo propio del cuestionario o un símbolo que lo identifique.

Introducción

Debe incluir:

- Propósito general del estudio.
- Motivaciones para el sujeto encuestado (importancia de su participación).
- Agradecimiento.
- Tiempo aproximado de respuesta (un promedio o rango).
- Espacio para que firme o indique su consentimiento (a veces se incluye al final o en ocasiones es innecesario).
- Identificación de quién o quiénes lo aplican.
- Explicar brevemente cómo se procesarán los cuestionarios y una cláusula de confidencialidad del manejo de la información individual.
- Instrucciones iniciales claras y sencillas (cómo responder en general, con ejemplos si se requiere).

Cuando el cuestionario se aplica mediante entrevista, la mayoría de tales elementos son explicados por el entrevistador. El cuestionario debe ser y parecer corto, fácil y atractivo.

A continuación, se presenta un ejemplo de carta introductoria y otro de instrucciones generales para responder al cuestionario.¹⁶

Ejemplo



Carta introductoria

Buenos días (tardes):

Estamos trabajando en un estudio que servirá para elaborar una tesis profesional acerca de la biblioteca de la Universidad de Celaya.

Quisiéramos pedir tu ayuda para que contestes algunas preguntas que no te tomarán mucho tiempo. Tus respuestas serán confidenciales y anónimas. No hay preguntas delicadas.

Las personas que fueron seleccionadas para el estudio no se eligieron por su nombre sino al azar.

Las opiniones de todos los encuestados serán sumadas e incluidas en la tesis profesional, pero nunca se comunicarán datos individuales.

Te pedimos que contestes este cuestionario con la mayor sinceridad posible. No hay respuestas correctas ni incorrectas.

Lee las instrucciones cuidadosamente, ya que existen preguntas en las que sólo se puede responder a una opción; otras son de varias opciones y también se incluyen preguntas abiertas.

¡Muchas gracias por tu colaboración!

¹⁶ El lenguaje es claramente el que se utiliza en México. Desde luego, cada quien debe adaptarlo a su propia cultura.

Ejemplo

Instrucciones de un cuestionario

ENCUESTA DEL CLIMA ORGANIZACIONAL

INSTRUCCIONES

Emplee un lápiz o un bolígrafo de tinta negra para responder el cuestionario. Al hacerlo, piense en lo que sucede la mayoría de las veces en su trabajo.

No hay respuestas correctas o incorrectas. Simplemente reflejan su opinión personal.

Todas las preguntas tienen cinco opciones de respuesta. Elija la que mejor describa lo que piensa usted. Solamente una opción.

Marque con claridad la opción elegida con una cruz o tache, o bien, una "paloma" (símbolo de verificación). Recuerde: NO se deben marcar dos opciones. Marque así:



Si no puede contestar una pregunta o si la pregunta no tiene sentido para usted, por favor pregúntele a la persona que le entregó este cuestionario y le explicó la importancia de su participación.

CONFIDENCIALIDAD

Sus respuestas serán anónimas y absolutamente confidenciales. Los cuestionarios serán procesados por personas externas. Además, como usted puede ver, en ningún momento se le pide su nombre.

De antemano: ¡muchas gracias por su colaboración!

Algunas apelaciones que podemos utilizar en la introducción se muestran en la tabla 9.5.¹⁷

► **Tabla 9.5** Ejemplos de apelaciones para incentivar la participación

Apelación	Ejemplo
Incentivo	"Al responder, usted recibirá..." (dinero, un obsequio, un boleto, etcétera).
Altruismo	"Los resultados servirán para resolver...", "El estudio ayudará a..." (problema social, mejora en la calidad de vida, solventar una necesidad comunitaria, etcétera).
Autoconcepto de la persona	"Usted es una de las pocas personas que puede señalar ciertas cuestiones...". "Debido a su experiencia (pericia, importancia, conocimientos, etc.) usted puede... y por ello le solicitamos..." (su opinión calificada, etcétera).
Interés por el conocimiento	"Le enviaremos una copia de los resultados..."
Intereses profesionales	"Los resultados serán útiles para conocer temas importantes en nuestra profesión..."
Ayuda-auxilio	"Necesitamos su apoyo para conocer...", "Los jóvenes requieren de ayuda para..."
Autoridad	Introducción acompañada de la firma de un líder o persona reconocida. O bien: "Doña Pola Castelán nos ha pedido que hagamos esta encuesta para conocer el problema de los niños...", "El científico...", "El empresario..."
Agradecimiento	"La comunidad de... estará muy agradecida por..."

También se insertan instrucciones a lo largo del cuestionario (normalmente con otra fuente tipográfica o bien en cursivas, para distinguirlas de las preguntas y respuestas), las cuales nos indican cómo contestar. Por ejemplo:

—¿Tiene este ejido o esta comunidad ganado, aves o colmenas que sean de propiedad colectiva? (Marque con una cruz su respuesta).

☐ Sí
(continúe)

☐ No
(pase a la pregunta 30)

—¿Se ha obtenido la cooperación de todo el personal o de la mayoría de éste para el proyecto de mejora de la calidad?

☒ Sí
(pase a la pregunta 26)

☒ No
(pase a la pregunta 27)

¹⁷ Basada en Mertens (2010).

Recordemos que en ocasiones se presentan tarjetas con las opciones de respuestas y se instruye al entrevistador para que las muestre a los participantes. Por ejemplo:

–Hablando de la mayoría de sus proveedores, en qué medida conoce usted... (*mostrar la tarjeta uno y marcar la respuesta en cada caso*).

Ejemplo

De pregunta con tarjeta de respuestas

	Completamente (5)	Bastante (4)	Regular (3)	Poco (2)	Nada (1)
• ¿Sus políticas de venta?					
• ¿Sus finanzas (estados financieros)?					
• ¿Los objetivos de su área de ventas?					
• ¿Sus programas de capacitación para vendedores?					
• ¿Número de empleados de su área de ventas?					
• ¿Problemas laborales?					
• ¿Sus métodos de producción?					
• ¿Otros clientes de ellos?					
• ¿Su índice de rotación de personal?					

Las instrucciones son tan importantes como las preguntas y es necesario que sean claras para los usuarios a quienes van dirigidas.

Agradecimiento final

Aunque haya agradecido de antemano, vuelva a agradecer la participación.

Formato, distribución de instrucciones, preguntas y categorías

Las preguntas deben estar organizadas para que sea más fácil de responder el cuestionario. Es importante asegurarnos de numerar páginas y preguntas.

La manera en que pueden distribuirse preguntas, categorías de respuesta e instrucciones es variada. Algunos prefieren colocar las preguntas a la izquierda y las respuestas a la derecha, con lo que se tendría un formato como el siguiente:

Ejemplo

Modelo de formato de distribución de preguntas

¿Considera a su jefe o superior inmediato como su amigo?	☐ Definitivamente sí	☐ Sí	☐ No	☐ Definitivamente no
¿Cuando tiene problemas se siente apoyado por su jefe o superior inmediato?	☐ Definitivamente sí	☐ Sí	☐ No	☐ Definitivamente no
¿Considera que su jefe o superior inmediato le orienta adecuadamente en su trabajo?	☐ Definitivamente sí	☐ Sí	☐ No	☐ Definitivamente no
¿Tiene una buena impresión de su jefe o superior inmediato?	☐ Definitivamente sí	☐ Sí	☐ No	☐ Definitivamente no

Otros dividen el cuestionario por secciones de preguntas y utilizan un formato horizontal.

Ejemplo



Modelo de formato horizontal

Presentación

PREGUNTAS SOBRE EL SUPERIOR INMEDIATO

¿Considera a su jefe o superior inmediato como su amigo?

☐ Definitivamente sí ☐ Sí ☐ No ☐ Definitivamente no

¿Cuando tiene problemas se siente apoyado por su jefe o superior inmediato?

☐ Definitivamente sí ☐ Sí ☐ No ☐ Definitivamente no

¿Considera que su jefe o superior inmediato le orienta adecuadamente en su trabajo?

☐ Definitivamente sí ☐ Sí ☐ No ☐ Definitivamente no

¿Tiene una buena impresión de su jefe o superior inmediato?

☐ Definitivamente sí ☐ Sí ☐ No ☐ Definitivamente no

PREGUNTAS SOBRE MOTIVACIÓN

Otros combinan diversas posibilidades, distribuyendo preguntas que miden la misma variable a través de todo el cuestionario. Cada quien es capaz de utilizar el formato que desee o juzgue conveniente, lo importante es que en su totalidad sea comprensible para el usuario: que las instrucciones, preguntas y respuestas se diferencien; que el formato no resulte visualmente tedioso y se lea sin dificultad.

Hoy en día es común elaborar cuestionarios electrónicos para dispositivos portátiles de diferentes tipos como laptops, tabletas, teléfonos móviles, etc., así como para páginas web y blogs que contienen fotografías, dibujos, secuencias de video y música. Son sumamente atractivos y existen decenas de programas para diseñarlos.

¿De qué tamaño debe ser un cuestionario?

No existe una regla al respecto, pero si el cuestionario es muy corto se pierde información y si resulta largo llega a ser tedioso. En este último caso, las personas se negarían a responder o, al menos, lo contestarían en forma incompleta. La abuela doña Margarita Castelán Sampieri repetía el refrán: “lo bueno y breve, doblemente bueno”. El tamaño depende del número de variables y dimensiones a medir, el interés de los participantes y la manera como se administre (de este punto se hablará en el siguiente apartado). Cuestionarios que toman más de 35 minutos suelen resultar fatigosos, a menos que los sujetos estén muy motivados para contestar (por ejemplo, cuestionarios de personalidad o cuestionarios para obtener un trabajo). Una recomendación que ayuda a evitar un cuestionario más largo de lo requerido es: *no hacer preguntas innecesarias o injustificadas*.

¿Cómo se codifican las preguntas abiertas?

Las preguntas abiertas se codifican una vez que conocemos todas las respuestas de los participantes a los cuales se les aplicaron, o al menos las principales tendencias de respuestas en una muestra de los cuestionarios aplicados. Es importante anotar que esta actividad es similar a “cerrar” una pregunta abierta por medio de la prueba piloto, pero el producto es diferente. En este caso, con la codificación de preguntas abiertas se obtienen ciertas categorías que representan los resultados finales.

El procedimiento consiste en encontrar y dar nombre a los patrones generales de respuesta (respuestas similares o comunes), listar estos patrones y después asignar un valor numérico o un símbolo a cada patrón. Así, un patrón constituirá una categoría de respuesta. Para cerrar las preguntas abiertas se sugiere el siguiente procedimiento:

1. Seleccionar determinado número de cuestionarios mediante un método adecuado de muestreo, que asegure la representatividad de los participantes investigados.
2. Observar la frecuencia con que aparece cada respuesta en las preguntas.
3. Elegir las respuestas que se presentan con mayor frecuencia (patrones generales de respuesta).
4. Clasificar las respuestas elegidas en temas, aspectos o rubros, de acuerdo con un criterio lógico, cuidando que sean mutuamente excluyentes.
5. Darle un nombre o título a cada tema, aspecto o rubro (patrón general de respuesta).
6. Asignarle el código a cada patrón general de respuesta.

Por ejemplo, en una investigación sobre mejora de la calidad se hizo una pregunta abierta: ¿de qué manera la alta gerencia busca obtener la cooperación del personal para el desarrollo del proyecto de calidad?

Las respuestas fueron múltiples, pero se encontraron los patrones generales de respuesta que se muestran en el ejemplo.

Ejemplo

De codificación de preguntas abiertas

Códigos	Categorías (patrones o respuestas con mayor frecuencia de mención)	Frecuencia de mención
1	Involucrar al personal y comunicarse con él	28
2	Motivación e integración	20
3	Capacitación en general	12
4	Incentivos/recompensas	11
5	Difundir el valor “calidad” o la filosofía de la empresa	7
6	Grupos o sesiones de trabajo	5
7	Posicionamiento del área de calidad o equivalente	3
8	Sensibilización en grupo	2
9	Desarrollo de la calidad de vida en el trabajo	2
10	Incluir aspectos de calidad en el manual de inducción	2
11	Poner énfasis en el cuidado de la maquinaria	2
12	Trabajar bajo un buen clima laboral	2
13	Capacitación “en cascada”	2
14	Otras	24

Como varias categorías o diversos patrones tenían únicamente dos frecuencias, se redujeron a sólo seis, tal como se ejemplifica a continuación.

Ejemplo

De reducción o agrupamiento de categorías

Códigos	Categorías (frecuencias)
1	Involucrar al personal y comunicarse con él (28)
2	Motivación e integración/mejoramiento del clima laboral (22)
3	Capacitación (14)
4	Incentivos/recompensas (11)
5	Difundir el valor “calidad” o la filosofía de la empresa (7)
6	Grupos o sesiones de trabajo (7)
7	Otras (33)

Al “cerrar” preguntas abiertas y codificarlas, debe tenerse en cuenta que un mismo patrón de respuesta puede expresarse con diferentes palabras. Por ejemplo, ante la pregunta: ¿qué sugerencias haría para mejorar al programa *Estelar*? Las respuestas: mejorar las canciones y la música, cambiar las canciones, incluir nuevas y mejores canciones, etc., se agruparían en la categoría o el patrón de respuesta *modificar la musicalización del programa*.

¿En qué contextos puede administrarse o aplicarse un cuestionario?

Los cuestionarios se aplican de dos maneras fundamentales: autoadministrado y por entrevista (personal o telefónica). En este apartado se comentan ambas y en el capítulo siete adicional “Recolección de los datos cuantitativos: segunda parte” (en el centro de recursos en línea: Material complementario → Capítulos), en la sección “Contextos de administración de los cuestionarios” se amplían algunos conceptos.



1. Autoadministrado

Autoadministrado significa que el cuestionario se proporciona directamente a los participantes, quienes lo contestan. No hay intermediarios y las respuestas las marcan ellos. Pero la forma de autoadministración puede tener distintos contextos: individual, grupal o por envío (correo tradicional, correo electrónico y página web o equivalente).

En el caso individual, el cuestionario se entrega al participante y éste lo responde, ya sea que acuda a un lugar para hacerlo (como ocurre cuando se llena un formulario para solicitar empleo) o lo conteste en su lugar de trabajo, hogar o estudio. Por ejemplo, si los participantes fueran una muestra de directivos de laboratorios farmacéuticos de Bogotá, se acudiría a sus oficinas y se les entregarían los cuestionarios. Los ejecutivos se administrarían ellos mismos el cuestionario y esperaríamos a que lo respondieran o lo recogeríamos otro día. El reto de esta última situación es lograr que los participantes devuelvan el cuestionario contestado completamente. Resulta conveniente que quien lo entregue posea habilidades para relacionarse con las personas, sea asertivo, y además se caracterice por una elevada persistencia. En nuestra experiencia, en distintos países de Iberoamérica, jóvenes de ambos géneros con buena capacidad comunicativa logran porcentajes de recuperación por encima de 90% en tiempos aceptables (una semana o menos). Y no es necesario que sean físicamente atractivos (aunque ayuda), más bien el éxito reside en su motivación y tenacidad. Asimismo, el mayor costo o gasto de esta clase de administración de los cuestionarios lo representa su distribución y recolección.

En el segundo caso, se reúne a los participantes en grupos (a veces pequeños —cuatro a seis personas—, otras en grupos intermedios —entre siete y 20 sujetos—, incluso en grupos grandes de 21 a 40 individuos). Por ejemplo, empleados (en encuestas de clima organizacional es muy común juntar a grupos de 25, entregarles el cuestionario, introducir al propósito del estudio y al instrumento, responder dudas y pedirles que al concluir lo depositen en una urna sellada, para mantener la confidencialidad), padres de familia (en reuniones escolares), televidentes (cuando asisten a un foro televisivo), alumnos (en sus salones de clase), etc. Es tal vez la forma más económica de aplicar un cuestionario.

Cuando se envía a los participantes para autoadministración, puede usarse correo postal público o mensajería privada, o bien, correo electrónico. También se les puede pedir que ingresen a una página web o equivalente para responderlo.

Los cuestionarios para autoadministración deben ser particularmente atractivos (a colores, en papel especial, con diseño original, etc.; si el presupuesto lo permite).

2. Por entrevista personal

Las *entrevistas* implican que una persona calificada (entrevistador) aplica el cuestionario a los participantes; el primero hace las preguntas a cada entrevistado y anota las respuestas. Su papel es crucial, resulta una especie de filtro.

El primer contexto que se revisará de una entrevista es el personal (“cara a cara”).

Normalmente se tienen varios entrevistadores, quienes deberán estar capacitados en el arte de entrevistar y conocer a fondo el cuestionario. No deberán sesgar o influir en las respuestas, por ejem-

plo, reservarse de expresar aprobación o desaprobación respecto de las respuestas del entrevistado, reaccionar de manera ecuánime cuando los participantes se alteren, contestar con gestos ambiguos cuando los sujetos busquen generar una reacción en ellos, etc. Su propósito es llevar a buen fin cada entrevista, evitando que decaiga la concentración e interés del participante, además de orientarlo en el tránsito del instrumento. Las explicaciones que proporcione deberán ser breves pero suficientes. Tiene que ser neutral, pero cordial y servicial. Asimismo, es muy importante que transmita a todos los participantes que no hay respuestas correctas o equivocadas. Por otra parte, su proceder debe ser lo más homogéneo posible (mismos señalamientos, presentación uniforme, etc.). Con respecto a las instrucciones del cuestionario, algunas son para el entrevistado y otras para el entrevistador. Este último debe recordar que al inicio se comenta: el propósito general del estudio, las motivaciones y el tiempo aproximado de respuesta, agradeciendo de antemano la colaboración.

El cuestionario por entrevista es el que consigue un mayor porcentaje de respuestas a las preguntas, su estimación es de 80 a 85% (León y Montero, 2003).

En relación con el perfil de los entrevistadores no hay un consenso; por ejemplo, Corbetta (2003) sugiere que sean mujeres casadas, amas de casa, de mediana edad, diplomadas y de clase media. León y Montero (2003) recomiendan que sean siempre profesionales. En nuestra experiencia, el tipo de entrevistador depende del tipo de persona entrevistada. Por ejemplo, que pertenezca a un nivel socioeconómico similar a la mayoría de la muestra, sea joven y haya cursado asignaturas de investigación, que posea facilidad de palabra y capacidad de socializar. Se reitera que los estudiantes de ambos géneros funcionan muy bien. Por tanto, es claro que para este fin deben evitarse personas inseguras o excesivamente tímidas.

La entrevista cuantitativa es diferente de la cualitativa.¹⁸ En relación a la primera, Moule y Goodman (2009) y Rogers y Bouey (2005), mencionan las siguientes características:

- a) El principio y final de la entrevista se definen con claridad. De hecho, tal definición se integra en el cuestionario.
- b) El mismo instrumento se aplica a todos los participantes, en condiciones lo más similares posible (estandarización).
- c) Se busca que sea individual, sin la intrusión de otras personas que pueden opinar o alterar de alguna manera la entrevista.
- d) Es poco a nada anecdótica (aunque en algunos casos es recomendable que el entrevistador anote cuestiones fuera de lo común como ciertas reacciones y negativas a responder).
- e) La mayoría de las preguntas suelen ser cerradas, con mínimos elementos rebatibles, ampliaciones y sondeos.
- f) El entrevistador y el propio cuestionario controlan el ritmo y la dirección de la entrevista.
- g) El contexto social no es un elemento a considerar, lo es solamente el ambiental.
- h) El entrevistador procura que su patrón de comunicación sea similar (su lenguaje, instrucciones, etcétera).

Asimismo, la capacitación de entrevistadores debe incluir cuestiones de comunicación no verbal básicas (control de gestos, manejo de silencios, etc.), además de todos los puntos que se revisaron anteriormente.

Cabe señalar que, cuando se trata de *entrevista personal*, las condiciones del lugar donde se realice son importantes (oficina, hogar o casa-habitación, sitio público, como centro comercial, parque, escuela, etc.). Pero, en cualquier caso, se aconseja que se busque un lugar lo más discreto, silencioso y privado que sea posible.

En estas entrevistas es común mostrar visualmente las opciones de respuesta a los entrevistados, mediante tarjetas, en especial cuando se incluyen más de cinco o son complejas. Pongamos de ejemplo la siguiente tarjeta.¹⁹

¹⁸ Las entrevistas cualitativas se revisan en el capítulo 14 “Recolección y análisis de los datos cualitativos”.

¹⁹ El ejemplo se ha simplificado por cuestiones de espacio, las opciones fueron obtenidas después de una prueba piloto.

Ejemplo

Tarjeta para mostrar al entrevistado cuando hay diversas opciones de respuesta

¿Cuáles considera usted que son los tres principales problemas en este municipio?

- | | | |
|--|---|--|
| ☐ Pandillerismo | ☐ Desempleo | ☐ Inseguridad en las calles |
| ☐ Venta de drogas, narcomenudeo | ☐ Falta de vivienda | ☐ Problemas en la recolección de basura |
| ☐ Pobreza | ☐ Falta de infraestructura (calles, puentes, etcétera) | ☐ Escasez de agua |
| ☐ Corrupción de funcionarios de la alcaldía | ☐ Empleo mal remunerado | ☐ Carencia de servicios de salud |

Hoy en día hay varios sistemas para sustituir al cuestionario de lápiz y papel, como la CAPI (*Computer-Assisted Personal Interviewing*), en la que el entrevistador muestra al participante una computadora portátil (*notebook* o *tableta*) o un teléfono móvil que contiene el cuestionario y este último responde guiado por el primero o directamente aquel anota las respuestas. Se usa teclado, sistema al tacto o pluma electrónica.

Casi siempre las **entrevistas** son individuales, aunque podrían aplicarse a un grupo pequeño (si ésta fuera la unidad de análisis). Es decir, el cuestionario lo responden entre todos sus miembros o parte de ellos (por ejemplo, cuestionarios para parejas o una familia, o un área de una empresa).

Los cuestionarios administrados usando dispositivos tienen la ventaja de que los datos se capturan y agregan a la base de datos de forma automática, de manera que en cualquier momento podemos hacer un corte y efectuar toda clase de análisis (ver tendencias, evaluar funcionamiento del instrumento, etc.). La desventaja es obvia: el costo, el cual es muy difícil de absorber por parte de un estudiante o maestro e incluso una institución.

Idealmente, después de una entrevista se puede preparar un informe que indique: si el participante se mostraba sincero, la manera como respondió, el tiempo que duró la entrevista, el lugar donde se realizó, las características del entrevistado, los contratiempos que se presentaron y la forma en que se desarrolló la entrevista, así como otros aspectos relevantes.

En entrevistas Si la pregunta se va a presentar oralmente, no debe contener más de cinco opciones de respuesta, ya que por encima de este límite se suelen olvidar las primeras.

3. Por entrevista telefónica

Obviamente, la diferencia con el anterior tipo de entrevista es el medio de comunicación, que en este caso es el teléfono (hogar, oficina, móvil o celular). Las entrevistas telefónicas son la forma más rápida y económica de realizar una encuesta.

Las habilidades requeridas de parte de los entrevistadores son parecidas a las de la entrevista personal, excepto que éstos no tienen que confrontarse “cara a cara” con los participantes (no importa la vestimenta ni el aspecto físico, pero sí la voz; su modulación y claridad son fundamentales). El nivel de rechazo varía de acuerdo con el entorno social y tiempo en que se realice (“saturación telefónica” de campañas comerciales y políticas de telemarketing, inseguridad pública, etcétera).

Una ventaja de este método reside en que se puede acceder a barrios inseguros y a hogares exclusivos donde se limita el ingreso, así como a lugares geográficamente lejanos.

La desventaja de las encuestas telefónicas es que están limitadas a unas cuantas preguntas o no se pueden efectuar mediciones complejas de variables o profundizar en ciertos temas. Pero una vez más, los datos se capturan y agregan a la base de datos de forma automática y se pueden hacer cortes de la información de manera inmediata y realizar toda clase de análisis.

Algunas recomendaciones para las entrevistas telefónicas y el tema de sistemas automatizados para realizarlas como el CATI (*Computer-Assisted Telephone Interviewing*), se revisan en el capítulo 7 adicional, “Recolección de los datos cuantitativos: segunda parte” (que se puede descargar del centro de recursos en línea).

Cuando se realizan entrevistas personales en el hogar o telefónicas se debe tomar en cuenta el *horario*. Ya que si efectuamos la visita o hablamos por teléfono sólo a una hora (digamos en la mañana), nos encontraremos con unos cuantos subgrupos de la población (por ejemplo, amas de casa).

Una variación de la administración de cuestionarios por teléfono es la siguiente: en un programa radial, televisivo o vía internet se solicita la opinión o respuesta de los receptores a una pregunta o algunas cuantas preguntas, éstos deben marcar un número telefónico y contestar las opciones de respuesta con las que concuerden más. El problema de estas encuestas reside en la muestra, que desde luego no es probabilística, sino que se conforma de voluntarios que cubren dos condiciones: tener teléfono y estar viendo o escuchando la emisión del programa. Este proceder nos conduce más que a un estudio, a un sondeo, lo cual no es un error en sí, lo grave es que se pretenda generalizar los resultados a una población.

Algunas consideraciones adicionales para la administración del cuestionario

Cuando se tiene población analfabeta, con niveles educativos bajos o niños que apenas comienzan a leer o no dominan la lectura, el método más conveniente de administración de un cuestionario es por entrevista. Aunque existen algunos cuestionarios muy gráficos que usan escalas sencillas para las opciones de respuestas. Como en el siguiente ejemplo.

Ejemplo



En desacuerdo



Neutral



De acuerdo

Con trabajadores de niveles de lectura básica se recomienda utilizar entrevistas o cuestionarios autoadministrados sencillos que se apliquen en grupos, proporcionándoles asesoría.

En algunos casos, con ejecutivos que difícilmente vayan a dedicarle a un solo asunto más de 20 minutos, se pueden utilizar cuestionarios autoadministrados o entrevistas telefónicas. Con estudiantes suelen funcionar los cuestionarios autoadministrados.

Algunas asociaciones realizan encuestas por correo y ciertas empresas envían cuestionarios a sus empleados mediante el servicio interno de mensajería o por correo electrónico. Cuando el cuestionario contiene unas cuantas preguntas (su administración toma entre cuatro y cinco minutos), la entrevista telefónica es una buena alternativa.

Ahora bien, sea cual fuere la forma de administración, siempre debe haber uno o varios supervisores que verifiquen que los cuestionarios se están aplicando correctamente.

La elección del contexto para administrar el cuestionario deberá ser muy cuidadosa y dependerá del presupuesto disponible, el tiempo de entrega de los resultados, el planteamiento del problema, la naturaleza de los datos y el tipo de participantes (edad, nivel educativo, etcétera).

A continuación, en la tabla 9.6, se comparan esquemáticamente las principales formas de administración de cuestionarios.

► **Tabla 9.6** Comparación de las principales formas de administración de cuestionarios

Método de administración	Tasa de respuesta	Presupuesto o costo (fuente que origina el mayor gasto)	Rapidez con que se administra	Profundidad de los datos obtenidos	Tamaño del cuestionario
Autoadministrado (individual)	Media	Medio (pago de recolectores)	Media	Alta	Cualquier tamaño razonable
Autoadministrado (grupal)	Alta	Bajo (sesiones)	Rápida	Alta	Cualquier tamaño razonable
Autoadministrado (envío por correo o paquetería)	Baja	Bajo por correo postal (envíos) Medio por paquetería (envíos)	Lenta	Alta	Cualquier tamaño razonable
Autoadministrado por correo electrónico o página web	Baja	Bajo (diseño electrónico)	Media	Alta	Cualquier tamaño razonable
Entrevista personal	Alta	Elevado (pago a entrevistadores y gastos de viaje)	Media	Alta	Cualquier tamaño razonable
Entrevista telefónica	Alta	Bajo (llamadas telefónicas y entrevistadores)	Rápida	Baja	Corto

Cuando los cuestionarios son muy complejos de contestar o de aplicar, suele utilizarse un manual que explica las instrucciones y cómo debe responderse o administrarse.

Escalas para medir las actitudes

4

Una **actitud** es una predisposición aprendida para responder coherentemente de una manera favorable o desfavorable ante un objeto, ser vivo, actividad, concepto, persona o símbolo (Kassin, Fein y Markus, 2013; Devine y Plant, 2013; Oskamp y Schultz, 2009; y Fishbein y Ajzen, 1975). Así, los seres humanos tenemos actitudes hacia muy diversos objetos, símbolos, etc.; por ejemplo, actitudes hacia el aborto, la política económica, la familia, un profesor, diferentes grupos étnicos, la ley, nuestro trabajo, una nación específica, los osos, el nacionalismo, un proceso productivo, nosotros mismos, etcétera.

Las actitudes están relacionadas con el comportamiento que mantenemos en torno a los objetos o conceptos a que hacen referencia. Si mi actitud hacia el aborto es desfavorable, probablemente no abortaría o no participaría en un aborto. Si mi actitud es favorable a un partido político, lo más probable es que vote por él en las próximas elecciones. Desde luego, las actitudes sólo son un indicador de la conducta, pero no la conducta en sí. Por ello, las mediciones de actitudes deben interpretarse como “síntomas” y no como “hechos” (Padua, 2000). Si detecto que la actitud de un grupo hacia la contaminación es desfavorable, esto no significa que las personas estén tomando acciones para evitar contaminar el ambiente, aunque sí es un indicador de que pueden adoptarlas en forma paulatina. La actitud es como una “semilla” que en ciertas condiciones suele “germinar en comportamiento”.

Las actitudes tienen diversas propiedades, entre las que destacan: dirección (positiva o negativa) e intensidad (alta o baja); estas propiedades forman parte de la medición.

Actitud Predisposición aprendida para responder coherentemente de manera favorable o desfavorable ante un objeto, ser vivo, actividad, concepto, persona o símbolo.



Los métodos más conocidos para medir por escalas las variables que constituyen actitudes son: el método de escalamiento de Likert, el diferencial semántico y la escala de Guttman. A continuación examinaremos los primeros dos, que son los utilizados con mayor frecuencia. En el capítulo 7 adicional del centro de recursos el línea: “Recolección de los datos cuantitativos, segunda parte”, se comenta el tercer método: Escalograma de Guttman.

Escalamiento de Likert

Escalamiento de Likert Conjunto de ítems que se presentan en forma de afirmaciones para medir la reacción del sujeto en tres, cinco o siete categorías.

Este método fue desarrollado por Rensis Likert en 1932; sin embargo, se trata de un enfoque vigente y bastante popularizado.²⁰ Consiste en un conjunto de ítems presentados en forma de afirmaciones o juicios, ante los cuales se pide la reacción de los participantes. Es decir, se presenta cada afirmación y se solicita al sujeto que externé su reacción eligiendo uno de los cinco puntos o categorías de la escala. A cada punto se le asigna un valor numérico. Así, el participante obtiene una puntuación respecto de la afirmación y al final su puntuación total, sumando las puntuaciones obtenidas en relación con todas las afirmaciones.

Las afirmaciones califican al objeto de actitud que se está midiendo. El objeto de actitud puede ser cualquier “cosa física” (un vestido, un automóvil...), un individuo (el presidente, un líder histórico, mi madre, mi sobrino Alexis, un candidato a una elección...), un concepto o símbolo (patria, sexualidad, la mujer vallenata —Colombia—, el trabajo), una marca (Adidas, Ford...), una actividad (comer, beber café...), una profesión, un edificio, etc. Por ejemplo, Kafer *et al.* (1992) generaron varias escalas para medir las actitudes hacia los animales y Meerkerk *et al.* (2009) desarrollaron un instrumento basado en escalas de Likert para determinar la severidad del uso compulsivo de internet.

Tales frases o juicios deben expresar sólo una relación lógica; además, es muy recomendable que no excedan de 20 palabras.

Ejemplo

De frase

Objeto de actitud medido
El voto

Afirmación
“Votar es una obligación de todo ciudadano responsable”

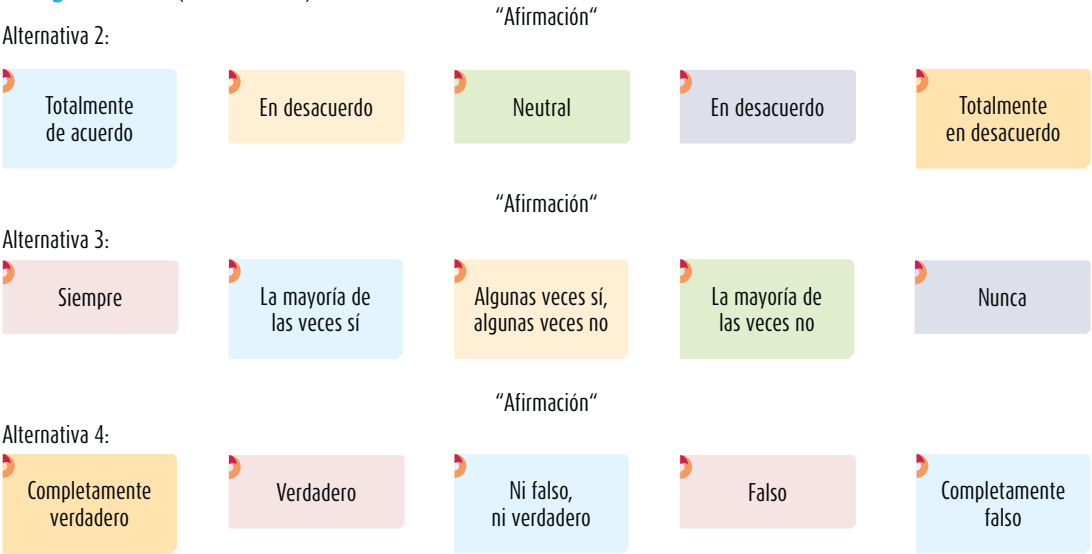
En este caso, la afirmación incluye ocho palabras y expresa una sola relación lógica (X—Y). Las opciones de respuesta o puntos de la escala son cinco e indican cuánto se está de acuerdo con la frase correspondiente.²¹ Las opciones más comunes se presentan en la figura 9.11. Debe recordarse que a cada una de ellas se le asigna un valor numérico (precodificado o no) y sólo puede marcarse una respuesta. Se considera un dato inválido si se marcan dos o más opciones.

Figura 9.11 Opciones o puntos en las escalas de Likert.



²⁰ Para conocer los orígenes de esta técnica, vea Likert (1976a o 1976b), Seiler y Hough (1976) y, particularmente, el libro original: Likert (1932).
²¹ Likert (1932), Clark (2000), Fabrigar y Wood (2006), Roberts y Jowell (2008), The SAGE Glossary of the Social and Behavioral Sciences (2009f) y Garson (2012).

Figura 9.11 (continuación)



Las opciones de respuesta o categorías pueden colocarse de manera horizontal, como en la figura 9.11, o en forma vertical.

- () Muy de acuerdo
- () De acuerdo
- () Ni de acuerdo ni en desacuerdo
- () En desacuerdo
- () Muy en desacuerdo

O bien, utilizando recuadros en lugar de paréntesis:

- ☐ Definitivamente sí
- ☐ Probablemente sí
- ☐ Indeciso
- ☐ Probablemente no
- ☐ Definitivamente no

Es indispensable señalar que el número de categorías de respuesta debe ser igual para todas las afirmaciones. Pero siempre respetando el mismo orden o jerarquía de presentación de las opciones para todas las frases (véase la tabla 9.7).

Tabla 9.7 Opciones jerárquicamente correctas e incorrectas en un ejemplo²²

Objeto de actitud: mi novia	
Correcto	Incorrecto (no se respeta la misma jerarquía en todos los ítems)
"Me gusta estar mucho con mi novia" ☐ Definitivamente sí ☐ Probablemente sí ☐ Indeciso ☐ Probablemente no ☐ Definitivamente no	"Me gusta estar mucho con mi novia" ☐ Probablemente sí ☐ Indeciso ☐ Definitivamente sí ☐ Probablemente no ☐ Definitivamente no

(continúa)

²² Es un ejemplo simple para ilustrar el concepto.

● **Tabla 9.7** (continuación)

Objeto de actitud: mi novia	
Correcto	Incorrecto (no se respeta la misma jerarquía en todos los ítems)
“Si por mí fuera, todos los días estaría con mi novia” ☐ Definitivamente sí ☐ Probablemente sí ☐ Indeciso ☐ Probablemente no ☐ Definitivamente no “Amo demasiado a mi novia” ☐ Definitivamente sí ☐ Probablemente sí ☐ Indeciso ☐ Probablemente no ☐ Definitivamente no	“Si por mí fuera, todos los días estaría con mi novia” ☐ Definitivamente sí ☐ Probablemente sí ☐ Probablemente no ☐ Definitivamente no ☐ Indeciso “Amo demasiado a mi novia” ☐ Definitivamente sí ☐ Definitivamente no ☐ Indeciso ☐ Probablemente no ☐ Probablemente sí

Dirección de las afirmaciones

Las afirmaciones pueden tener dirección: *favorable o positiva y desfavorable o negativa*. Y esta dirección es muy importante para saber cómo se codifican las alternativas de respuesta.

Si la afirmación es *positiva* significa que califica favorablemente al objeto de actitud; de este modo, cuanto más de acuerdo con la frase estén los participantes, su actitud será igualmente más favorable.

Ejemplo

“El Ministerio de Hacienda ayuda al contribuyente a resolver sus problemas en el pago de impuestos”.

En este ejemplo, si estamos “muy de acuerdo” con la afirmación implica una actitud más favorable hacia el Ministerio de Hacienda que si estamos solamente “de acuerdo”. En cambio, si estamos “muy en desacuerdo” significa una actitud muy desfavorable. Por tanto, *cuando las afirmaciones son positivas se califican comúnmente de la siguiente manera:*

(5) Muy de acuerdo

(4) De acuerdo

(3) Ni de acuerdo ni en desacuerdo

(2) En desacuerdo

(1) Muy en desacuerdo

Es decir, en el ejemplo, estar más de acuerdo implica una puntuación mayor.

Pero si la afirmación es *negativa*, significa que califica desfavorablemente al objeto de actitud, y cuanto más de acuerdo estén los participantes con la frase, implica que su actitud es menos favorable, esto es, más desfavorable.

Ejemplo

“El Ministerio de Hacienda se caracteriza por obstaculizar al contribuyente en el pago de impuestos”.

En este segundo ejemplo, si estamos “muy de acuerdo” implica una actitud más desfavorable que si estamos de “acuerdo”, y así en forma sucesiva. En contraste, si estamos “muy en desacuerdo” significa una actitud favorable hacia el Ministerio de Hacienda. Rechazamos la frase porque califica negativamente al objeto de actitud. Un ejemplo cotidiano de afirmación negativa sería: “Luis es un mal amigo”. Cuanto más de acuerdo estemos con el juicio, nuestra actitud hacia Luis será menos favorable. Es decir, estar más de acuerdo implica una puntuación menor. *Cuando las afirmaciones son negativas se califican al contrario de las positivas.*

- (1) Totalmente de acuerdo
- (2) De acuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) En desacuerdo
- (5) Totalmente en desacuerdo

En la figura 9.12 se presenta un ejemplo de una escala de Likert para medir la actitud hacia un organismo tributario.²³

● **Figura 9.12** Muestra de una escala de Likert.

Las afirmaciones que voy a leer son opiniones con las que algunas personas están de acuerdo y otras en desacuerdo. Voy a pedirle que me diga, por favor, qué tan de acuerdo está usted con cada una de estas opiniones:	
1. El personal de la Dirección General de Impuestos Nacionales es grosero al atender al público.	
1. Muy de acuerdo	4. En desacuerdo
2. De acuerdo	5. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
2. La Dirección General de Impuestos Nacionales se caracteriza por la deshonestidad de sus funcionarios.	
1. Muy de acuerdo	4. En desacuerdo
2. De acuerdo	5. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
3. Los servicios que presta la Dirección General de Impuestos Nacionales habitualmente son muy buenos.	
5. Muy de acuerdo	2. En desacuerdo
4. De acuerdo	1. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
4. La Dirección General de Impuestos Nacionales informa claramente sobre cómo, dónde y cuándo pagar los impuestos.	
5. Muy de acuerdo	2. En desacuerdo
4. De acuerdo	1. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
5. La Dirección General de Impuestos Nacionales es muy lenta en la devolución de impuestos pagados en exceso.	
1. Muy de acuerdo	4. En desacuerdo
2. De acuerdo	5. Muy en desacuerdo
3. Ni de acuerdo, ni en desacuerdo	
6. La Dirección General de Impuestos Nacionales informa oportunamente sobre cómo, dónde y cuándo pagar los impuestos.	
5. Muy de acuerdo	2. En desacuerdo
4. De acuerdo	1. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
7. La Dirección General de Impuestos Nacionales tiene normas y procedimientos bien definidos para el pago de impuestos.	
5. Muy de acuerdo	2. En desacuerdo
4. De acuerdo	1. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	
8. La Dirección General de Impuestos Nacionales tiene malas relaciones con la gente porque cobra impuestos muy altos.	
1. Muy de acuerdo	4. En desacuerdo
2. De acuerdo	5. Muy en desacuerdo
3. Ni de acuerdo ni en desacuerdo	

Como puede observarse en la figura 9.12, las afirmaciones 1, 2, 5 y 8 son negativas (desfavorables); y las afirmaciones 3, 4, 6 y 7 son positivas (favorables).

²³ El ejemplo fue utilizado en un país latinoamericano y su confiabilidad total fue de 0.89; aquí se presenta una versión muy reducida de la escala original. El nombre del organismo tributario que aquí se utiliza es ficticio.

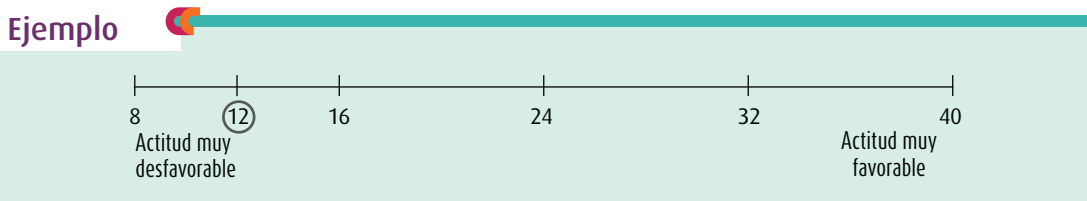
Forma de obtener las puntuaciones

Las puntuaciones de las escalas de Likert se obtienen sumando los valores alcanzados respecto de cada frase. Por ello se denomina *escala aditiva*. La figura 9.13, la cual se basa en la figura 9.12, constituiría un ejemplo de cómo calificar una escala de Likert.

Figura 9.13 Muestra de puntuaciones de la escala de Likert.

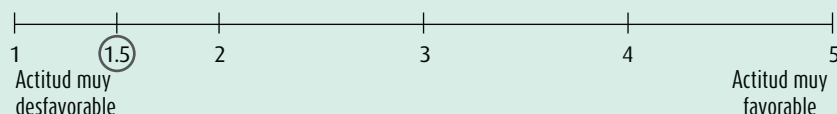
1. El personal de la Dirección General de Impuestos Nacionales es grosero al atender al público.	☒ 1. Muy de acuerdo	☐ 4. En desacuerdo
	☐ 2. De acuerdo	☐ 5. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
2. La Dirección General de Impuestos Nacionales se caracteriza por la deshonestidad de sus funcionarios.	☐ 1. Muy de acuerdo	☐ 4. En desacuerdo
	☒ 2. De acuerdo	☐ 5. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
3. Los servicios que presta la Dirección General de Impuestos Nacionales habitualmente son muy buenos.	☐ 5. Muy de acuerdo	☐ 2. En desacuerdo
	☐ 4. De acuerdo	☒ 3. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
4. La Dirección General de Impuestos Nacionales informa claramente sobre cómo, dónde y cuándo pagar los impuestos.	☐ 5. Muy de acuerdo	☐ 2. En desacuerdo
	☐ 4. De acuerdo	☐ 1. Muy en desacuerdo
	☒ 3. Ni de acuerdo ni en desacuerdo	
5. La Dirección General de Impuestos Nacionales es muy lenta en la devolución de impuestos pagados en exceso.	☒ 1. Muy de acuerdo	☐ 4. En desacuerdo
	☐ 2. De acuerdo	☐ 5. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
6. La Dirección General de Impuestos Nacionales informa oportunamente sobre cómo, dónde y cuándo pagar los impuestos.	☐ 5. Muy de acuerdo	☐ 2. En desacuerdo
	☐ 4. De acuerdo	☒ 3. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
7. La Dirección General de Impuestos Nacionales tiene normas y procedimientos bien definidos para el pago de impuestos.	☐ 5. Muy de acuerdo	☒ 2. En desacuerdo
	☐ 4. De acuerdo	☐ 1. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
8. La Dirección General de Impuestos Nacionales tiene malas relaciones con la gente porque cobra impuestos muy altos.	☒ 1. Muy de acuerdo	☐ 4. En desacuerdo
	☐ 2. De acuerdo	☐ 5. Muy en desacuerdo
	☐ 3. Ni de acuerdo ni en desacuerdo	
Valor = 1 + 2 + 1 + 3 + 1 + 1 + 2 + 1 = 12		

Una puntuación se considera alta o baja según el número de ítems o afirmaciones. Por ejemplo, en la escala para evaluar la actitud hacia el organismo tributario, la puntuación mínima posible es de ocho (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1) y la máxima es de 40 (5 + 5 + 5 + 5 + 5 + 5 + 5 + 5), porque hay ocho afirmaciones. La persona del ejemplo obtuvo 12. Su actitud hacia el organismo tributario es más bien bastante desfavorable; veámoslo gráficamente:



Si alguien hubiera tenido una puntuación de 37 ($5 + 5 + 4 + 5 + 5 + 4 + 4 + 5$) su actitud se calificaría como sumamente favorable. En *las escalas de Likert* a veces se califica el promedio resultante en la escala mediante la sencilla fórmula PT/NT (donde PT es la puntuación total en la escala y NT es el número de afirmaciones), y entonces una puntuación se analiza en el continuo 1-5 de la siguiente manera, con el ejemplo de quien obtuvo 12 en la escala ($12/8 = 1.5$).

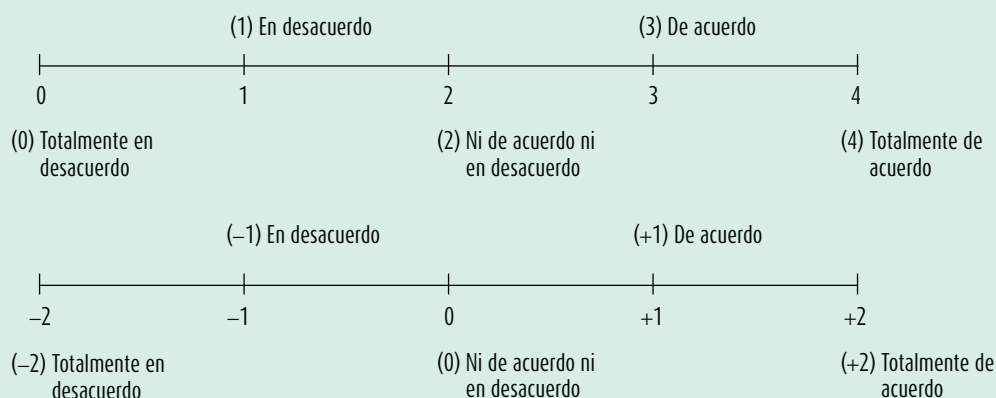
Ejemplo



La escala de Likert es, en sentido estricto, una medición ordinal; sin embargo, es común que se le trabaje como si fuera de intervalo. Creswell (2005), Pell (2005) y Carifio y Perla (2007) señalan que puede “forzarse” (no considerarse) a un nivel de medición por intervalos porque ha sido probada en múltiples ocasiones durante décadas y las personas están acostumbradas a escalar sus verbalizaciones con números. Pero otros autores, como Vega (2013), González, Carbonell y Santana (2011) y Jamieson (2004), opinan que tiene que concebirse como ordinal y analizarse como tal. Para profundizar en esta polémica, véase Garson (2012), Grashaw (2011), González *et al.* (2011), Brill (2008), Achyar (2008), Carifio y Rocco (2007 y 2008) y Hodge y Gillespie (2003 y 2007), así como a Fabrigar y Wood (2006).²⁴ Una solución con escalas de Likert puede ser efectuar tanto análisis para variables ordinales como de intervalos y ver si hay coincidencia (por ejemplo, si pretendemos correlacionar dos escalas de Likert, utilizar el coeficiente de Pearson y el de Spearman u otro y contrastar ambos valores).²⁵

Asimismo, a veces se utiliza un intervalo de 0 a 4 o de -2 a $+2$, en lugar de 1 a 5. Pero esto no importa porque se cambia el marco de referencia de la interpretación. Veámoslo gráficamente.

Ejemplo



Simplemente se ajusta el marco de referencia; pero el rango se mantiene y las categorías continúan siendo cinco.

²⁴ Vega (2013) argumenta que la escala de Likert es ordinal porque sus intervalos son desconocidos. Los autores a favor de aceptar trabajarla en un nivel por intervalos señalan que las opciones de respuesta siempre deben ir acompañadas de sus respectivos valores numéricos. De cualquier modo, debe tenerse en mente que lo que se “mide” son percepciones y lo que se considera más que números exactos son tendencias y las puntuaciones deben contextualizarse.

²⁵ Desde luego, no siempre hay pruebas equivalentes en ambos niveles de medición. Si hay duda, sugerimos consultar con su tutor o profesor de metodología y analizar en cada caso qué procede. Los alumnos se encontrarán con ambas posturas. Pero al final, lo importante es hacer investigación y aprender de esta experiencia.



Otras condiciones sobre la escala de Likert

A veces se disminuye o se incrementa el número de categorías, sobre todo cuando los participantes potenciales tienen una capacidad muy limitada de discriminación o, por el contrario, muy amplia. Por ejemplo, Lee y Guerin (2009) en su estudio comentado en capítulos previos, midieron sus variables o criterios de la satisfacción de la calidad del diseño ambiental del área de trabajo mediante dimensiones que incluían ítems perceptuales en el formato de Likert, cuyas escalas son de siete puntos que van de “muy satisfecho” (+3) a “muy insatisfecho” (-3), lo mismo para la satisfacción general sobre el espacio de trabajo.

Ejemplo



1 De acuerdo	0 En desacuerdo	
3 De acuerdo	2 Ni de acuerdo ni en des- acuerdo	1 En desacuerdo
7 Totalmente de acuerdo	6 De acuerdo	5 Indeciso, pero más bien
4 Indeciso, ni de acuerdo ni en desacuerdo	3 Indeciso, pero más bien en desacuerdo	
2 En desacuerdo	1 Totalmente en desacuerdo	

Si los participantes tienen poca capacidad de discriminar se pueden considerar dos o tres categorías. Por el contrario, si son personas con un nivel educativo elevado y gran capacidad de discriminación, pueden incluirse siete o más categorías. Pero debe recalcarse que el número de categorías de respuesta tiene que ser el mismo para todos los ítems. Si son tres, son tres categorías para todos los ítems o las afirmaciones. Si son cinco, son cinco categorías para todos los reactivos. En ocasiones se elimina la opción o categoría intermedia y neutral (ni de acuerdo ni en desacuerdo, neutral, indeciso...) para comprometer al sujeto o forzarlo a que se pronuncie de manera favorable o desfavorable.

Asimismo, como señalan Hodge y Gillespie (2003), algunos entrevistados gradúan su intensidad en un continuo que va del “fuertemente de acuerdo” a “neutral” y hasta el “fuertemente en desacuerdo”, mientras que otros entienden esta categoría central como un “no sé” o “no aplica”. Estos individuos ven al punto neutral o medio como una extensión de la dimensión de contenido, considerándolo como una opción de respuesta cuando no poseen suficiente información. En este caso, resulta apropiado ignorar tales respuestas cuando se calcula la puntuación total (Raaijmakers *et al.*, 2000 y Hodge y Gillespie, 2003). Si después de la prueba piloto se observa que una cuarta parte o más de los respondientes tiende a irse a la categoría neutral en un ítem, es necesario revisarlo e incluso eliminarlo. Si esto ocurre en varios reactivos, se debe eliminar tal categoría o revisar a fondo la escala.

Un aspecto muy importante de la escala de Likert es que asume que los ítems o las afirmaciones miden la actitud hacia un único concepto subyacente. En el caso de que se midan actitudes hacia varios objetos, deberá incluirse una escala por objeto, porque aunque se presenten conjuntamente, se califican por separado. *En cada escala de Likert se considera que todos los ítems tienen igual peso.*

Cómo se construye una escala de Likert

En términos generales, una escala de Likert se construye generando un elevado número de afirmaciones que califiquen al objeto de actitud y se administran a una muestra piloto para obtener las puntuaciones del grupo en cada ítem o frase. Estas puntuaciones se correlacionan con las del grupo a toda la escala (la suma de las puntuaciones de todas las afirmaciones), y las frases o reactivos, cuyas

puntuaciones se correlacionen significativamente con las puntuaciones de toda la escala, se seleccionan para integrar el instrumento de medición. Asimismo, debe calcularse la confiabilidad y validez de la escala.

Preguntas en lugar de afirmaciones

En la actualidad, la escala original con frases se ha extendido a preguntas y observaciones, como se puede apreciar en el siguiente ejemplo para evaluar al conductor de un programa televisivo.

Ejemplo

¿Cómo considera usted al conductor del programa...?

- | | | |
|---|--|------------------------------------|
| ☐ 5 Muy buen conductor | ☐ 4 Buen conductor | ☐ 3 Regular |
| ☐ 2 Mal conductor | ☐ 1 Muy mal conductor | |

Otro ejemplo sería un conjunto de preguntas formuladas en una investigación para analizar la relación de compraventa en empresas mexicanas (Paniagua, 1985). De ella se presenta un fragmento en la tabla 9.8.²⁶

► **Tabla 9.8** Ejemplo de la escala de Likert aplicada a varias preguntas

¿Para elegir a sus proveedores qué tan importante es...	Indispensable (5)	Sumamente importante (4)	Medianamente importante (3)	Poco importante (2)	No se toma en cuenta (1)
el precio?	5	4	3	2	1
la forma de pago (contado/ crédito)?	5	4	3	2	1
el tiempo de entrega?	5	4	3	2	1
el lugar de entrega?	5	4	3	2	1
la garantía del producto?	5	4	3	2	1
el prestigio del producto?	5	4	3	2	1
el prestigio de la empresa proveedora?	5	4	3	2	1
el cumplimiento del proveedor con las especificaciones?	5	4	3	2	1
la información que sobre el producto proporcione el proveedor?	5	4	3	2	1
el tiempo de trabajar con el proveedor?	5	4	3	2	1
la entrega del producto en las condiciones acordadas?	5	4	3	2	1
la calidad del producto?	5	4	3	2	1
Las respuestas se califican del modo que ya hemos comentado.					

²⁶ Estas preguntas se han seguido utilizando en estudios más actuales, siguen siendo vigentes.

La escala en la pregunta

En ocasiones la escala se incluye en la pregunta. Mertens (2010) las denomina *preguntas actitudinales*, por ejemplo: ¿está usted fuertemente a favor, más bien a favor, más bien en contra o fuertemente en contra del aborto cuando la mujer ha sido violada?

En la pregunta se eliminó la categoría central o intermedia. Pero estas interrogantes suelen limitarse a entrevistas de unas cuantas preguntas, porque requieren cierta capacidad de memorización.

Método de completar las frases

Hodge y Gillespie (2003) desarrollaron una derivación del escalamiento clásico de Likert, en la cual se incluyen frases incompletas respecto al objeto de actitud y a éstas se les agrega un continuo que sirve como base para las respuestas claves. Estos autores plantearon un continuo con 11 puntos o categorías (0 a 10) y que se “ancla” en cada extremo con terminaciones respecto a la frase a la que hacen referencia, las cuales representan la ausencia del constructo (cero) y la máxima “cantidad” o “presencia” de éste (10). Sostienen que los participantes usan un número para guiarse en sus respuestas, y la frase introductoria los orienta en el continuo. Se les pide que circunscriban o marquen el número que refleje mejor su respuesta. El constructo se mide por ítems que enfatizan la fuerza del atributo. Los números trabajan en concordancia con las frases para implicar el grado de presencia de éste. El ejemplo, lo sería la actitud intrínseca hacia la religión.²⁷

Ejemplo



Escala en la que se completan las afirmaciones

Actitud intrínseca hacia la religión

1. Mis creencias religiosas afectan:

Ningún aspecto de
mi vida

Absolutamente todos los
aspectos de mi vida

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

2. Estoy consciente de la presencia de Dios...

Continuamente

Nunca

10	9	8	7	6	5	4	3	2	1	0
----	---	---	---	---	---	---	---	---	---	---

3. Respecto a las preguntas que tengo sobre la vida, mi religión responde...

Absolutamente a ninguna de
mis preguntas

Absolutamente a todas mis
preguntas

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

4. Mi religión es...

El motivo más importante de
mi vida dirige todos los
demás aspectos

No es un factor en mi
vida

10	9	8	7	6	5	4	3	2	1	0
----	---	---	---	---	---	---	---	---	---	---

²⁷ Adaptado al español y modificado después de una prueba piloto (Hodge y Gillespie, 2003, p. 52).

5. Leo libros, busco información en internet y veo programas que se relacionen con mi fe...

Nunca

Todos los días,
sin fallar

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

6. Busco momentos para meditar y pensar sobre mi religión y Dios...

Todos los días,
sin fallar

Nunca

10	9	8	7	6	5	4	3	2	1	0
----	---	---	---	---	---	---	---	---	---	---

Al construir una escala de Likert, debemos asegurarnos de que las afirmaciones y alternativas de respuesta serán comprendidas por los participantes a los que se les aplicará y que éstos tendrán la capacidad de discriminación requerida. Ello se evalúa cuidadosamente en la prueba piloto. Las escalas pueden ser autoadministradas o aplicadas mediante entrevistas; en este último caso, es recomendable mostrar al entrevistado una tarjeta donde se presenten las alternativas de respuestas o categorías. Asimismo, las escalas de Likert pueden integrarse dentro de un cuestionario.²⁸

Diferencial semántico

El **diferencial semántico** fue planteado por Osgood, Suci y Tannenbaum (1957) para explorar las dimensiones del significado.²⁹ Consiste en una serie de adjetivos extremos que califican al objeto de actitud, ante los cuales se solicita la reacción del participante. Es decir, éste debe calificar al objeto de actitud a partir de un conjunto de adjetivos bipolares; entre cada par de éstos, se presentan varias opciones y la persona selecciona aquella que en mayor medida refleje su actitud.

Diferencial semántico Serie de pares de adjetivos extremos que sirven para calificar al objeto de actitud, ante los cuales se pide la reacción del sujeto, al ubicarlo en una categoría por cada par.

Ejemplo

Escala bipolar

Objeto de actitud: candidato "A"

justo : _____ : _____ : _____ : _____ : _____ : _____ : _____ : injusto

Debe observarse que los adjetivos son "extremos" y que entre ellos hay siete opciones de respuesta. Cada participante califica al candidato "A" en términos de esta escala de adjetivos bipolares.

Osgood, Suci y Tannenbaum (1957) nos indican que si el respondiente considera que el objeto de actitud se relaciona *muy estrechamente* con uno u otro extremo de la escala, la respuesta se marca así:

justo: **X** : _____ : _____ : _____ : _____ : _____ : _____ : injusto

O de la siguiente manera:

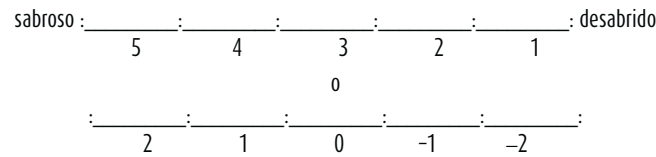
justo: _____ : _____ : _____ : _____ : _____ : _____ : **X** : injusto

²⁸ Para compenetrarse con la técnica de Likert desde una perspectiva estadística, se recomienda a: Dittrich, Francis, Hatzinger y Katzenbeisser (2007); Beal y Dawson (2007); y Adelson y McCoach (2010).

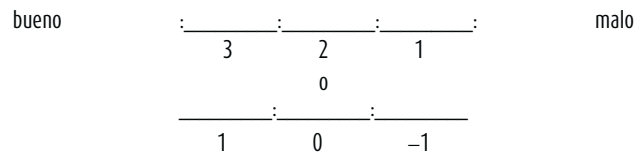
²⁹ Para profundizar en el diferencial semántico, véase: Osgood, Suci y Tannenbaum (1957, 1976a y 1976b), así como Heise (1976) y Al-Hindawi (2013).

● **Figura 9.14** (continuación)

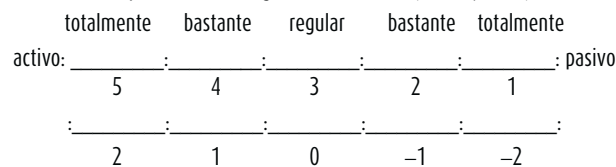
En los casos en que los respondientes tengan menor capacidad de discriminación, se pueden reducir las categorías a cinco opciones. Por ejemplo:



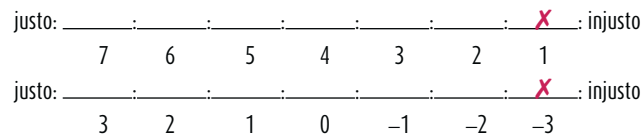
o aun a tres opciones (lo cual es poco común):



También pueden agregarse calificativos a los puntos o las categorías de la escala (Babbie, 2012).



Codificar de 1 a 7 o de -3 a 3 no tiene importancia, siempre y cuando estemos conscientes del marco de interpretación. Por ejemplo, si una persona califica al objeto de actitud: candidato “A” en la escala justo-injusto, marcando la categoría más cercana al extremo “injusto”, la puntuación sería “1” o “-3”.



En un caso la escala oscila entre 1 y 7, y en el otro caso entre -3 y 3. Si deseamos evitar el manejo de números negativos utilizamos la escala de 1 a 7.

El diferencial semántico (DS) se ha utilizado en diversas situaciones para evaluar “objetos” de actitud. Por ejemplo, Lilja *et al.* (2004) emplearon un instrumento con 57 pares de adjetivos bipolares con la finalidad de apreciar la actitud de un grupo de enfermeras respecto a ciertos pacientes psiquiátricos y su orientación hacia ellos (enfocadas en el “ser humano” y en establecer una relación genuina y duradera, o bien centradas en simplemente corregir la conducta “defectuosa” del paciente). Shields (2007) aplicó el DS para examinar las actitudes y opiniones del personal de apoyo y los padres en torno al cuidado de niños hospitalizados en cuatro países (dos desarrollados: Australia y Gran Bretaña, y dos subdesarrollados: Indonesia y Tailandia). Salcuni *et al.* (2007) usaron esta técnica en Italia a fin de evaluar las representaciones que hacen los padres respecto de sus hijos (6 a 11 años de edad). Mientras que Bauer (2008) lo utilizó para determinar actitudes hacia la Química (como ciencia y materia) por parte de estudiantes universitarios.

Otro estudio es el de Friborg, Martinussen y Rosenvinge (2006), quienes midieron mediante una escala tipo Likert y un diferencial semántico la resiliencia en alumnos de licenciatura (capacidad de los individuos para sobreponerse de acontecimientos desestabilizadores, condiciones de vida difíciles, periodos de dolor emocional y traumas psicológicos). Asimismo, Miller *et al.* (2008), en una investigación con pacientes que padecían la enfermedad de Parkinson y que tenía el objetivo de examinar los cambios percibidos en la propia comunicación y con el especialista en salud que los atendía, y relacionarlos con la inteligibilidad del habla, el género, la edad y otras enfermedades, recurrieron al diferencial semántico como instrumento de medición. También, Geake y Gross (2008) emplearon el DS para detectar las actitudes subconscientes de docentes de educación básica hacia alumnos “superdotados” en Inglaterra, Escocia y Australia; y Gluth, Ebner y Schmiedek (2010) validaron una escala para medir actitudes hacia los adultos jóvenes y viejos mediante el diferencial semántico en Alemania.

Su interpretación depende del número de ítems o pares de adjetivos. Asimismo, en ocasiones se califica el promedio obtenido en la escala total.

$$\left(\frac{\text{puntuación total}}{\text{número de ítems}} \right)$$

Se pueden utilizar distintas escalas o diferenciales semánticos para medir actitudes hacia varios objetos. Por ejemplo, es posible medir con cuatro pares de adjetivos la actitud hacia el candidato “A”, con otros tres pares de adjetivos la actitud respecto de su plataforma ideológica, y con otros seis pares de adjetivos la actitud hacia su partido político. Tenemos tres escalas, cada una con distintos pares de adjetivos para medir la actitud en relación con tres diferentes conceptos (“objetos de actitud”).

El *diferencial semántico* es una escala de medición ordinal, aunque es común que se le trabaje como si fuera de intervalo (Key, 1997), por las mismas razones de Likert.

Escalograma de Guttman

El **escalograma de Guttman** es otra técnica para medir las actitudes y al igual que la escala de Likert, se fundamenta en afirmaciones o juicios respecto del concepto u objeto de actitud, ante los cuales los participantes deben externar su opinión seleccionando uno de los puntos o categorías de la escala respectiva. Una vez más, a cada categoría se le otorga un valor numérico. Así, el participante obtiene una puntuación respecto de la afirmación y al final su puntuación total, sumando las puntuaciones obtenidas en relación con todas las afirmaciones.

La diferencia con el método de Likert es que las frases tienen diferentes intensidades (se escalan por tal intensidad), por ejemplo, la siguiente afirmación en relación con el aborto (actitud evaluada): “si en estos momentos me embarazara, jamás abortaría”; es más intensa que esta otra: “si una de mis mejores amigas se embarazara, nunca le recomendaría abortar”, y a su vez, esta última resulta más intensa que la afirmación: “si una compañera del salón de clases se embarazara, probablemente no le recomendaría abortar”. Es decir, se basa en el principio de que algunos ítems indican en mayor medida la fuerza o intensidad de la actitud.

Por cuestiones de espacio, el escalograma de Guttman no se comenta en este capítulo, sino en el capítulo 7 adicional del centro de recursos en línea: “Recolección de los datos cuantitativos: segunda parte” (que se puede encontrar y descargar en: Material complementario → Capítulos → Capítulo 7).

Escalograma de Guttman Técnica para medir las actitudes, que al igual que la escala de Likert, se fundamenta en juicios ante los cuales los participantes deben externar su opinión seleccionando uno de los puntos o categorías de la escala respectiva.



Otros métodos cuantitativos de recolección de los datos

En la investigación disponemos de otros métodos sumamente útiles para recolectar los datos de las diferentes ciencias y disciplinas, como los cuestionarios y las escalas de actitudes, los cuales solamente se enuncian en este capítulo, pero se comentan con mayor profundidad en el capítulo 7 adicional: “Recolección de los datos cuantitativos: segunda parte”, que se puede descargar del centro de recursos en línea.³⁰ Entre tales técnicas se encuentran:



1. Análisis de contenido cuantitativo

Es una técnica para estudiar cualquier tipo de comunicación de una manera “objetiva” y sistemática, que cuantifica los mensajes o contenidos en categorías y subcategorías, y los somete a análisis estadístico.

Sus usos son muy variados; por ejemplo: evaluar el grado de carga de contenido sexual de uno o varios programas televisivos; estudiar las apelaciones y características de campañas publicitarias (digamos, de perfumes femeninos de costo elevado) en los medios de comunicación colectiva (radio,

³⁰ Como se ha insistido, el método o instrumento de recolección de los datos depende de diversos factores como el planteamiento del problema, las hipótesis y las definiciones operacionales.



televisión, periódicos y revistas); comparar estrategias propagandísticas de partidos políticos en internet; conocer discrepancias ideológicas entre varios periódicos al tratar un tema como el terrorismo internacional; determinar la evolución de cierta clase de pacientes que asisten a psicoterapia al analizar sus escritos y expresiones verbales; cotejar el vocabulario aprendido por niños que se exponen más al uso de la computadora en comparación con infantes que la utilizan menos; conocer y contrastar la posición de diversos presidentes latinoamericanos en cuanto al problema del desempleo; comparar estilos de escritores que se señalan como parte de una misma corriente literaria; o analizar la calidad y profundidad de la información de internet sobre un virus.

Una investigación de este tipo es la de Guillaume y Bath (2008), quienes estudiaron la cobertura y el tratamiento que se daba en la prensa británica a la información sobre las vacunas para el sarampión, las paperas y la rubéola durante un periodo de dos meses. Hall y Wright (2008) aplicaron el análisis de contenido para examinar opiniones judiciales. Incluso, Christie y Fleischer (2010) lo utilizaron para determinar los diseños y métodos más utilizados en estudios de evaluación publicados en revistas científicas.

2. Observación

Este método de recolección de datos consiste en el registro sistemático, válido y confiable de comportamientos y situaciones observables, a través de un conjunto de categorías y subcategorías. Útil, por ejemplo, para analizar conflictos familiares, eventos masivos (como la violencia en los estadios de fútbol), la aceptación-rechazo de un producto en un supermercado, el comportamiento de personas con capacidades mentales distintas, la adaptación de operarios a una nueva maquinaria, etc. Haynes (1978) menciona que es el método más utilizado por quienes se orientan conductualmente.

Como muestras de este tipo de investigación podemos citar a Regina *et al.* (2008), quienes, utilizando una técnica conocida como la lista de verificación de la conducta autista, compararon las observaciones de profesionales de la salud en torno a los comportamientos de niños autistas brasileños con las observaciones de sus madres. Asimismo, Franco, Rodrigues y Balcells (2008) evaluaron la pedagogía de instructores de ejercicios físicos y aeróbicos en tres gimnasios de Portugal, al analizar por observación clases grabadas en video. Labus, Keefe y Jensen (2003) revisaron estudios para indagar sobre la relación entre los autorreportes de intensidad del dolor y las observaciones directas de la conducta producida por tal dolor.

3. Pruebas estandarizadas e inventarios

Estas pruebas o inventarios miden variables específicas, como la inteligencia, la personalidad en general, la personalidad autoritaria, el razonamiento matemático, el sentido de vida, la satisfacción laboral, el tipo de cultura organizacional, el estrés preoperatorio, la depresión posparto, la adaptación al colegio, intereses vocacionales, la jerarquía de valores, el amor romántico, la calidad de vida, la lealtad a una marca de algún producto, etc. Hay miles.

Asimismo, hay un tipo de pruebas que evalúan proyecciones de los participantes y determinan su estado en una variable, con elementos cuantitativos y cualitativos: las pruebas proyectivas como el *test de Rorschach* (que presenta manchas de tinta en tarjetas o láminas blancas numeradas a los sujetos y éstos relatan sus asociaciones e interpretaciones en relación con tales manchas).

4. Datos secundarios (recolectados por otros investigadores)

Implica la revisión de documentos, registros públicos y archivos físicos o electrónicos. Por ejemplo, si nuestra hipótesis fuera: “la violencia manifiesta en la Ciudad de México es mayor que en la ciudad de Caracas”; entonces acudiríamos a las alcaldías de las ciudades para solicitar datos relacionados con la violencia, como número de asaltos, violaciones, robos a casa-habitación, asesinatos, etc. (datos generales, por distrito y habitante). También obtendríamos información de los archivos de los hospitales y las diferentes procuradurías o cuerpos policiacos. Un caso de una investigación cuyo método de recolección se fundamentó en datos secundarios fue el que a continuación se comenta.

Ejemplo



Un grupo de investigadores efectuó (en 2008 y 2009) un estudio para explorar el impacto que tienen las becas otorgadas o gestionadas por una institución de educación superior sobre el desarrollo académico de los alumnos beneficiarios y su deserción escolar.³¹

Los investigadores solicitaron a las diferentes direcciones información de los estudiantes respecto a su promedio general en la carrera, nivel socioeconómico, estatus respecto a la beca (becado-no becado), tipo de beca (institucional, otorgada por el Ministerio de Educación, por organismo privado, con fondos del gobierno estatal), monto de la beca, estatus académico del alumno (regular, irregular, desertor), semestre que cursa, género y edad, entre otras cuestiones. Consideraron los últimos cinco años escolares. Con tal información construyeron una base de datos (con más de medio millón de registros) y efectuaron análisis. Entre otras cuestiones encontraron que el promedio de los becarios era muy superior al de los no becarios y la deserción escolar era mínima entre los primeros, casi inexistente. Pero no encontraron una relación entre el monto de la beca y el promedio general de la carrera (acumulado). Asimismo, descubrieron que las mujeres tenían en general mejor promedio que sus compañeros.

Comparar indicadores económicos de países de la Comunidad Europea, analizar la relación comercial entre dos naciones, cotejar el número, eficacia y tipo de casos atendidos por diferentes hospitales, contrastar la eficacia con que se insertan en el mundo laboral los egresados de una carrera de distintas universidades, evaluar las tendencias electorales en un país, antes y después de un suceso crítico (como lo fueron los actos terroristas en Madrid en 2004 o el fallecimiento del presidente de la República Bolivariana de Venezuela Hugo Chávez en 2013), son ejemplos donde la recolección y análisis de datos secundarios son la base de la investigación.

5. Instrumentos mecánicos o electrónicos

Sistemas de medición por aparatos, como el detector de mentiras o polígrafo, que considera la respuesta galvánica de la piel (en investigaciones sobre crímenes); la pistola láser, que mide la velocidad a la que circula un automóvil desde un punto externo al vehículo (en estudios sobre el comportamiento de conductores); instrumentos que captan la actividad cerebral (evaluaciones médicas y psicológicas); el escáner, que mide con exactitud el cuerpo de un ser humano y ubica la talla ideal para confeccionar toda su ropa o vestuario (en investigaciones para diseñar los uniformes de los soldados); la medición electrónica de distancias, etcétera.

6. Instrumentos y procedimientos específicos propios de cada disciplina

En todas las áreas de estudio se han generado valiosos métodos para recolectar datos sobre variables específicas. Por ejemplo, en la comunicación organizacional se utilizan formatos para evaluar el uso que hacen los ejecutivos de los medios de comunicación interna (teléfono, reuniones, internet, etc.), así como herramientas para conocer procesos de comunicación en la empresa (la auditoría en comunicación). Para el análisis de grupos se usan los sistemas sociométricos y el análisis de redes. La medición de la presión arterial implica un procedimiento preciso, etcétera.

Por ejemplo, Pérez, Arango y Agudelo (2009) en su experimento cuyo objetivo fue determinar el efecto que tienen los factores: experiencia del operario, tipo de dobladora, clase de material utilizado y su grosor sobre la longitud y el ángulo de doblado de las piezas de metal producidas, para la recolección de sus datos establecieron “que las piezas fueran dobladas transversalmente una longitud de 15 mm a un ángulo de 90°, y según estos parámetros ambos operarios doblaron retales de láminas en los dos materiales seleccionados (acero laminado en frío y acero galvanizado) con los grosores especificados (0.8 mm y 1 mm) y en cada una de las dos dobladoras. De esta forma se obtuvieron los datos para las dos variables respuesta medidas: longitud de doblado y ángulo de doblado, las cuales se midieron usando pie de rey y goniómetro” (Pérez, Arango y Agudelo, 2009, p. 148).

³¹ No se menciona el nombre de la institución ni del estudio porque ésta solicitó el anonimato.



“Como se especificó, las dimensiones de las piezas dobladas se describen por la longitud y el ángulo de doblado. Luego, se tienen dos variables respuesta para analizar” (Pérez, Arango y Agudelo, 2009, p. 149):

Y_{ijklm} = longitud de doblado obtenida en la dobladora i , por el operario j , sobre la lámina de metal con material k y grosor l , en la réplica m .

Z_{ijklm} = ángulo de doblado obtenido en la dobladora i , por el operario j , sobre la lámina de metal con material k y grosor l , en la réplica m .

¿Puede utilizarse más de un tipo de instrumento de recolección de datos?

Cada día es más común ver estudios en los que se utilizan diferentes métodos de recolección de datos. En los estudios cuantitativos no resulta extraño que se incluyan varios tipos de cuestionarios al mismo tiempo que pruebas estandarizadas y recopilación de contenidos para análisis estadístico u observación. Incluso, al utilizar diversos instrumentos se ayuda a establecer la validez de criterio. No solamente se puede, sino que es conveniente, hasta donde lo permita el presupuesto para investigar.

¿Cómo se codifican las respuestas a un instrumento de medición?

Una vez recolectados los datos, deben codificarse. Ya hemos dicho que las categorías de un ítem, indicador o pregunta requieren codificarse con símbolos o números; y esto debe hacerse, porque de lo contrario no se efectuaría ningún análisis o sólo se contaría el número de respuestas o casos en cada categoría (por ejemplo, 25 contestaron “sí” y 24 respondieron “no”).³² Comúnmente, el investigador se interesa en realizar análisis más allá de un conteo de casos por categoría, y éstos se llevan a cabo por medio de la computadora u ordenador. Para ello es necesario transformar las respuestas en símbolos o valores numéricos. Los datos deben resumirse, codificarse y prepararse para el análisis. También se comentó que las categorías pueden ir o no precodificadas (incluir la codificación en el instrumento de medición) y que las preguntas abiertas no están precodificadas. Desde luego, los aparatos incluyen las categorías o lectura de la escala en el propio instrumento (por ejemplo, termómetros, acelerómetros, básculas, electrocardiógrafos, etcétera).

Los valores perdidos y su codificación

Cuando las personas no responden a un ítem, contestan incorrectamente (por ejemplo, marcan dos opciones, cuando las alternativas eran mutuamente excluyentes) o no puede registrarse la información (por ejemplo, si tenemos varias simulaciones y en una de ellas no se pudo obtener el dato de una de las variables), se crean una o varias categorías de valores perdidos y se les asignan sus respectivos códigos.

Ejemplo



Sí = 1
No = 0
No contestó = 3
Contestó incorrectamente = 4

Sí = 1
No = 0
Valor perdido por diversas razones = 9

Asimismo, tenemos el caso de preguntas que **no** aplican a ciertos participantes, en estas situaciones debe considerarse y codificarse la categoría: “no aplica” o “no corresponde”. Por ejemplo, si un

cuestionario administrado mediante entrevista a mujeres contuviera las siguientes dos preguntas: ¿durante el último mes realizó alguna compra en la tienda de ropa femenina “Ensueños”?³³ y ¿me podría indicar qué artículos o prendas compró?, y una entrevistada nos contestara a la primera pregunta que “no” (no había comprado en la tienda), anotaríamos esta categoría, y obviamente no haríamos la segunda pregunta, sino que marcaríamos la opción “no aplica” (la pregunta).

Los valores perdidos pueden reducirse con instrumentos que motiven al participante y no sean muy largos, con instrucciones claras y capacitación a los entrevistadores. Un alto número de valores perdidos (más de 10%) indica que el instrumento tiene problemas. Lo adecuado es que no supere 5% respecto del total de posibles datos o valores.

En la forma tradicional, la codificación de las respuestas a preguntas o afirmaciones implica cuatro pasos que comentaremos brevemente sólo para que se refuercen algunos conceptos:

1. Establecer los códigos de las categorías o alternativas de respuesta de los ítems o preguntas

Cuando todas las categorías fueron precodificadas y no se tienen preguntas abiertas, este primer paso no es necesario, ya se efectuó. Si las categorías no fueron precodificadas o se tienen preguntas abiertas, deben asignarse los códigos o la codificación a todas las categorías de los ítems. Por ejemplo:

Pregunta no precodificada:

¿Practica usted algún deporte por lo menos una vez a la semana?

☐ Sí ☐ No

Se codifica:

1 = Sí 0 = No

Frase no precodificada:

“Creo que estoy recibiendo un salario justo por mi trabajo”.

() Totalmente de acuerdo () De acuerdo () Ni de acuerdo ni en desacuerdo

() En desacuerdo () Totalmente en desacuerdo

Se codifica:

5 = Totalmente de acuerdo

4 = De acuerdo

3 = Ni de acuerdo ni en desacuerdo

2 = En desacuerdo

1 = Totalmente en desacuerdo

El tema sobre la codificación de preguntas abiertas ya se expuso.

2. Elaborar el libro de códigos incluyendo todos los ítems, uno por uno

Una vez que están codificadas todas las categorías de los ítems, se procede a elaborar el “libro de códigos”, el cual describe la localización de las variables y los códigos asignados a las categorías *en una matriz o base de datos*. Los elementos comunes de un libro de códigos son: variables de la investigación, preguntas, indicadores o ítems; categorías, códigos (números o símbolos utilizados para asignarse a las categorías) y número de columna en la matriz de datos a que corresponde cada ítem.

Supongamos que tenemos una escala de Likert con tres ítems (frases):

1. “La Dirección General de Impuestos Nacionales informa oportunamente sobre cómo, dónde y cuándo pagar los impuestos”.

(5) Muy de acuerdo

(4) De acuerdo

³³ Nombre ficticio.

- (3) Ni de acuerdo ni en desacuerdo
- (2) En desacuerdo
- (1) Muy en desacuerdo
2. “Los servicios que presta la Dirección General de Impuestos Nacionales habitualmente son muy buenos”.
- (5) Muy de acuerdo
- (4) De acuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (2) En desacuerdo
- (1) Muy en desacuerdo
3. “La Dirección General de Impuestos Nacionales se caracteriza por la deshonestidad de sus funcionarios”.
- (1) Muy de acuerdo
- (2) De acuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) En desacuerdo
- (5) Muy en desacuerdo

El libro de códigos sería el que se muestra en la tabla 9.9.

● **Tabla 9.9** Ejemplo de un libro o documento de códigos con una escala de actitud tipo Likert (tres ítems)

Variable	Ítem	Categorías	Códigos	Columnas
Actitud hacia la Dirección General de Impuestos Nacionales	Frase 1 (informa)	— Muy de acuerdo	5	1
		— De acuerdo	4	
		— Ni de acuerdo ni en desacuerdo	3	
		— En desacuerdo	2	
		— Muy en desacuerdo	1	
	Frase 2 (servicios)	— Muy de acuerdo	5	2
		— De acuerdo	4	
		— Ni de acuerdo ni en desacuerdo	3	
		— En desacuerdo	2	
		— Muy en desacuerdo	1	
	Frase 3 (deshonestidad)	— Muy de acuerdo	1	3
		— De acuerdo	2	
		— Ni de acuerdo ni en desacuerdo	3	
		— En desacuerdo	4	
		— Muy en desacuerdo	5	

Es decir, el libro de códigos es una especie de manual para interpretar la matriz de datos (la cual, como veremos más adelante, es una matriz en Excel, SPSS —Paquete Estadístico para las Ciencias Sociales—, Minitab o cualquier otro programa similar).

3. Efectuar físicamente la codificación

El tercer paso del proceso es la codificación física de los datos, es decir, llenar la matriz de datos con los valores implicados en las respuestas o las lecturas al instrumento de medición (transferir éstas a la matriz).

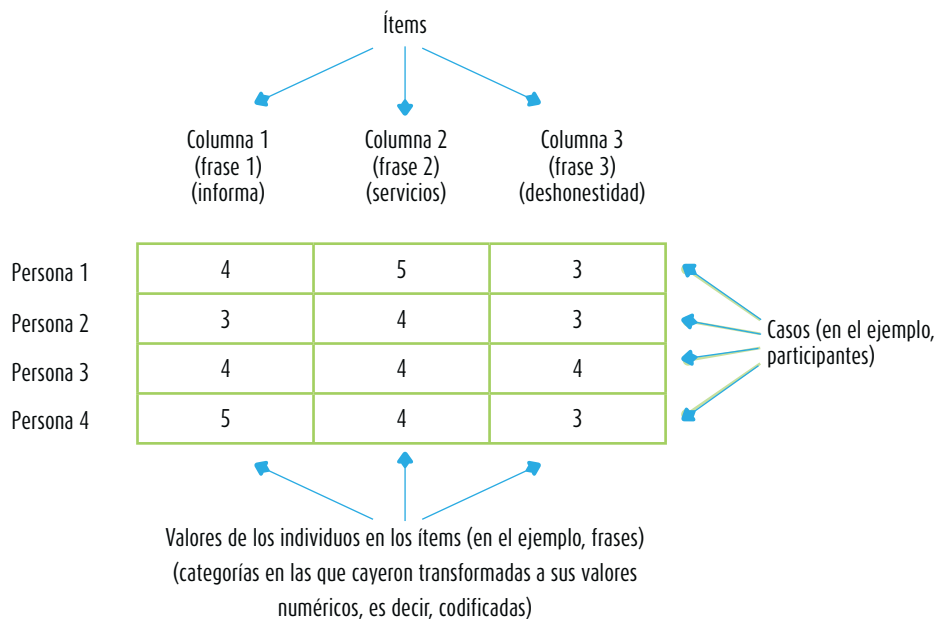
Veamos un ejemplo simplificado con la escala de Likert de tres ítems aplicada a cuatro individuos (figura 9.17).

● **Figura 9.17** Ejemplo de aplicación de tres ítems a cuatro sujetos.

Persona 1		
A continuación...		
1. "La Dirección General de Impuestos Nacionales informa oportunamente sobre cómo, dónde y cuándo pagar los impuestos".		
(5) Muy de acuerdo	(X) De acuerdo	(3) Ni de acuerdo ni en desacuerdo (2) En desacuerdo (1) Muy en desacuerdo
2. "Los servicios que presta la Dirección General de Impuestos Nacionales son en general muy buenos".		
(X) Muy de acuerdo	(4) De acuerdo	(3) Ni de acuerdo ni en desacuerdo (2) En desacuerdo (1) Muy en desacuerdo
3. "La Dirección General de Impuestos Nacionales se caracteriza por la deshonestidad de sus funcionarios".		
(1) Muy de acuerdo (2) De acuerdo	(X) Ni de acuerdo ni en desacuerdo	(4) En desacuerdo (5) Muy en desacuerdo
El participante obtuvo:		
	4 (de acuerdo)	
	5 (muy de acuerdo)	
	3 (ni de acuerdo ni en desacuerdo)	
Persona 2		
Obtuvo respectivamente:		
	3 (ni de acuerdo ni en desacuerdo)	
	4 (de acuerdo)	
	3 (ni de acuerdo ni en desacuerdo)	
Persona 3		
Obtuvo respectivamente:		
	4	
	4	
	4	
Persona 4		
Obtuvo respectivamente:		
	5	
	4	
	3	

De acuerdo con el libro de códigos de la tabla 9.9 y las respuestas a la escala, tendríamos la matriz de la figura 9.18.

● **Figura 9.18** Ejemplo de matriz de datos para el libro de códigos de la tabla 9.9.



4. Guardar los datos codificados (casos) en un archivo permanente

Dicho de otro modo, guardar la matriz como documento de SPSS®, Excel, Minitab o equivalente, y por supuesto, darle un nombre que lo identifique.

Codificación con un programa de análisis estadístico

Pero hoy en día los investigadores ya no lo hacen de la manera descrita, sino que la codificación la efectúan directamente, transfiriendo los valores registrados en los instrumentos aplicados (cuestionarios, escalas de actitudes, pruebas de laboratorio o equivalentes) a un archivo/matriz de un programa computarizado de análisis estadístico (IBM SPSS®, Minitab o equivalente). O bien, si no se cuenta con el programa, los datos se capturan en un documento de Excel (matriz) y luego se trasladan a un archivo del programa de análisis. Veamos el proceso en IBM SPSS®, pero antes es necesario hacer algunas aclaraciones:

- Se abre el programa IBM SPSS®, como cualquier otro, y si se trata de un archivo existente con los datos codificados (matriz completada), pues lo abrimos y hacemos los análisis pertinentes. Si vamos a crear un nuevo archivo o base de datos, elegimos la opción: “Archivo” → “nuevo” → “datos” y comenzamos a ejercer tal función.
- IBM SPSS® y programas equivalentes tienen dos matrices o ventanas: *a) vista de las variables* (variable view) y *b) vista de los datos* (data view). Ambas aparecen como pestañas (simulando carpetas o folders) ubicadas en la parte inferior de la pantalla hacia nuestro lado izquierdo.

Vista de datos	Vista de variables
----------------	--------------------

- La “vista de variables” representa el sistema de codificación o libro o documento de códigos electrónico (constituye una matriz). Los renglones o filas significan ítems, indicadores o reactivos, y las columnas representan características, propiedades o atributos de cada ítem. A los ítems o indicadores en estos programas se les denomina “variables” de la matriz, a veces coinciden con el concepto de variable que se tiene en la investigación (por ejemplo, género, es una variable de la investigación y un renglón o fila en la “vista de variables”) y en otras ocasiones son simplemente un ítem de una variable del estudio.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columna	Alineación	Medida	Rol
Ítem 1 o variable 1 de la matriz											
Ítem 2 o variable 2 de la matriz											
Ítem k...											

Como dijimos, las columnas son propiedades del ítem o indicador que debemos definir:

1. Nombre de cada ítem o variable de la matriz: lo asignamos nosotros (obviamente debe reflejar al ítem o reactivo al que hace referencia). Por ejemplo: género, edad, p1 (pregunta uno), ingresos, pas (presión arterial sistólica), etcétera.



● **Figura 9.20** Ejemplo de la vista de los datos en IBM SPSS® (final de la tierra).

	Cuantitativo1	Duración	Características	Importancia	Selección	Titulación	Retenencia	Raza	Operación	Importancia2	Importancia3
1	3	16.1	3.2	1	2	4	1	3	3.00		
2	5	12.2	2.2	1	1	3	3	3	2.50		
3	3	11.3	2.3	2	2	1	1	3	2.50		
4	3	18.2	5.2	2	1	1	1	3	5.00		
5	5	8.2	3.1	1	1	2	1	3	5.00		
6	3	1.2	3.1	1	1	2	1	3	3.00		
7	3	3.2	4.1	2	1	2	2	3	3.00		
8	3	6.3	4.4	3	2	3	1	4	4.00		
9	3	7.4	4.5	4	2	4	2	4	4.00		
10	3	8.2	5.6	3	2	2	2	4	4.50		
11	3	1.2	3.5	3	2	1	2	4	3.00		
12	3	12.2	2.6	3	2	2	2	5	3.50		
13	3	13.1	1.4	3	1	1	1	5	3.00		
14	3	2.3	3.3	2	2	2	1	5	4.00		
15	3	1.4	2.2	2	2	1	1	5	4.00		
16	3	8.6	4.2	1	3	2	1	5	3.00		
17	3	24.2	6.2	1	4	2	1	5	4.00		
18	3	12.1	2.1	1	4	2	3	5	2.00		
19	3	2.2	3.1	1	2	3	2	5	3.00		
20	3	9.2	1.1	1	3	1	1	5	3.00		
21	3	8.2	2.1	2	3	2	1	5	2.00		
22	3	8.3	4.2	2	1	1	1	5	3.00		
23	1	7.4	4.3	3	2	2	1	5	3.00		
24	3	6.2	5.2	2	2	2	1	4	4.50		
25	5	9.2	3.1	1	1	3	1	5	4.00		
26	5	7.2	5.4	3	1	4	1	5	5.00		
27	1	6.1	5.1	2	1	2	2	4	4.50		
28	5	6.3	4.3	1	3	1	3	4	4.00		

Errores de codificación

Al teclear los valores en la vista de los datos, se pueden cometer errores, es humano. Por ejemplo, que en un ítem o variable de la matriz donde solamente se tenían dos categorías, aparezca en uno o más casos una no contemplada (imaginemos que tenemos el ítem *género* con las opciones: 1 = masculino y 2 = femenino, y alguien teclea un “3” o un “8”, esto es un error de codificación; o bien que en una escala con tres opciones: 1. en desacuerdo, 2. neutral y 3. de acuerdo, se presenten valores como: “w”, “#”, ¿qué es eso?). Los errores de codificación tienen que corregirse. Esto puede hacerse:

- Revisando físicamente la vista de los datos y haciendo los cambios pertinentes.
- En IBM SPSS® con la función “ordenar o clasificar casos” —*sort*— (en “Datos” o “Data”) y, de este modo, visualizar valores que no correspondan a cada variable o ítem de la matriz.
- Ejecutando el “análisis de frecuencias” en el menú “Analizar”, “Estadísticos descriptivos” y “Frecuencias”, y una vez obtenidos los resultados, se observará en qué variables de la matriz (columnas) hay valores que no deberían estar, para efectuar las correcciones necesarias.³⁴

Cabe señalar que los *valores perdidos* no son errores de codificación, porque al registrarlos como tales, le estamos informando al programa que son precisamente valores perdidos y podemos o no excluirlos del análisis de frecuencias (se incluyen si queremos saber cuántos no contestaron o lo hicieron incorrectamente). De los análisis inferenciales se descartan automáticamente.

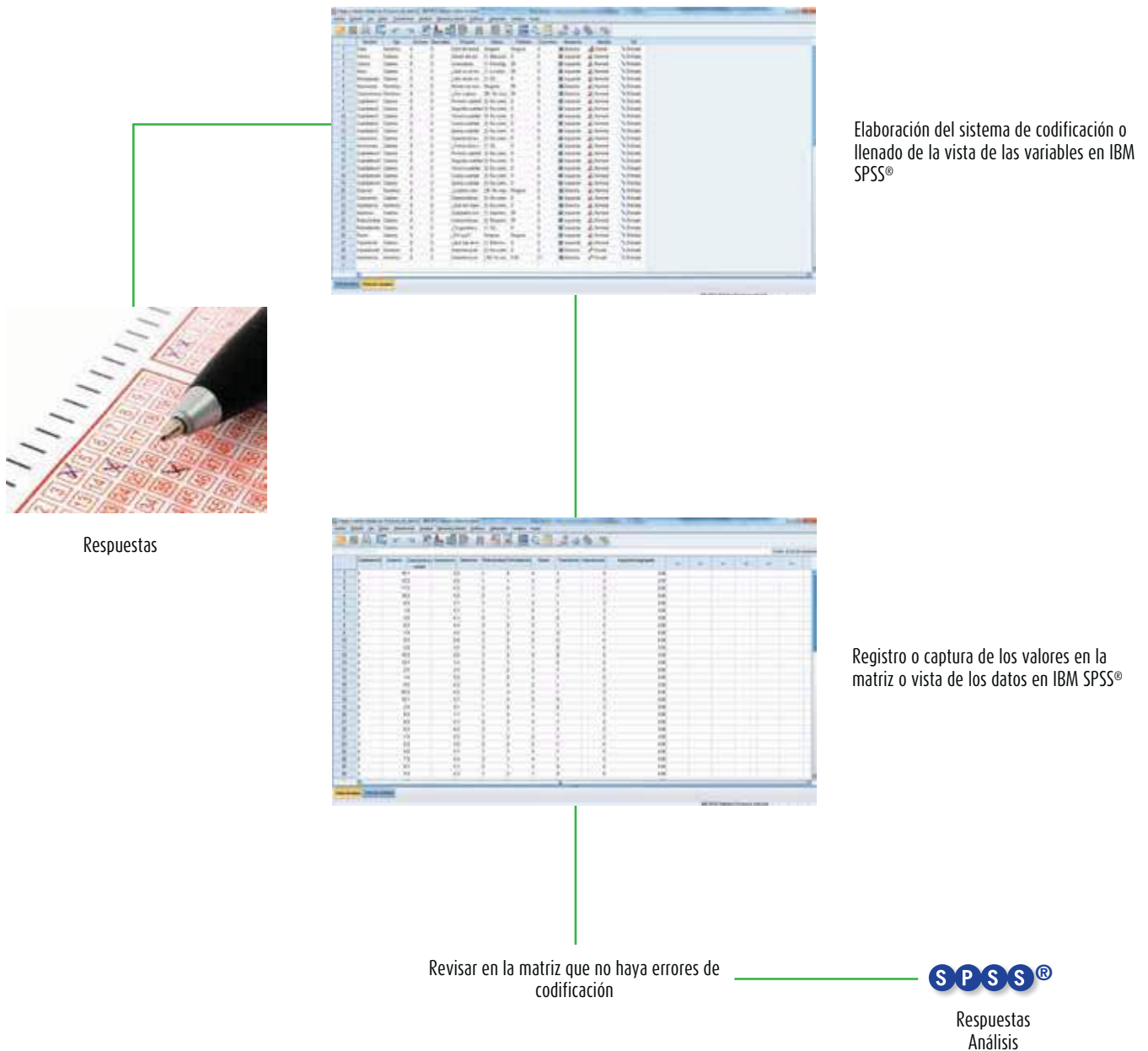
En ambas vistas se muestran las opciones para ejecutar las funciones de IBM SPSS®, como por ejemplo: analizar datos y elaborar gráficas, las cuales se comentarán en el siguiente capítulo, el 10: “Análisis de los datos cuantitativos” y con mayor profundidad en: Material complementario → Manuales → Manual “Introducción al IBM SPSS®” del centro de recursos en línea. Este manual lo llevará por el proceso. Además, hay cientos de páginas sobre este paquete y la de la propia empresa (<http://www-01.ibm.com/software/mx/analytics/spss/>).³⁵

En resumen, el proceso sería el que se muestra en la figura 9.21.

³⁴ Desde luego, como SPSS® se actualiza permanentemente los comandos pueden variar, mas no las funciones.

³⁵ Asimismo, busque en su país o región al representante de IBM SPSS® Inc.

● **Figura 9.21** Proceso de codificación y preparación de los datos para su análisis en IBM SPSS®.



Si queremos capturar los datos en nuestra PC, tableta u otro dispositivo electrónico personal y no disponemos de IBM SPSS® (solamente en nuestra universidad, centro de cómputo público o empresa), podemos hacerlo en una matriz de Excel y luego copiarlos y pegarlos en la vista de los datos de IBM SPSS®. Pues las columnas (A, B, C, D, etc.) corresponden a las variables de la matriz, y los renglones o filas, son los casos, al igual que en IBM SPSS®.

Como todo archivo, debe guardarse y respaldarse, implica nuestros datos y el sistema de codificación.

Cuando se utilizan dispositivos electrónicos para capturar los datos (como tabletas, terminales remotas, lectores ópticos, cuestionarios electrónicos, palms, etc.), obviamente no requerimos teclear los datos, éstos pasan directamente de la fuente a la matriz o base de datos.

En este capítulo, por razones didácticas, se presentaron matrices pequeñas, pero en la investigación pueden tenerse el número de columnas que se requieran.

Resumen



- Recolectar los datos implica: a) seleccionar uno o varios métodos o instrumentos disponibles, adaptarlos o desarrollarlos, esto depende del enfoque que tenga el estudio, así como del planteamiento del problema y de los alcances de la investigación; b) aplicar el o los instrumentos, y c) preparar las mediciones obtenidas o los datos recolectados para analizarlos correctamente.
- En el enfoque cuantitativo, recolectar los datos es equivalente a medir.
- Medir es el proceso de vincular conceptos abstractos con indicadores empíricos, mediante clasificación o cuantificación.
- En toda investigación cuantitativa medimos las variables contenidas en la(s) hipótesis.
- Cualquier instrumento de recolección de datos debe cubrir tres requisitos: confiabilidad, validez y “objetividad”.
- La confiabilidad se refiere al grado en que la aplicación repetida de un instrumento de medición, a los mismos individuos u objetos, produce resultados iguales.
- La validez se refiere al grado en que un instrumento de medición mide realmente las variables que pretende medir.
- Se pueden aportar tres tipos principales de evidencia para la validez cuantitativa: evidencia relacionada con el contenido, evidencia relacionada con el criterio y evidencia relacionada con el constructo.
- Los factores que principalmente pueden afectar la validez son: la improvisación, utilizar instrumentos desarrollados en el extranjero y que no han sido validados para nuestro contexto, poca o nula empatía con los participantes y los factores de aplicación.
- No hay medición perfecta, pero el error de medición debe reducirse a límites tolerables.
- La confiabilidad cuantitativa se determina al calcular el coeficiente de fiabilidad.
- Los coeficientes de fiabilidad cuantitativa varían entre 0 y 1 (0 = nula confiabilidad, 1 = total confiabilidad).
- Los métodos más conocidos para calcular la confiabilidad son: a) medida de estabilidad, b) formas alternas, c) mitades partidas y d) consistencia interna.
- La evidencia sobre la validez de contenido se obtiene al contrastar el universo de ítems frente a los ítems presentes en el instrumento de medición.
- La evidencia sobre la validez de criterio se obtiene al comparar los resultados de la aplicación del instrumento de medición frente a los resultados de un criterio externo.
- La evidencia sobre la validez de constructo se puede determinar mediante el análisis de factores y al verificar la teoría subyacente.
- Los pasos para elaborar un instrumento de medición son:
 1. Redefiniciones fundamentales sobre propósitos, definiciones operacionales y participantes.
 2. Revisar la literatura, particularmente la enfocada en los instrumentos utilizados para medir las variables de interés.
 3. Identificar el conjunto o dominio de conceptos o variables a medir e indicadores de cada variable.
 4. Tomar decisiones en cuanto a: tipo y formato; utilizar uno existente, adaptarlo o construir uno nuevo, así como el contexto de administración.
 5. Construir el instrumento.
 6. Aplicar la prueba piloto (para calcular la confiabilidad y validez iniciales).
 7. Desarrollar su versión definitiva.
 8. Entrenar al personal que va a administrarlo.
 9. Obtener autorizaciones para aplicarlo.
 10. Administrar el instrumento.
- Posteriormente, se preparan los datos para el análisis.
- En la investigación disponemos de diversos instrumentos de medición.
 1.  Escalas de actitudes: de Likert, diferencial semántico y escalograma de Guttman (este último se encuentra comentado en el capítulo 7 adicional que puede descargarse del centro de recursos en línea).
 2. Cuestionarios (autoadministrado, por entrevista personal, por entrevista telefónica, internet y por correo).
 3. Recolección y análisis de contenido cuantitativo (capítulo 7 adicional).
 4. Observación cuantitativa (capítulo 7 adicional).
 5. Pruebas estandarizadas (capítulo 7 adicional).
 6.  Archivos y otras formas de medición (capítulo 7 adicional).
- Las respuestas a un instrumento de medición se codifican.
- Actualmente, la codificación se efectúa transfiriendo los valores registrados en los instrumentos aplicados (cuestionarios, escalas de actitudes o equivalentes) a un archivo o matriz de un programa computarizado de análisis estadístico (IBM SPSS®, Minitab o equivalente).
- En la tabla 9.10 se resumen algunos instrumentos tratados en el capítulo.

► **Tabla 9.10** Concentrado de instrumentos para la recolección de datos (véase más en el centro de recursos en línea)

Métodos	Propósito general básico	Ventajas	Retos
Cuestionarios/ Escala de actitudes/ Pruebas estandarizadas	• Obtener de manera relativamente rápida datos sobre las variables. • Propios para actitudes, expectativas, opiniones y variables que pueden medirse mediante expresiones escritas o que el mismo participante puede ubicarse en las categorías de las variables (autobúsqueda).	• Puede ser anónimo. • Poco costosa su aplicación individual. • Relativamente fácil de responder. • Relativamente fácil de analizar y comparar. • Puede administrarse a un considerable número de personas. • Normalmente disponemos de versiones previas para escoger o basarnos en éstas.	• Regularmente no se obtiene realimentación detallada de parte de los respondientes. • Se evalúan actitudes y proyecciones, no comportamientos (mediciones indirectas). • El manejo del lenguaje puede ser una fuente de sesgos e influir en las respuestas. • Son impersonales. • No nos proporcionan información sobre el individuo, excepto en las variables medidas.
Observación	• Recolectar información no obstrusiva respecto a conductas y procesos.	• Se puede adaptar a los eventos tal y como ocurren. • Se evalúan hechos, comportamientos y no mediciones indirectas.	• Dificultad para interpretar conductas. • Complejidad al categorizar las conductas observadas. • Puede ser obstrusiva y provocar sesgos si es "participante". • Puede ser costosa.
Análisis de contenido	• Recolectar información no obstrusiva respecto de mensajes.	• Se puede adaptar a los eventos tal como ocurren. • Se evalúan mediciones indirectas.	• Dificultad para interpretar mensajes. • Complejidad al categorizar los mensajes.

Conceptos básicos



- Análisis cuantitativo de contenido (capítulo 7 del centro de recursos en línea)
- Archivo de datos
- Categorías
- Codificación
- Codificador
- Coeficiente alfa de Cronbach
- Coeficiente de confiabilidad
- Confiabilidad
- Contexto de administración del instrumento
- Cuestionarios
- Diferencial semántico
- Escala de Likert
- Escalas de actitudes
- Escalograma de Guttman (capítulo 7 del centro de recursos en línea)
- Entrevista
- Evidencia relacionada con el constructo
- Evidencia relacionada con el contenido
- Evidencia relacionada con el criterio
- Instrumento de medición
- Matriz de datos
- Medición
- Medida de estabilidad
- Método de formas alternas
- Método de mitades partidas
- Niveles de medición
- Observación cuantitativa (capítulo 7 del centro de recursos en línea)
- Pruebas estandarizadas (capítulo 7 del centro de recursos en línea)
- Pruebas proyectivas (capítulo 7 del centro de recursos en línea)
- Recolección de datos
- Unidad de análisis
- Validez
- Vista de las variables
- Vista de los datos

Ejercicios



1. Busque una investigación cuantitativa en algún artículo de una revista científica, en la cual se incluya información sobre la confiabilidad y la validez del instrumento de medición. ¿El instrumento es confiable? ¿Qué tan confiable?
2. Responda y explique con ejemplos la diferencia entre confiabilidad y validez.

¿Qué técnica se utilizó para determinar la confiabilidad? ¿Es válido? ¿Cómo se determinó la validez?

- 

3. Defina ocho variables e indique su nivel de medición.

4. Defina una variable de cada nivel de medición.

5. Suponga que alguien intenta evaluar la actitud hacia el presidente de la República. Entonces construya un cuestionario tipo Likert con 20 ítems para medir dicha actitud e indique cómo se calificaría la escala total (10 ítems positivos y 10 negativos). Por último, señale la dimensión que cada ítem pretende medir de dicha actitud (credibilidad, presencia física, etcétera).

6. Construya un cuestionario para medir la variable que considere conveniente (con un mínimo de 10 preguntas o ítems) e incluya preguntas demográficas adicionales. Aplíquelo a 20 conocidos suyos; elabore el libro de códigos y la matriz de datos, mínimo en Excel.

7. ¿Cómo mediría la variable “hostilidad” mediante observación y cómo por medio de una escala de actitudes? (Debe leer el apartado de observación en el capítulo 7 adicional del centro de recursos en línea).
8. Genere un planteamiento del problema en el que utilice por lo menos dos tipos de instrumentos cuantitativos para recolectar datos.

9. Construya una matriz de datos en Excel, Minitab o SPSS (o una tabla en Word) con 15 casos y las siguientes variables: género, edad, deporte preferido para practicar, deporte preferido para observar, escuela de procedencia (pública/privada), tipo de música que más le agrada, si está o no de acuerdo con la política económica del gobierno actual, partido por el que votó en la últimas elecciones municipales y líder histórico que más admira. Que los casos (participantes) sean sus compañeros de clase (el ejercicio implica levantar datos y codificarlos, desde luego).

10. Elabore uno o varios instrumentos para el ejemplo de estudio que ha desarrollado hasta ahora en el proceso cuantitativo (incluya la codificación).

Ejemplos desarrollados³⁶

La relación entre la personalidad y las enfermedades

Como vimos en el capítulo 6 (“Formulación de hipótesis”), la personalidad se midió a través de una serie de escalas (véase la tabla 9.11).

Las variables “enfermedades padecidas en el presente y el pasado”, “factores de riesgo a la salud”, “hábitos de vida” y las demográficas fueron medidas a través de un cuestionario de autorreporte.

Tabla 9.11 Medición de la personalidad de Yousfi *et al.* (2004)

Dimensiones	Prueba o escala	Número de ítems y confiabilidad (Alfa-Cronbach)
Factor R	Escalas del Factor R de Grossarth-Maticek:	
	Inhibición de la expresión centrada en uno mismo (disposición para el cáncer)	28 ítems ($\alpha = .90$)
	Barreras a la expresión centrada en uno mismo (disposición para las enfermedades cardíacas)	25 ítems ($\alpha = .92$)
	Psicopatología	20 ítems ($\alpha = .83$)
	Salud/autonomía	31 ítems ($\alpha = .88$)
	Racionalismo	23 ítems ($\alpha = .91$)
	Tendencias antisociales	13 ítems ($\alpha = .76$)
Optimismo	Prueba de la Orientación de Vida: LOT	8 ítems ($\alpha = .55$)

(continúa)

³⁶ Por cuestiones de espacio, se comentan brevemente. El primer ejemplo presenta las mediciones hechas (pruebas o escalas aplicadas, con su respectiva confiabilidad) y los contextos de administración. El segundo ejemplo aborda un aspecto de la recolección: el procedimiento y las variables centrales; el tercero, el instrumento de medición; y el cuarto, la consideración y comparación de instrumentos (escalas).

► **Tabla 6.4** (continuación)

Dimensiones	Prueba o escala	Número de ítems y confiabilidad (Alfa-Cronbach)
<i>Locus (centro) del control sobre enfermedades: interno y externo.</i>	Cuestionario FEGK	16 ítems para el locus interno y 13 para el externo ($\alpha = .79$ y $.81$, respectivamente)
<i>Urgencia de tiempo y activación perpetua</i>	Escala TUPA e Inventario de Expresión para Tratar el Estado de Ira	13 ítems para TUPA ($\alpha = .65$) y 24 para el inventario con tres subescalas de ira interna, ira externa y control de la ira ($\alpha = .82$, $.83$ y $.81$ respectivamente)
<i>Soporte social</i>	Escala SOZU K-22	22 ítems ($\alpha = .91$)
<i>Depresión</i>	Escala de von Zerssen	16 ítems ($\alpha = .85$)
<i>Sentido de coherencia</i>	Escala SOC-HD.	19 ítems ($\alpha = .80$)
<i>Hostilidad (factores: agresión, irritabilidad y envidia)</i>	Cuestionario de Agresión y Hostilidad de Saltz-Epstein Inventario de Irritabilidad y Envidia de Buss-Durkee	Agresión: 8 ítems ($\alpha = .62$) Hostilidad: 11 ítems ($\alpha = .66$) 8 ítems ($\alpha = .63$)
<i>Psicoticismo (extraversión, neuroticismo y deseabilidad social)</i>	Inventario de la Personalidad de Eysenck (EPI)	57 ítems en total con $\alpha = .69$, $.85$ y $.57$ respectivamente
<i>Control social exagerado</i>	Escala de Modo de Vida —WOLS—	21 ítems ($\alpha = .76$)

Respondieron 2 047 participantes a los instrumentos en grupos de 20 (autoadministrados), con asistencia para dudas y comentarios (en las instalaciones de la universidad). A otros 2 306 que habían expresado su consentimiento, se les enviaron los instrumentos por correo y se recibieron por este medio. Al resto, también se les mandaron por correo pero sin preguntarles previamente si deseaban participar (solicitud única). Es decir, en todos los casos el contexto fue “autoadministrado”.

No se encontraron diferencias significativas en las respuestas entre los diferentes contextos de autoadministración (en grupos, por correo y consentimiento previo y por correo sin autorización anterior).

La televisión y el niño

Se aplicó un cuestionario en una muestra total de 2 112 niños y niñas de la Ciudad de México, de acuerdo con la estrategia de muestreo planteado. Las variables medidas fueron: uso de medios de comunicación colectiva, tiempo de exposición a la televisión, preferencia de contenidos televisivos (programas), bloques de horarios de exposición a la televisión (mañana, tarde o noche), comparación de la televisión con otras fuentes de entretenimiento, actividades que realiza mientras observa la televisión, condiciones de exposición a la televisión (solo-acompañado), autonomía en la elección de los programas, control de los padres sobre la actividad de ver televisión, usos y gratificaciones de la televisión, y datos demográficos.

El cuestionario es descriptivo y fue evaluado y validado por 10 expertos en investigación sobre la relación niño-televisión. Se elaboró una versión piloto con 100 participantes (50 niñas y 50 niños), la cual se probó y ajustó. No hubo escalas con varios ítems, por lo que no se calculan coeficientes de confiabilidad.

La pareja y relación ideales

Se desarrolló un instrumento inicial para recolectar los datos con base en preguntas. Por ejemplo: “Pensando en tu relación ideal, ¿cuáles serían las características que más te gustaría que tuviera esa relación?”, o bien: “¿Qué cualidades te gustaría que tuviera tu novio(a) ideal?”. Sin embargo, en la prueba piloto con 100 estudiantes se observó que era mejor sustituir las preguntas por afirmaciones que fueran completadas por los participantes (por ejemplo: “Pensando en tu relación ideal, las características que más te gustaría que tuviera esa relación serían...”). Resultaban más comprensibles para ellos y fueron respondidas con mayor precisión. Así, se aplicó por entrevista el siguiente cuestionario:³⁷

Cuestionario sobre la pareja y relación ideales

El objetivo de esta encuesta es conocer tu opinión acerca de las relaciones y parejas sentimentales que has tenido, tienes y tendrás, así como de tu concepción de una pareja ideal, por lo cual te agradeceríamos contestar lo que se te indica a continuación, pensando y contestando según corresponda en cada caso. Recuerda que tus respuestas son totalmente confidenciales.

³⁷ Las opciones de respuesta también se han reducido por espacio (por ejemplo: carreras).

Edad: _____ años

Género: 1) Masculino _____ 2) Femenino _____

Indica la carrera que estudias actualmente:

Psicología ()	Turismo ()	Comunicación ()
Medicina ()	Administración ()	Mercadotecnia ()
Arquitectura ()	Contaduría Pública ()	Derecho ()
Ing. Industrial ()	Ing. en Sistemas Computacionales ()	Comercio Internacional ()
Otra (mencionar) _____		

1. Para ti, ¿qué es un novio(a)? Un novio(a) es: _____

2. Para ti, ¿qué es una relación de noviazgo? Un noviazgo es: _____

Pasado:

3. ¿Has tenido novio(a)? 1) Sí _____ 2) No _____

4. ¿Con cuántos(as) novios(as) has durado más de un mes? _____

5. Las cualidades que más te gustaban del novio(a) más importante que has tenido en el pasado son:

Anota de la más importante (1) a la menos importante (5).

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

6. Pensando en tu relación pasada más importante, las características que más te gustaban de la relación de pareja eran (*no hablamos de tu pareja, sino de la relación de noviazgo*):*Anota de la más importante (1) a la menos importante (5).*

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

Actualmente:

7. ¿Tienes novio(a)? 1) Sí _____ 2) No _____

8. Las cualidades que más te gustan de tu novio(a) son:
Anota de la más importante (1) a la menos importante (5).

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

9. ¿Cuántos meses llevas con tu novio(a) actual? _____

10. Pensando en tu relación actual, las características que más te gustan de la relación de pareja (*no hablamos de tu pareja, sino de la relación de noviazgo*):*Anota de la más importante (1) a la menos importante (5).*

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

11. ¿Qué tan importante es en tu vida tu pareja actual?

5) Sumamente importante 4) Muy importante 3) Importante 2) Poco importante 1) No tiene importancia

Ideal:

12. Piensa en tu novio(a) ideal y menciona las cualidades que te gustaría que tuviera:

Anota de la más importante (1) a la menos importante (5).

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

13. Pensando en tu relación ideal, las características que más te gustaría que tuviera esa relación (*no hablamos de tu pareja, sino de la relación de noviazgo*) serían:*Anota de la más importante (1) a la menos importante (5).*

1) _____, 2) _____, 3) _____,

4) _____, 5) _____

Futuro:

14. En tu futuro, ¿te gustaría o no tener una relación de pareja para toda la vida?

1. Sí _____ 2. No _____ 3. No sé _____

15. ¿Por qué? _____

16. En tu futuro, ¿qué tipo de relación de pareja duradera a largo plazo te gustaría establecer, tener o formar?

(Marca una opción)

1. Matrimonio civil

2. Matrimonio religioso

3. Matrimonio religioso y civil

4. Unión libre (vivir juntos sin estar casados)

5. Llevar una relación de pareja sin vivir juntos

6. Otra: _____

Gracias por tu colaboración.

El abuso sexual infantil**Escala cognitiva**

El instrumento Children's Knowledge of Abuse Questionnaire-Revised (CKAQ-R), fue traducido al español y adaptado para preescolares. En esta escala modificada, se eliminaron los elementos redundantes y los que evaluaban las actitudes ante los desconocidos, con la tesis que quienes agreden sexualmente a los menores son en su gran mayoría personas cercanas. Además, se simplificaron las preguntas formuladas negativamente, como "¿algunas veces está bien no hacer lo que nos pide un adulto?", que tienden a ser confusas para los preescolares. El CKAQ-Español puede tener un puntaje máximo de 22 y cada reactivo posee evaluación dicotómica, dando un punto por cada respuesta correcta. Sigue el mismo esquema y protocolo que el CKAQ original. Cada pregunta puede ser contestada como "sí", "no", o "no sé" y su evaluación es dicotómica (correcto o incorrecto). Incluye cuestiones para medir el desarrollo cognitivo y actitudes asertivas ante contactos positivos y negativos, chantaje emocional, disociación de los contactos con la afectividad y pedir ayuda ante el abuso.

El estudio de la confiabilidad interna se efectuó con el modelo Kurder Richardson 20 (KR-20), en la versión adaptada de Cron-

banch para reactivos dicotómicos. Tal estudio se realizó con el total de casos ($n = 150$). Se obtuvo un alfa de 0.69, lo que representa un nivel moderadamente aceptable.

Escala conductual

Después del estudio de diversas escalas, se decidió partir del RPP para el desarrollo del instrumento conductual. Las razones de esta decisión se basaron en que el RPP se ha aplicado a muestras grandes ($n = 670$). Por otro lado, evalúa en acción los patrones seguidos por los agresores, así da la oportunidad de analizar las reacciones de los niños y sus habilidades de protección “en vivo”. Además, no aborda al niño o niña de manera burda o aterradora, se enfoca en los preámbulos del abuso, en donde se censa la posibilidad. Por estas razones, este instrumento nos parece de los más acertados por su evaluación conductual, su aproximación a lo que un(a) niño(a) puede vivir en su cotidianidad en cuanto a sus acercamientos incómodos y evaluar sus recursos asertivos, seguridad emocional y habilidades de autoprotección.

Uno de los inconvenientes de este protocolo es que no se disponen de valores psicométricos que lo avalen. Por lo que no hay comparativos para los resultados que de esta investigación se obtengan.

Partiendo del RPP original, se le hizo una adaptación mediante la traducción y adecuación al contexto mexicano. A esta escala le llamaremos Role Play-México. Uno de los inconvenientes que se le cuestiona al RPP es que sólo puede aplicarse uno a uno. Es decir, no se puede aplicar a grupos de infantes en conjunto. Sin embargo, en el caso de preescolares esto no aplica, porque en general las pruebas administradas a grupos, requieren del desarrollo de las habilidades lectoescritoras, un estado no dominado en la etapa preescolar. Por tanto, tal inconveniente es intrascendente en el caso de estudio. Otra desventaja que se le atribuye, es que la escala no incluye elementos que evalúen la actitud de los menores ante los contactos positivos, para determinar si los PPASI (programas de prevención del abuso sexual infantil) generan un efecto nocivo de suspicacia indiscriminada ante cualquier contacto. Por lo cual, se decidió incluir un par de reactivos para evaluar esta posibilidad en el Role Play-México (RP-México). Estos reactivos incluyen por ejemplo, abrazos por los padres o felicitaciones. Se desarrolló, también, una prueba paralela a dicha adaptación, a la que llamamos Evaluación de la Prevención del Abuso (EPA).

En la escala RPP se tiene un puntaje máximo de 14 puntos, evalúa la negación verbal y no verbal de seis escenas “en vivo”. Es decir, donde el evaluador actúa y se le pide al niño que responda a la pregunta: ¿qué diría y haría? en una situación planteada. Además, en los tres reactivos donde se aborda el chantaje emocional y la coerción, se otorga un punto extra si el participante muestra intención de denunciar el evento. En el caso de la evaluación del RP-México y de la EPA se considera un total de ocho escenarios “en vivo”, seis de tipo abusivo y dos de contactos no abusivos. El puntaje de ambas escalas (RP-México y EPA) tiene un máximo de 40 puntos. Al igual que el RPP, evalúa la asertividad verbal y conductual, pero se amplía la evaluación con la intención de denuncia del evento abusivo, cubriendo la necesidad de mejorar el sistema de medición con la persistencia de los infantes de pedir ayuda hasta obtenerla. Mide además las siguientes subescalas: 1) reconocimiento de contactos, tanto positivos como negativos, y 2) las habilidades de asertividad verbal (*qué decir*), no verbal (*qué hacer*) y la persistencia en la intención de denuncia ante algún incidente abusivo (*denuncia*). Los 40 puntos se derivan de la suma de un punto por cada acierto en la asertividad verbal (ocho máximo), un punto por cada asertividad conductual (máximo ocho), un punto por cada intención de denuncia de los contactos inapropiados (seis máximo) y un punto por cada persona a quién denunciarían, hasta un máximo de tres por cada escenario de contacto inapropiado (18 puntos máximo).

Se desarrolló el análisis de confiabilidad tanto temporal como interno. Se aplicó *test-retest* de acuerdo con un método de formas paralelas al administrarse el RP-México y el EPA. La correlación entre ambas pruebas alcanzó un buen nivel y fue significativo ($r = 0.75$, $p < 0.01$), lo que avala la utilización de estos instrumentos de forma paralela. El *test-retest* se aplicó en un subgrupo ($n = 44$) del grupo de control ($n = 79$). Este estudio confirma que hay correlación *test* y *retest* entre cada instrumento RPP, RP-México y en todas las subescalas, los índices van de 0.59 a 0.78, todas con $p < 0.01$. El instrumento RP-México tiene una correlación ($r = 0.75$) equivalente a la reportada en otros instrumentos similares (WIST, PSQ). Este índice muestra un grado de estabilidad temporal aceptable, dado el tamaño de la muestra.

Los investigadores opinan

Recolección de datos cuantitativos

Dentro del modelo de investigación cuantitativa, la etapa de recolección de los datos resulta de vital importancia para el estudio, de ella dependen tanto la validez interna como externa.

La validez interna de una investigación depende de una adecuada selección o construcción del instrumento con el cual se va a recolectar la información deseada, la teoría que enmarca el estudio tiene que conjugar perfectamente con las características teóricas y empíricas del instrumento; si esto no ocurre, se corre el

riesgo de recolectar datos que a la postre pueden ser imposibles de ser interpretados o discutidos, la teoría y los datos pueden caminar por distintas direcciones. Un ejemplo muy sencillo para graficar este problema sería hacer hipótesis y teorizar en torno a la personalidad sobre la base de una de las teorías de los rasgos y usar un instrumento proyectivo para recolectar los datos. Lo correcto sería que la misma teoría sustente los planteamientos hipotéticos y teóricos, así como fundamente el instrumento. Si bien el ejemplo puede resultar un tanto simple y grosero, en el



nivel de las investigaciones de pregrado, este problema resulta bastante común y le es muy difícil manejarlo al estudiante promedio.

Del mismo modo, la recolección de los datos se relaciona con la validez externa del estudio, por cuanto la generalización depende de la calidad y cantidad de los datos que recolectamos. Por ello, en estudios cuantitativos resulta importante determinar una muestra adecuada, que tenga representatividad en el tamaño y que a la vez refleje la misma estructura existente en la población. Sin una buena muestra de datos, no se puede generalizar; y si se corre este riesgo, el investigador podría llevar sus conclusiones más allá de la realidad, cuando lo que se desea es reflejar la realidad.

Una idea clave, para no tropezar con asuntos insalvables en este momento de la investigación o para no tomar decisiones que conduzcan al error, es hacer un buen proyecto de investigación. En la etapa de la planificación debe quedar claramente establecido y justificado qué instrumento se va a utilizar; cómo, dónde y a quiénes se les aplicará; qué instrucciones se les va a brindar a los sujetos o participantes; qué datos son los que se someterán a tratamiento y cuáles otros no serán tomados en cuenta; cómo se van a tratar los mismos y cómo se llegará desde los datos a la teoría.

EDWIN SALUSTIO SALAS BLAS

Universidad de Lima

Perú