

**Construcción de Cuestionarios y Escalas:  
El Proceso de la Producción de Información Cuantitativa**  
Dr © Rodrigo Asún Inostroza

**1.- Las características de la tarea de construcción de instrumentos cuantitativos**

*"Lo que nos piden no es imposible,  
Pero está tan cerca que lo parece"*  
Robert Heinlein

Sólo es posible hacer investigación cuantitativa si se dispone de información válida registrada en forma numérica. Esta primera afirmación, que es verdaderamente muy obvia, tiene dos corolarios que son un poco más interesantes y sirven de punto de partida para este capítulo.

- A pesar que actualmente tanto las técnicas de análisis de información, como los sistemas de procesamiento de la información, permitirían realizar análisis cuantitativos de muchos fenómenos sociales, en una parte importante de éstos no estamos en condiciones de utilizar esta herramienta de indagación producto de que no disponemos de registros cuantitativos ni somos capaces de construirlos. Esta es, por ejemplo, nuestra principal limitación cuando intentamos hacer investigación cuantitativa de fenómenos históricos, ya que sobre muchos temas no existen las bases de datos que nos permitirían realizar nuestros análisis.
- La calidad de una investigación cuantitativa depende muy directamente de la calidad de los datos que se analicen, y a su vez, estos últimos dependen del instrumento con el que hayan sido registrados. Por ello, uno de los elementos críticos que determinan la calidad de una investigación cuantitativa es el instrumento con el cual se registra cuantitativamente una información.

Evidentemente, una gran parte de la realidad social no se nos aparece directamente en forma cuantitativa. Tenemos que "traducirla" al lenguaje numérico utilizando procedimientos de operacionalización, que incluyen la aplicación de un instrumento cuantitativo.

No obstante, este proceso de traducción es posible de realizar de tres maneras analíticamente muy diferentes.

En primer lugar, es posible que la información que el investigador social utilice para realizar su investigación no haya sido producida por él mismo. Por ejemplo es posible que un funcionario de un ministerio encargado de un sistema de consultas telefónicas registre diariamente en una "ficha" las preguntas que se le han hecho en relación con cierto tema. A partir de esta información un investigador social puede realizar una clasificación de las preguntas efectivamente formuladas en una serie simple de "tipos de preguntas" y asignarle un número a cada tipo. Luego, a partir de la lectura y asignación de un número a cada pregunta concreta es posible construir una base de datos susceptible de analizar cuantitativamente.

¿Qué características centrales tiene un proceso como es descrito?. Creemos que fundamentalmente tres: a) lo registrado usualmente es un hecho, tal como este ha sido percibido por quien lo registra, b) el objetivo inicial que ha motivado el registro de la información no ha sido realizar una investigación social, sino más bien llevar el control de un cierto proceso, c) si bien usualmente la información

registrada en la ficha no tiene forma cuantitativa, el proceso para obtener un registro numérico (llamado proceso de codificación), tiende a no ser complejo, ya que el instrumento de registro facilita que se describa el hecho de manera muy sintética y esquematizada.

Las investigaciones que utilizan esta estrategia son llamadas "investigaciones con datos secundarios", dando a entender que la información analizada no ha sido producida expresamente para la investigación ni por el investigador en persona. Este tipo de investigaciones tienen sus propias complejidades y potencialidades, las que no serán desarrolladas en este capítulo.

Lo que acá nos interesa es trabajar sobre las investigaciones con datos primarios, es decir, cuya información cuantitativa básica ha sido producida por el propio investigador. En este segundo caso nos encontramos con dos entradas posibles.

Es frecuente que el investigador (o sus ayudantes) registren los hechos que ellos mismos han observado en una determinada situación (registro que también se realiza a través de una ficha, llamada "de observación"), usualmente también en este caso se debe utilizar procesos de codificación, hasta llegar a obtener la base de datos definitiva.

En este segundo caso (llamadas "investigaciones observacionales") también estamos en presencia de un estudio sobre hechos (aunque no directamente, sino tal como son estos percibidos por los observadores). Este tipo de estudios y la forma de diseñar fichas de observación, tampoco será objeto de este capítulo.

Finalmente, una tercera entrada para realizar estudios cuantitativos es que el investigador produzca por si mismo la información que va a analizar, pero que su fuente de información no sean fenómenos percibidos por sus sentidos, sino las declaraciones que determinados sujetos hacen sobre el tema de estudio en una conversación directa o indirecta con el investigador (o sus ayudantes).

Ahora bien, con el fin de que esta conversación sea posible de analizar en forma cuantitativa, debe ser dirigida y registrada con un instrumento que permita sólo (o principalmente) que el sujeto entrevistado exprese opiniones en forma cuantitativa. Estos instrumentos son los llamados genéricamente "cuestionarios", los que pueden incluir escalas de actitud o aptitud en su interior.

Evidentemente, ya que lo que posteriormente se analizará será solo lo registrado en estos instrumentos, la calidad de este tipo de investigación cuantitativa dependerá crucialmente de la calidad del cuestionario. Si incluimos en el preguntas o respuestas inadecuadas, la información que obtendremos no responderá a los objetivos de estudio que nos hemos trazado.

No obstante lo anterior y a pesar de lo crítico que es el proceso de construcción de cuestionarios para la investigación cuantitativa, esta es una tarea que ha sido usualmente poco trabajada en la literatura. En general, las exposiciones sobre este tema se detienen en generalidades sobre la forma de redactar las preguntas, sin intentar explicar la complejidad y las etapas del proceso de construcción de cuestionarios (véanse por ejemplo Azofra, 1999 y Cea, 1998).

No es casualidad que la forma habitual de trabajar este tema en los textos de investigación cuantitativa sea a través de ejemplos, lo que indica que quizás una de las mejores maneras de aprender a construir cuestionarios es observar como lo han hecho otros. La paradoja es que una estrategia metodológica

basada en la esquematización analítica de los procedimientos (como lo es la metodología cuantitativa), tenga en su base una tarea que parece más un oficio o artesanía que un proceso ordenable en etapas.

Nuestra impresión es que precisamente debido a la dificultad que existe de esquematizar el proceso de construcción de cuestionarios, es que este proceso ha sido un poco dejado de lado por los investigadores cuantitativos, siempre reacios a profundizar en los temas menos estandarizables.

Además, no está claro el tipo de experticia necesaria para construir cuestionarios. Evidentemente, un investigador cuantitativo con experiencia en análisis de datos podrá ayudar, ya que más o menos sabe que tipo de preguntas resultan fáciles o difíciles de interpretar una vez obtenidas las respuestas de las personas o tiende a tener una idea del tipo de preguntas que generan que los respondientes se abstengan de responder, se molesten o no comprendan lo preguntado.

No obstante, ese tipo de conocimientos no bastan. Si el investigador no tiene idea sobre el tema tratado no sabrá que es importante preguntar (por ejemplo, si pretende estudiar el nivel de racismo de las personas y no conoce teorías modernas sobre el tema, difícilmente sabrá que es lo importante de preguntar hoy en día). Por ello, el constructor del cuestionario no sólo debe saber de metodología cuantitativa, sino también ser un experto teórico sobre el tema que trabajará.

Más aún, la construcción del cuestionario supone el tratar de traducir los conceptos teóricos del investigador en preguntas y respuestas comprensibles por las personas a las cuales se estudiará, por lo que conocer el lenguaje, cultura y formas de expresión de los sujetos investigados también es muy relevante.

En síntesis, se requieren entonces habilidades o experticias de orden metodológico, teórico y socio/antropológico. Con frecuencia no encontramos a una sola persona que cumpla con todas estas experticias, por tanto, se impone la necesidad de un trabajo en equipo.

Por otro lado, no debemos olvidar que el ejercicio de construcción de un cuestionario implica también un proceso de seleccionar que es lo que se va a preguntar del conjunto de preguntas y temas posibles de investigar. Este ejercicio de recorte de la realidad implica no sólo una actividad técnica, sino también una toma de posición ideológica (consciente o inconsciente) por parte del investigador, el que no sólo obtendrá información sobre su objeto de estudio, sino que "producirá" dicha información a partir de su propio punto de vista, contribuyendo a la construcción de una imagen social sobre sus sujetos y tema de investigación (ver como ejemplo: Muñoz, 2004).

Finalmente, y en gran parte como consecuencia de todo lo anterior, construir cuestionarios es una tarea clave para la investigación social cuantitativa, pero difícil e ingrata.

Difícil, porque implica compatibilizar muchas demandas divergentes, e ingrata, porque a las personas externas este oficio se les aparece como muy simple, de tal manera que casi cualquier persona que estudie un cuestionario será capaz de criticar ciertas preguntas y creará que lo puede hacer mejor.

Por ello, ¿qué recomendaciones generales se les puede hacer a quienes, valientemente, se apresten a enfrentar la tarea de construir un cuestionario en ciencias sociales?. En general, me parece de buen sentido seguir las siguientes recomendaciones:

- Privilegiar siempre el trabajo en equipo: un cuestionario siempre será mejor cuantas más personas contribuyan a su formulación. La existencia de diversas personas, cada una aportando una diferente experticia y punto de vista, sin duda enriquecerán el producto final.
- Utilizar como base instrumentos ya existentes: siempre es más fácil la crítica y perfeccionamiento de lo ya existente que la creación de algo nuevo. Por ello, la primera fase de la creación de un nuevo instrumento debiera ser el análisis de los cuestionarios ya existentes sobre el tema. Con un poco de suerte, algunas preguntas que se quiere realizar ya estarán construidas.
- No esperar perfección: este es uno de esos tipos de trabajos que siempre se pueden perfeccionar, por lo tanto, dados los recursos, profesionales y tiempo disponible, se debe intentar construir el mejor instrumento posible, sin esperar que el resultado sea totalmente perfecto. En toda encuesta siempre hay preguntas que no funcionan y que se dejan sin analizar, lo importante es que exista un buen núcleo de preguntas fiables, válidas e interesantes.

## 2.- ¿Qué es un cuestionario?

*"Si toda herramienta pudiera hacer el trabajo que quisiera,  
no serían necesarios ni aprendices para los maestros,  
ni siervos para los señores"*  
Aristóteles

Un cuestionario es un dispositivo de investigación cuantitativo consistente en un conjunto de preguntas que deben ser aplicadas a un sujeto (usualmente individual) en un orden determinado y frente a las cuales este sujeto puede responder adecuando sus respuestas a un espacio restringido o a una serie de respuestas que el mismo cuestionario ofrece. El cuestionario también incluye, en sí mismo o en una hoja de respuestas especial, espacios en donde es posible registrar las opciones de respuesta del sujeto. El objetivo general de un cuestionario es "medir" el grado o la forma en que los sujetos encuestados poseen determinadas variables o conceptos de interés (sus opiniones, creencias, conductas que recuerdan haber realizado, características demográficas, capacidades matemáticas, etc.).

A partir de esta definición podemos profundizar algunas de las principales características de los cuestionarios como dispositivos de investigación cuantitativos:

- Tal como queda claro en la definición, un cuestionario es un instrumento diseñado para dirigir una conversación no horizontal, en que uno de los participantes (encuestador) cumple el rol de formular las preguntas pre establecidas y presentar las opciones de respuesta, mientras que el sujeto encuestado sólo escoge sus respuestas en función de las posibilidades que se le presentan.
- Si bien en un cuestionario pueden incluirse muchos tipos de preguntas y respuestas, éstas últimas generalmente cumplen con la propiedad de estar pre-redactadas, quedándole al sujeto encuestado sólo la tarea de escoger la opción que desea dentro de las que se le presentan. No obstante, a modo de excepción, también es posible incluir en un cuestionario preguntas abiertas o semi abiertas, en que al sujeto se le permite expresar su opinión de manera más libre. No

obstante, aún en este caso, tanto la forma de redacción de las preguntas, como el espacio concedido en el cuestionario para registrar la respuesta es muy limitado y desincentiva la formulación de respuestas muy complejas.

- El sentido de presentar al sujeto respuestas pre establecidas es facilitar el proceso de "medición", ya que las respuestas simples y sobre todo las respuestas pre-establecidas, facilitan de gran manera la tarea de asignar números a las respuestas de las personas. De esta manera se cumple el principal objetivo de la aplicación de un cuestionario: medir la presencia de determinadas variables en las personas objeto de estudio.
- La conversación que se entabla al aplicar un cuestionario es tan estructurada que incluso es posible eliminar o alejar al encuestador del sujeto encuestado, cosa que ocurre en los cuestionarios que están diseñados para ser autoadministrados (desaparece el encuestador, sustituido por instrucciones escritas) o en las encuestas telefónicas, en que el encuestador y el encuestado están sólo virtualmente en el mismo espacio. Otra consecuencia de esta propiedad es que la tarea de encuestar se simplifica tanto que no requiere habilidades demasiado complejas (básicamente sólo se requiere saber leer y escribir y la capacidad de desarrollar empatía y confianza en el encuestado), por lo que su costo encuesta es bastante menor que el de las técnicas cualitativas.
- Ya que el objetivo de un cuestionario es "medir" el grado o la forma en que los entrevistados poseen determinadas variables, cada pregunta que se incluya en él debe estar pensada en función de su capacidad de cumplir este objetivo. Por ello, las preguntas constituyen la forma en que un investigador "traduce" un concepto que desea medir (o una porción de éste) en una pregunta y sus respectivas categorías de respuesta, comprensibles y respondibles por los encuestados.

En síntesis, podemos entender un cuestionario como un instrumento que, dirigiendo una conversación vertical hacia tópicos pre establecidos, traduce conceptos relativamente complejos derivados de la teoría del investigador, a preguntas y respuestas tan simples, que permiten que el encuestado exprese el grado o forma en que posee determinada variable o concepto en forma numérica, o al menos fácilmente cuantificable.

Este proceso de traducción se denomina "operacionalización" y ya que está detrás de la redacción y del sentido de todas las preguntas de un cuestionario, lo desarrollaremos con cierto detalle a continuación.

### **3.- La lógica de las preguntas y respuestas: el proceso de operacionalización.**

*Cuanto más precisos sean los cálculos,  
más tienden a difuminarse los conceptos”*  
R. S. Mulliken

La operacionalización es una estrategia general de investigación que se desarrolló en el momento en que las ciencias comenzaron a trabajar con conceptos no observables directamente (los que se denominan conceptos "latentes"). Ya que la lógica general del análisis científico consiste en que la resolución de las preguntas de investigación se debe hacer siempre con referencia a información

empírica, resultó evidente que era necesario un procedimiento que permitiese hacer observable aquello que no lo era en un primer momento.

En términos generales, el proceso de operacionalización consiste en: a) definir cuidadosamente un concepto que se pretende medir y que no es observable directamente (al menos en el estado de las técnicas de observación en el momento del estudio), b) derivar supuestos que relacionen el concepto a medir con otra serie de fenómenos, usualmente más simples, que serían consecuencias observables de la existencia en determinado grado del concepto latente (estas consecuencias se denominan "indicadores"), c) medir el nivel de presencia de estos indicadores, d) deducir de dicho número el grado en que el objeto estudiado posee la variable latente.

Un ejemplo de este proceso consiste por ejemplo en medir la edad de los árboles a partir de los anillos que se observan en su tronco, o la antigüedad de un fósil por el grado en que determinados elementos radiactivos se han transmutado en otros.

Evidentemente, en estos ejemplos queda claro que la validez y precisión de la medición obtenida depende de lo directa que sea la relación entre el concepto latente y el indicador disponible. Si algunos árboles produjeran dos anillos al año, mientras que otros de la misma especie produjeran sólo uno, la precisión y la validez de la estimación obtenida sería bastante relativa.

Se puede afirmar que la construcción de un cuestionario es una de las posibles aplicaciones del proceso de operacionalización: el investigador desea medir una serie de conceptos no observables directamente y para ello acude a indicadores empíricamente observables: las respuestas que da un sujeto a una encuesta telefónica o las marcas en el papel que deja en una encuesta autoadministrada.

Pero, ¿Cómo opera el proceso de operacionalización en el caso de los cuestionarios?

En primer lugar, cualquier investigador que pretende estudiar un tema en forma cuantitativa posee un conjunto de interrogantes, conceptos, e hipótesis, derivadas usualmente de una o varias teorías sobre el tema que pretende estudiar.

El primer paso en la construcción de un cuestionario consiste en determinar, a partir de estas preguntas, el listado de variables que desea conocer de los sujetos investigados.

Evidentemente estas variables tendrán distinto grado de complejidad (por ejemplo, el investigador podrá querer conocer la edad de su muestra, así como el nivel de autoritarismo familiar que ellos poseen), diversa naturaleza (algunas variables serán "objetivas" en el sentido de referirse a hechos, como podrá ser el voto en las últimas elecciones, o claramente subjetivas, como el nivel de satisfacción con su último encuentro sexual), o cumplirán distinta función en la investigación (algunas variables serán valiosas por si mismas o muy centrales para la investigación, como por ejemplo la calidad de la vivienda para un estudio sobre satisfacción habitacional, mientras que otras sólo tendrán sentido al estudiar su grado de asociación con las variables principales o son poco relevantes por si solas para el logro de los objetivos del estudio, como por ejemplo la pregunta por el sexo del jefe de hogar, o los años que lleva casada la pareja).

En cualquier caso, independiente de estas diferencias, sobre todos estos temas deberán incluirse preguntas (y respuestas) en el cuestionario. En términos generales, en un cuestionario las preguntas y

respuestas cumplen la función de los indicadores, mientras que las variables que queremos medir se entienden como conceptos latentes, no accesibles directamente al investigador.

No importa cuan simple es el concepto que el investigador desee medir (edad de las personas, por ejemplo), si se incluye una pregunta sobre esos temas en el cuestionario, es porque la información buscada no está disponible directamente para el investigador (es un concepto "latente"), teniéndose que pasar por lo que "indica" (responde) el sujeto para obtener la información buscada. Es por ello que en aquellos cuestionarios aplicados por un encuestador no se pregunta el sexo de la persona: no resulta necesario preguntarlo, la registra previa o posteriormente el encuestador. En otras palabras, generalmente el sexo no es un concepto latente, como si lo es la edad. Si fuera posible conocer con la suficiente exactitud la edad de una persona con sólo mirarla, tampoco se preguntaría en los cuestionarios (por ello, si deseamos conocer la edad en rangos gruesos -niño, adulto, anciano- tampoco es necesario preguntarla).

No obstante, el proceso por el cual se operacionaliza en preguntas y respuestas un concepto latente depende en gran medida de la complejidad de dicho concepto.

Pero... ¿qué entendemos por un concepto complejo y por uno simple?. Un concepto simple es aquel conocido y utilizado regularmente por las personas a las que se desea investigar. En otras palabras, en la cultura cotidiana de estos sujetos se tiende a utilizar el concepto, lo que permite suponer un nivel relativamente alto de comprensión de su significado y que este significado es compartido socialmente. Además, ya que el concepto es utilizado en las conversaciones cotidianas, se puede haber instalado en el colectivo la noción de que existen "grados" en que las personas lo poseen.

Ya que en el caso de estos conceptos simples se puede suponer un acuerdo entre el lenguaje del investigador y de los encuestados sobre el concepto/variable a medir, se puede obtener muy directamente la información buscada sobre la presencia o no del concepto en los sujetos o el grado en que estos lo poseen.

Por ejemplo, es posible preguntar directamente "¿cuántos años tienes?" cuando lo que se quiere conocer es el concepto latente "edad (medida en número de años) de las personas". Nótese que en este caso el concepto simple utilizado es de naturaleza demográfica y "objetivo" (en el sentido de hacer referencia a un suceso que podría verificarse por medios físico - biológicos o documentales, si dispusiéramos del tiempo y recursos necesarios), pero los conceptos/variables simples no tienen porqué ser sólo de este tipo.

Por ejemplo, ya que en el habla común se habla de "enamorarse" y de los grados de enamoramiento que las personas han tenido en su vida (concepto claramente subjetivo), es perfectamente posible preguntar en un cuestionario de relaciones de pareja: "¿Te sientes enamorado de tu pareja actual?", o "¿Cuán enamorado te encuentras actualmente?" (permitiendo las respuestas: mucho, más o menos, poco, no estoy enamorado).

Con esto no se quiere decir que la validez de las respuestas a preguntas simples de naturaleza "objetiva" o demográfica sea igual que las respuesta a preguntas simples de naturaleza subjetiva. Por el contrario, el análisis de las segundas es mucho más ambiguo que el de las primeras, pero eso no quita que es perfectamente posible preguntar de esa manera y esperar que, dentro de ciertos límites, las respuestas de los sujetos reflejen cierta autopercepción de los estados internos de las personas. En cualquier caso

el problema de la validez también afecta a las preguntas demográficas u objetivas (aunque quizás en grado menor y por causas distintas), mal que mal, ¿que puede impedir que las personas nos mientan sobre su edad o su voto en las elecciones pasadas o que no recuerden adecuadamente un hecho determinado?.

También es importante hacer notar acá que el proceso de preguntar (y ofrecer alternativas de respuesta), produce siempre un cierto grado de "cristalización" en la opinión de las personas. En otras palabras, el sólo hecho de preguntar genera que el sujeto se ponga en una situación en que no se había situado con anterioridad o frente a alternativas diferentes a la forma en que antes pensaba sobre el tema. Por ejemplo: quizás una persona nunca se había interrogado sobre su "nivel de satisfacción sexual" o al menos no lo había hecho en términos de los grados (muy satisfecho, satisfecho, insatisfecho o muy insatisfecho), que le propone el cuestionario.

Queda claro entonces que la realidad que estudiamos está siendo construida al mismo tiempo que la observamos. Por ello, cuando aplicamos un cuestionario estamos "produciendo" información y no sólo "recogiendo información", como se tendía a señalar en los manuales de investigación cuantitativa de los años 80.

Fuera de las consecuencias de orden ético que implica el fenómeno de la cristalización (quizás un cuestionario sobre satisfacción de relaciones de pareja genera algunas tomas de conciencia sobre hechos anteriormente no reflexionados), no cabe duda que el estudio de un fenómeno en que ha existido más cristalización será menos válido que otro en que la cristalización es mínima.

Entonces, ¿de que depende el grado de cristalización?. Básicamente de lo que hemos denominado simplicidad del concepto. Si en el lenguaje cotidiano de las personas se habla del concepto de manera similar a como podemos preguntarlo y a como lo quiere conocer el investigador (es decir, si el lenguaje de los sujetos coincide con el del investigador y por tanto, se puede preguntar en forma directa sobre el tema), la cristalización será mínima. Mientras más se alejen ambos lenguajes, más producirá cristalización el cuestionario.

Ejemplo claro de no cristalización es la pregunta por la edad de las personas. Ya que lo que quiere conocer el investigador (años que ha vivido el sujeto) es igual a la forma en que las personas hablan del tema (preguntándose con frecuencia "oye, y tú, ¿cuántos años tienes"), el investigador sólo tiene que repetir la pregunta del habla cotidiana para conocer lo que desea. Por el contrario, si quisiéramos conocer la edad en una cultura que no acostumbrara a llevar cuenta de sus años de vida o que midiera su edad por otro criterio (años bisiestos u osos que ha cazado, por ejemplo), la pregunta del investigador perdería todo sentido y sería imposible de realizar.

Esto es válido para conceptos simples, pero... ¿cómo se operacionalizan los conceptos complejos?. Ya hemos dicho que entenderemos por conceptos complejos aquellos que el habla cotidiana de las personas no utiliza, o al menos no lo hace de manera similar a como los entiende el investigador.

En estos casos el proceso de operacionalización puede ser entendido como un proceso de "traducción" entre el lenguaje del investigador hacia el habla de los sujetos, es decir, hacia la forma en que los sujetos son capaces de "expresar" el concepto o variable que el investigador quiere conocer.

En este caso el proceso de operacionalización tiene dos fases: a) una primera parte teórica y b) una segunda empírica.

En la primera parte, el investigador debe definir cuidadosamente el concepto y luego intentar fraccionarlo en unidades teóricas menores y más simples (las que a su vez también tienen que ser definidas). Si aún esas unidades son demasiado complejas debe repetir el proceso de desagregación hasta llegar a conceptos (que este caso podríamos llamar "subconceptos" o "subdimensiones", tales que por su simplicidad sean comprensibles y hablables por los encuestados).

Por ejemplo, si se desea medir un concepto como el "nivel de individuación de las personas", debería primero que nada definirse el concepto (digamos que se acepta la siguiente definición: "se entenderá por nivel de individuación el grado en que las personas toman o desean tomar sus decisiones de acuerdo a sus propios deseos e inclinaciones, sin seguir los comportamientos esperados de acuerdo a los grupos sociales a los que pertenecen"). Evidentemente este concepto es demasiado lejano al habla popular, como para poder preguntar directamente "¿cuán individuado estás tú?".

Lo que se debe realizar es dividir el concepto en sus "subconceptos" o "dimensiones" componentes. Por ejemplo, se puede entender la individuación como dividida en tres dimensiones: a) el grado en que el sujeto está dispuesto a romper normas sociales explícitas, b) el grado en que el sujeto siente que tiene deberes morales que restringen sus opciones conductuales y, c) el grado en que las personas perciben su biografía como producto de sus decisiones personales (versus percibirla como determinada social o estructuralmente).

Si consideramos que a este nivel los "subconceptos" son preguntables, podemos pasar a formular preguntas que midan cada uno de ellos. Si por el contrario, creemos que aún no hemos llegado al nivel en que los sujetos pueden "expresar" su nivel de individuación, debemos continuar con la desagregación teórica, fraccionando (y definiendo) cada uno de los subconceptos.

En el caso del ejemplo que estamos siguiendo el investigador consideró que el concepto ya estaba suficientemente desagregado, por lo que pasó a la fase empírica: redactar las preguntas con que se medirá cada subconcepto. En el ejemplo que estamos siguiendo, el PNUD (2002) decidió medir el grado en que el sujeto está dispuesto a romper las normas sociales con las preguntas: i) ¿Está dispuesto a seguir adelante con una idea aunque vaya en contra de la opinión de sus padres?, ii) ¿Está dispuesto a seguir adelante con una idea aunque vaya en contra de la opinión de su pareja? y iii) ¿Está dispuesto a seguir adelante con una idea aunque vaya en contra de la opinión de la iglesia?. Resulta claro que responder en contra de estas instituciones implica estar más individualizado que quien no sea capaz de dar esa respuesta.

Por su parte, el segundo subconcepto: grado en que el sujeto siente que tiene deberes morales que restringen sus opciones conductuales, fue medido con la pregunta: ¿cómo le gustaría ser recordado?, con las siguientes opciones de respuesta: a) Como alguien que se entregó a los demás y fue querido por ellos, b) Como alguien que salió adelante contra viento y marea, c) Como alguien que fue fiel a sus sueños y vivió de acuerdo a lo que se propuso, d) Como alguien que siempre supo cumplir con su deber. Evidentemente la respuesta "c" indica mucho mayor monto de individuación que las otras.

Finalmente, el tercer subconcepto: grado en que las personas perciben su biografía como producto de sus decisiones personales, fue medido con la pregunta: ¿Cuál de las siguientes frases lo representa

mejor?: a) En la vida uno tienen que hacer lo que hay que hacer, y b) Yo analizo mi vida y veo que hacer. En este caso la opción "b" indica mayor individuación que la opción "a".

En todos estos ejemplos debe notarse que el supuesto de fondo es que las personas ya están en condiciones de dar respuestas a preguntas formuladas en este nivel, es decir, las preguntas representan la forma en que las personas pueden "expresar" el nivel que poseen de un concepto tan complejo como "individuación".

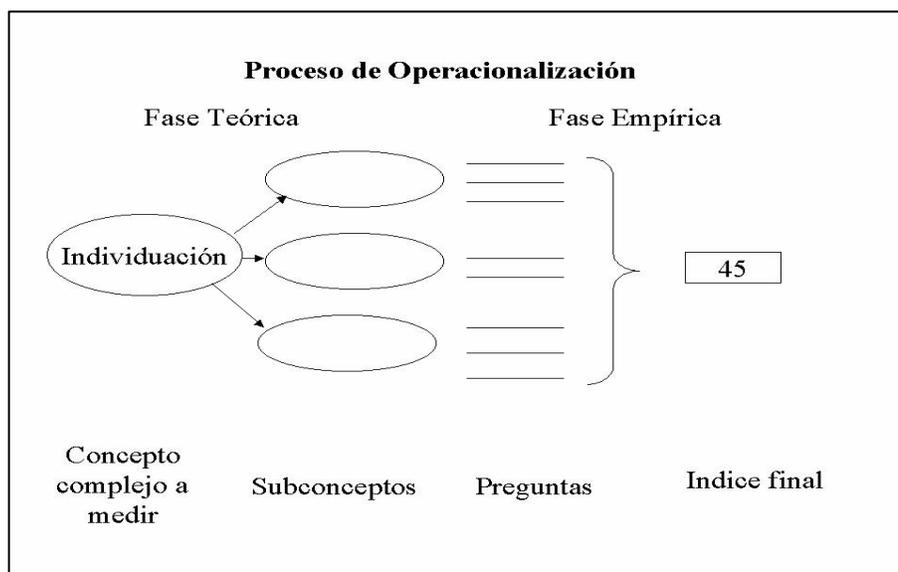
Una vez que se han formulado las preguntas (y sus opciones de respuesta) requerimos generar un procedimiento que permita integrar las respuestas que el sujeto dará a cada una de las preguntas de manera de obtener una sola puntuación que indique el nivel de "individuación" de las personas.

Este proceso empírico de reunión de las preguntas en un solo puntaje se denomina construcción de índices o escalas (dependiendo del tipo de preguntas que estemos utilizando, como se explicará más adelante). En términos concretos, un índice es una fórmula matemática que combina de alguna manera las respuestas de las personas en una sola puntuación.

En el ejemplo que estamos siguiendo, el índice podría ser asignar a la persona 1 punto de individuación cada vez que el responde utilizando la alternativa que indica poseer en alto grado este concepto y 0 punto cada vez que opta por otras alternativas. Por ejemplo, se podría asignar un punto cada vez que el sujeto se manifieste dispuesto a ir contra la opinión de sus padres, su pareja o su iglesia, cuando preferiría ser recordado como alguien que fue fiel a sus sueños, y cuando señala que para tomar decisiones analiza su vida y decide.

En este caso un sujeto que haya tomado todas las opciones "individualizadas" obtendría 5 puntos que en este caso sería la máxima individuación. Por otro lado, si el decidiera tomar las otras opciones podría obtener un mínimo de 0 puntos (que en este caso se interpretaría como nada o muy poco individualizado). Evidentemente la forma (o fórmula) concreta que se utilice para construir el índice puede cambiar, pero lo central es que siempre la idea es reflejar con un solo puntaje el grado en que el sujeto posee el concepto que deseamos medir.

Una representación gráfica del concepto de operacionalización la encontramos en la siguiente lámina:



Si bien en este ejemplo hemos utilizado un concepto de carácter "subjetivo" para mostrar el proceso de operacionalización de un concepto complejo, esta operatoria no es de ninguna manera privativa de este tipo de conceptos.

Por el contrario, si un investigador deseara conocer por ejemplo la "calidad de la vivienda" que poseen los habitantes de una determinada ciudad (un concepto que dice relación con algo totalmente objetivo), podría dividir el concepto en los siguientes subconceptos: a) calidad de la materialidad de la vivienda, b) nivel de acceso a servicios básicos, c) disponibilidad de servicios sociales en las cercanías. En el primer subconcepto podría preguntarse por el material del techo, del piso o de las paredes de la vivienda (especificando que materiales indican buena, regular y mala calidad), mientras que en el segundo subconcepto podría preguntarse por la disponibilidad de agua, luz, gas, en el interior de la vivienda. Finalmente en el tercero se podría preguntar por la presencia de colegios, carabineros, plazas, etc. en un radio de 5 manzanas alrededor de la vivienda. Posteriormente con todas estas preguntas se podría construir un índice que nos daría la calidad de las viviendas de las personas.

Resulta interesante notar que si bien el concepto "calidad de la vivienda" podría parecerse lo suficientemente simple como para poder ser preguntado directamente a los sujetos (con una pregunta como: ¿De qué calidad es la vivienda que habitas?, con las alternativas a) muy buena, b) regular, c) mala), al decidir utilizar el procedimiento de operacionalización para variables complejas el investigador supone que la forma en que las personas expresan el concepto "calidad de la vivienda" no es el mismo que él desea conocer (quizás las personas no toman en cuenta aspectos que a él le interesen, como la infraestructura comunitaria, o tal vez hablan de este concepto de una manera demasiado simple).

En términos generales, esta forma de medir los conceptos nos entrega más información (más detallada) que la medición directa de variables simples, pero tiene un mayor costo: requiere muchas más preguntas en el cuestionario, exige más tiempo al sujeto para responder y requiere un trabajo teórico previo más fino por parte del investigador.

Además, sigue siendo un proceso indirecto en que interpretamos en cierto sentido las respuestas que los sujetos entregan en el cuestionario y esto no sólo implica que estamos expuestos al fenómeno de la cristalización o de la falta de veracidad de las respuestas de las personas, sino también a que mucho del sentido del concepto que queremos medir se puede "perder" en el proceso de definirlo operacionalmente (es decir a través de las preguntas que realmente les formulamos a las personas).

Para expresar lo anterior con un ejemplo, podríamos preguntarnos ¿quién puede asegurar que el sentido completo y global de la "individuación", tal como está definida teóricamente, es recogida por las 5 preguntas que decidimos incluir en el cuestionario?, ¿es sólo eso individuación?, ¿no nos faltarán preguntas?.

La verdad es que debemos reconocer que el proceso de operacionalización siempre implica un grado de distorsión o mutilación del sentido teórico de un concepto. Evidentemente, si logramos formular preguntas más relacionadas con el concepto y más comprensibles por la población, introduciremos menos distorsión, pero ella siempre existirá en algún grado. Sinceramente, nunca logramos medir totalmente el concepto que buscamos, sólo obtenemos mejores o peores interpretaciones empíricas.

Esto último también es cierto cuando operacionalizamos conceptos simples: también acá puede haber diferencias de interpretación entre las preguntas de los investigadores y las respuestas de los encuestados, aunque aquello dependerá obviamente del grado de simplicidad del concepto (sobre la pregunta ¿qué edad tienes?, es poco posible que existan diferencias de interpretación, pero sobre la pregunta ¿cuál es tu estado civil actual?, efectivamente pueden haber situaciones complejas no fáciles de encasillar en las alternativas presentadas.

Al describir el proceso de operacionalización de conceptos a través de cuestionarios hemos dejado abiertas una serie de cuestiones, que trataremos en las próximas secciones: ¿cómo se redactan las preguntas y las respuestas a los cuestionarios?, ¿cómo decidir cuantas preguntas utilizar al medir cada concepto o subconcepto?, ¿qué son exactamente los índices y las escalas?.

#### **4.- La formulación de las preguntas y respuestas: tipología y recomendaciones de redacción.**

*"Bienaventurados los que se arriesgan a dar consejos,  
porque serán sabios a costa de los errores ajenos"*  
Juan Manuel Serrat

Las preguntas y sus alternativas de respuesta constituyen los elementos constituyentes de los cuestionarios y como tales sus características definirán gran parte de la utilidad o calidad de éstos.

Ahora bien, ¿Qué tipos de preguntas se pueden incluir en los cuestionarios?. La verdad es que en un cuestionario se pueden formular muy diversas preguntas, que pueden ser clasificadas desde varios puntos de vista. Nos parece que tres de estos puntos de vista son especialmente esclarecedores de los tipos de preguntas posibles de incluir y los desafíos específicos que hay que vencer para lograr preguntas de calidad.

En primer lugar, desde el contenido preguntado, en un cuestionario se pueden incluir:

- a) **Preguntas relativas a hechos:** son aquellas preguntas en que se inquiera respecto a una conducta o fenómeno sucedido o que se encuentra sucediendo efectivamente. Ejemplo típico de este tipo de preguntas son: ¿Por quien a votado usted en las últimas elecciones? (y se le ofrecen al sujeto los nombres de los candidatos efectivos), ¿con quien vive usted? (y se le presentan al encuestado las referencias a diferentes familiares y no familiares), ¿cuál es su edad? (y se deja espacio para que el sujeto escriba el número de años que ha cumplido).

En este caso la clave de la formulación de buenas preguntas consiste en redactarlas de manera de hacerlas lo más claras y unívocas posibles y presentar al sujeto las opciones reales de conductas posibles. Por ejemplo, si se pretende conocer las personas que forman el hogar del encuestado se debe explicar con claridad que se entenderá por hogar de manera que no se incluyan personas que habitando la misma vivienda sean parte de otro hogar. Además, las alternativas de respuesta deben permitir incluir a personas familiares o no familiares, permitiendo que sean bien registradas aquellas formas de hogar no tradicionales, como estudiantes o amigos que conforman un hogar.

En cualquier caso nunca se debe olvidar que independiente de que estemos preguntando por hechos, lo que recibimos de los sujetos no son esos hechos, sino la percepción, recuerdo o lo que nos desea transmitir el sujeto. En otras palabras, lo que obtenemos no es el hecho real, sino un

indicador de éste que pasa por el tamiz de lo que el sujeto puede o quiere reportarnos. Si notamos que tanto el recuerdo como la percepción están fuertemente influidas tanto por las características personales como sociales de la persona encuestada, podremos darnos cuenta que independiente de la voluntad que tenga la persona de respondernos la verdad, lo que tendremos será, en el mejor de los casos una cierta interpretación de ésta.

- b) **Preguntas de aptitud:** son aquellas que miden el nivel de habilidad, capacidad o conocimiento que los sujetos poseen sobre un tema determinado. Ejemplos de estas preguntas son aquellas incluidas en los tests de inteligencia o de habilidad verbal, matemática o conocimientos (pruebas de selección universitaria o prueba SIMCE, por ejemplo).

La principal diferencia entre estas preguntas y las otras es que en éstas existen alternativas de respuestas correctas e incorrectas y si el sujeto elige las primeras demuestra un mayor nivel de dominio que si escoge las segundas. La clave de la formulación de buenas preguntas de aptitud es que ellas tengan niveles diferentes de dificultad, de manera que las personas poco hábiles tengan posibilidades de responder algunas adecuadamente, mientras que las personas de más habilidad tengan algunas lo suficientemente difíciles como para que sólo los mejores puedan responderlas correctamente. De esta manera se logrará distribuir a los sujetos en todas las puntuaciones posibles.

Ya que usualmente de la respuesta a estos cuestionarios se deducen mayores o menores oportunidades de desarrollo vital para las personas (obtener un trabajo, aprobar un curso o entrar a una carrera universitaria), acá no nos enfrentamos con el problema del falseamiento voluntario de las respuestas (es poco habitual que alguien falle un test voluntariamente).

- c) **Preguntas de actitud, opinión, expectativas o intenciones conductuales:** son aquellas en que se pretende conocer lo que un sujeto haría en una determinada situación, su opinión, sus deseos o actitudes. En general son preguntas relacionadas con lo usualmente denominado "subjetivo".

En este caso, ya que lo que se pretende conocer no está "anclado" en los conocimientos ni en las conductas o hechos reales del sujeto, la formulación de las preguntas debe cuidar muy especialmente el problema de la deseabilidad social de las respuestas, es decir, la tendencia de las personas a responder lo que ellos creen es socialmente correcto o esperado por el encuestador. En el caso de las preguntas relativas a hechos si bien es posible que el sujeto no responda la verdad o no tenga un recuerdo fiel de éste, se esperaría que el hecho real de alguna manera produzca que algunas personas hagan referencia a él (mentir implica un costo que no todas las personas están dispuestas a asumir), este efecto "ancla" no lo encontramos en las preguntas subjetivas, por lo que es mucho más fácil que las personas distorsionen sus actitudes u opiniones reales para conformarse a lo socialmente deseable o correcto.

Por todo lo anterior, en estas preguntas se debe poner especial atención a evitar palabras cargadas positiva o negativamente, a redactar las preguntas de modo de validar tácitamente todas las posibles respuestas y a que ninguna pregunta sea más "deseable" que las otras. Más adelante profundizaremos en estas recomendaciones.

Una segunda forma de clasificar a las preguntas de un cuestionario es aquella que las ordena de acuerdo al tipo de alternativas de respuesta que incluyen. Desde esta perspectiva se puede distinguir entre:

- a) **Preguntas abiertas:** son aquellas preguntas en que el investigador no presenta al sujeto alternativas fijas de respuesta. En cambio se deja un espacio pequeño o mediano para que sea el mismo sujeto el que estructure su respuesta.

Este tipo de preguntas si bien tienen la importante ventaja de no "cristalizar" o restringir la opinión de los sujetos a una serie de alternativas prefijadas, tienen muchos más costos que las preguntas cerradas: hay que construir con posterioridad las categorías de respuesta que serán numeradas (proceso de codificación), lo que consume mucho tiempo y energías y además toman más tiempo de responder por parte del sujeto.

Estas razones hacen que las preguntas abiertas sean más bien raras en los cuestionarios y se reserven sólo para situaciones en que sea realmente necesario obtener la opinión no distorsionada de las personas y/o el investigador no tenga ninguna posibilidad de construir alternativas de respuestas previamente a la aplicación del cuestionario.

Además, dado el costo de tiempo y energías que tiene codificar largas preguntas abiertas, se tiende a preferir la utilización de preguntas abiertas de respuesta breve (una sola palabra o frase) que las de respuesta extensa (uno o más párrafos).

- b) **Preguntas cerradas de respuesta múltiple:** son aquellas en que se presenta al sujeto una serie de alternativas de respuesta, de las cuales el encuestado puede elegir más de una.

En general este tipo de pregunta se utiliza para estudiar temas no excluyentes, es decir, en donde optar por una alternativa no tiene porque significar que se deje de lado las otras. Por ejemplo cuando a un grupo de alumnos se les pregunta que tipo de deporte les gustaría practicar, con el objeto de planificar las actividades extraprogramáticas de un año académico. Evidentemente si alguien desea hacer pesas, eso no excluye que al mismo tiempo (y quizás con la misma intensidad) desee practicar tenis, por lo que se puede permitir que éste elija más de una alternativa.

Estas preguntas soportan varios formatos posibles de restricción de las alternativas de respuesta: algunas veces se entrega total libertad al encuestado para escoger todas las alternativas que desee, mientras que en otros casos se limitan las elecciones posibles a sólo 2 u otro número. También es posible que se pida jerarquizar la preferencia (asignando un 1 a la alternativa más deseada, un 2 a la siguiente y así sucesivamente).

Este tipo de pregunta si bien no tiene tantos costos como las preguntas abiertas, de todas maneras plantea algunas complejidades al momento de analizar los resultados o digitar las respuestas en la base de datos, por lo que no se la utiliza con demasiada frecuencia, a pesar que tiende a reflejar con más claridad la complejidad de las elecciones de las personas, las que rara vez son tan excluyentes.

- c) **Preguntas cerradas de respuesta simple:** son aquellas preguntas, mayoritarias en los cuestionarios habituales, en que se permite escoger sólo una alternativa de respuesta, traspasando al propio sujeto la complejidad de escoger la más deseada dentro de las que se le presentan.

Se espera que las alternativas de respuesta que se presentan a las personas tengan dos cualidades: que sean excluyentes (es decir, que no se traslapen sus sentidos, de manera que marcar una de ellas

no implique también marcar otra) y que sean exhaustivas, es decir, que las alternativas presentadas cubran todas las posibles (o al menos las principales) respuestas frente al tema preguntado. Para lograr la segunda propiedad es que muchas preguntas incorporan la opción "otra respuesta", pero esta no es realmente la solución a presentar al sujeto las principales respuestas posibles frente a un tema, ya que pocas personas parecen dispuestas a marcar la alternativa "otra", prefiriendo distorsionar su opinión escogiendo una alternativa que realmente no refleja lo que verdaderamente piensan.

Por otro lado, este tipo de pregunta tiene la dificultad que usualmente simplifica en exceso las intenciones, opiniones, actitudes, recuerdos, situaciones, en que están o creen los sujetos, ya que posiblemente éstos tienen preferencia por más de una opción o incluso pueden no estar de acuerdo con ninguna de las presentadas (en este caso se obliga a escoger a la persona la alternativa "menos mala" de las que se le presentan).

En otras palabras, con este tipo de preguntas no recogemos la opinión o situación verdadera (y compleja usualmente) de las personas frente a un tema, sino más bien la alternativa más cercana a su opinión o situación real, dentro de las presentadas. Esto implica que siempre hay algún grado de distorsión al obtener información de este tipo de preguntas. Por ejemplo, si nosotros le preguntamos a alguien las personas con que vive en su hogar y en este hogar hay un integrante que pasa parte del año en otro lugar (digamos un tío que estudia algunos meses al año fuera de casa o viaja mucho, por ejemplo), la persona entrevistada tendrá que decidir si marca o no la respuesta "tío o tía" en las alternativas presentadas. Cualquiera sea su respuesta, la respuesta del sujeto simplifica una situación bastante más compleja.

Ahora bien, a cambio de esa distorsión y/o simplificación, este tipo de preguntas son muy fáciles de codificar, de digitar en el cuestionario y de analizar estadísticamente, por lo que tienden a ser preferidas por los investigadores.

En términos conceptuales, lo que hacen estas preguntas es traspasar al encuestado la tarea de simplificar su compleja situación u opinión, escogiendo de las respuestas o situaciones simples que se le presentan. Se espera que al hacer este proceso de simplificación la persona sea capaz de escoger lo más relevante, frecuente o cercano a su opinión real, por lo que el investigador analiza solamente esta esencia simplificada.

Finalmente, una tercera vía para clasificar las preguntas que se pueden incluir en un cuestionario consiste en clasificarlas de acuerdo al nivel de medición que implican. Desde este punto de vista se puede distinguir entre:

- a) **Preguntas de respuesta nominal dicotómica:** Son aquellas en que el encuestado debe escoger sólo entre dos alternativas de respuestas. En general estas alternativas presentan opciones antagónicas (verdadero/falso, si/no, acuerdo/desacuerdo, etc.) que obligan al sujeto a simplificar fuertemente la situación estudiada y a escoger entre dos polos.

El principal defecto de estas preguntas es su baja "sensibilidad" frente a las actitudes o conductas de las personas, es decir, dos personas con actitudes diferentes aunque similares en su sentido marcarán exactamente la misma respuesta, impidiendo que el investigador detecte las diferencias reales que existen entre las opiniones de los sujetos. Por ejemplo, dos personas marcarán la misma

respuesta aunque una persona crea fuertemente que la pregunta es "falsa", mientras que otra piense que es mayoritariamente "falsa", aunque no es su totalidad..

Dada esta extrema simplificación, las preguntas nominales dicotómicas son muy rápidas de responder y muy adecuadas para encuestas telefónicas, autoadministradas o población con nivel educativo más bajo (que tradicionalmente tiende a expresar sus juicios de manera más categórica). Por otro lado, las personas de nivel educativo más alto tienden a sentirse incómodas con este formato, reclamando que la situación no es tan polar.

En todo caso, este formato permite formular muchas preguntas en el cuestionario y realizar numerosos análisis estadísticos, por lo que tiende a ser muy utilizada en test orientados para población general.

- b) **Preguntas de respuesta nominal no dicotómica:** son aquellas preguntas en que el investigador construye una serie acotada de respuestas que no se encuentran ordenadas entre sí. En este caso se requiere que el investigador conozca bastante del tema que está preguntando y sobre todo de las formas en que la población podría responder a la consulta.

Esto último es necesario ya que una de las razones por las cuales puede fallar una pregunta en un cuestionario es que las personas no encuentren en las alternativas presentadas una opción lo suficientemente atractiva como para marcarla (ya sea porque las opciones presentadas son poco pertinentes o no exhaustivas).

En este caso pueden ocurrir tres fenómenos, todos los cuales indican errores de medición: a) un conjunto de personas, al no encontrar una alternativa a su gusto, dejarán la pregunta sin contestar. De esta manera porcentajes altos de no respuestas a determinadas preguntas son indicadores de problemas en la pregunta, b) otro conjunto de personas podría optar por utilizar la respuesta "otros" (si es que la pregunta lo incluye). Al igual que el caso anterior, demasiada utilización de esta respuesta indica problemas en la conformación de las alternativas de respuesta a la pregunta, finalmente, c) otro conjunto de personas podría marcar una alternativa que no sea realmente de su agrado, distorsionando los resultados de la investigación. Este tercer fenómeno es el más grave, ya que no es posible detectarlo y el investigador puede extraer consecuencias equivocadas de las intenciones, conductas u opiniones de las personas.

Por ejemplo, veamos la siguiente pregunta: ¿podría decirme la religión o iglesia a la que usted se siente más cercano?, con las alternativas de respuesta: a) católica, b) evangélica, c) mormona, d) Otra iglesia cristiana, e) judía, f) otra iglesia no cristiana, g) no soy religioso, h) no sé.

En este ejemplo, el supuesto que es que las iglesias o religiones presentadas son las principales para la población que se está estudiando. Si por alguna casualidad esta pregunta fuera traducida a alguna lengua árabe y aplicada en Tunes, funcionaría mal, ya que la mayor parte de la población no encontraría la opción que realmente le gustaría marcar.

Se puede hipotetizar que este conjunto de personas de religión musulmana se inclinarían a las alternativas d), f), h) o derechamente dejaría la pregunta sin respuesta. Todas estas conductas en el fondo nos indican que dejamos fuera una alternativa relevante al redactarlas. Peor aún sería que un porcentaje de los respondientes (quizás cansados por una entrevista excesivamente larga) marcaran

a), b), c) o e), ya que esta distorsión no sería detectada. La solución clara a este problema sería incluir la alternativa "musulmana" a las alternativas de respuesta.

Otro desafío que enfrentamos al redactar alternativas de respuesta a este tipo de preguntas es lograr que cada alternativa tenga algún sentido interpretativo establecido de antemano. Por ejemplo observemos la pregunta: ¿cuál de las siguientes situaciones le cuesta más?, con las alternativas de respuesta: a) reclamar o hacer reclamos a otros, b) pedir ayuda a otras personas, c) expresar ante otros sus sentimientos, d) ponerle límites a la gente y decir que no y, e) conversar intimidades con su pareja.

Si bien acá aparece el desafío de la pertinencia de las respuestas, ya que a las personas podrían no gustarle ninguna de estas alternativas, queremos ahora poner el acento en otro tema: que cada una de las alternativas presentadas debe "significar" o "indicar" algo para el investigador. Por ejemplo, posiblemente las alternativas b) y d) son indicadores de dificultades de sociabilidad en relaciones de comunidad, las alternativas c) y e) expresan dificultades de sociabilidad en relaciones íntimas, mientras la alternativa a) indica dificultades de sociabilidad en relaciones de sociedad.

Lo importante acá es tener claro de antemano que alternativas representan o indican que sentidos o conceptos, ya que adjudicar sentidos con posterioridad a la aplicación del cuestionario no siempre es fácil e incluso puede llevarnos a concluir que nos faltó alguna alternativa que indicara algún concepto teóricamente relevante.

- c) **Preguntas de respuesta ordenada:** son aquellas preguntas en que las alternativas de respuesta constituyen un continuo ordenado desde la máxima aceptación, frecuencia o cercanía a un tema, hasta el máximo rechazo, lejanía o mínima frecuencia de ocurrencia del fenómeno.

También es usual, aunque no indispensable, que la pregunta no esté redactada en forma interrogativa, sino como una afirmación, frente a la cuál el sujeto expresa su grado de cercanía.

Ejemplo típico de estas preguntas la constituyen muchos test sociológicos, en que las personas deben señalar su grado de acuerdo frente a cada frase presentada, o el grado de frecuencia con que ocurre cierta conducta (las alternativas usualmente utilizadas en este caso son: a) muy de acuerdo, b) de acuerdo, c) indeciso, d) en desacuerdo y, e) muy en desacuerdo; también se puede encontrar: a) muy frecuentemente, b) a veces, c) rara vez y, d) nunca). Otra forma de expresar relaciones de frecuencia de algunas conductas puede ser: a) todos los días, b) hasta tres veces a la semana, c) una vez a la semana, d) entre una y dos veces al mes y d) casi nunca.

En estas preguntas los problemas de pertinencia de las respuestas, su exhaustividad y su carácter excluyente están resueltos, ya que como el continuo cubre todas las posibles actitudes frente a la frase, las personas siempre podrán sentirse más o menos cercanas a alguna de las alternativas de respuesta.

En cualquier caso dos discusiones existen sobre este tipo de preguntas: a) ¿cuántas alternativas incluir? y ¿se debe redactar una alternativa intermedia?.

Sobre lo primero, la conclusión más fuerte es que no es necesario presentar demasiadas alternativas para tener una medición fiable de las actitudes o conductas de las personas, por lo que se sugiere un

mínimo de dos o tres alternativas y un máximo absoluto de siete. Por ello, la mayor parte de estas preguntas incluyen entre cuatro y cinco alternativas, que parece ser una opción razonable.

En todo caso, se debe considerar que mientras más bajo sea el nivel educativo de las personas, más se sentirán incómodas con muchas alternativas, por lo que un cuestionario dirigido hacia esos segmentos de la población debe tener pocas alternativas (digamos unas tres).

Además, se debe tomar en cuenta como se aplicará el cuestionario, ya que si bien el soporte autoadministrado permite más opciones de respuesta, una entrevista telefónica impone las menos alternativas posibles.

Con respecto a la alternativa intermedia, el debate es aún menos claro. No existe una recomendación precisa, aunque se sabe que colocar una alternativa intermedia tiene fundamentalmente dos defectos: a) puede permitir a las personas no tomar posición o "fugarse" de tomar opciones difíciles o poco deseables (esto es especialmente relevante al medir conceptos complicados como racismo, intolerancia, conductas sexuales, etc.) y, b) no siempre encontramos en el español una palabra que exprese claramente el sentido intermedio que queremos. Por ejemplo "indeciso", "no se", "en medio", "de acuerdo y en desacuerdo", "?" (que son las más utilizadas), no necesariamente implican estar en la mitad del continuo que representan las otras alternativas de respuesta.

- d) **Preguntas de respuesta intervalar:** son aquellas en que las alternativas de respuesta están constituidas por un continuo numérico frente al cual el sujeto debe posicionarse.

Usualmente son los mismos números lo que le entregan sentido a la respuesta o el investigador le coloca una etiqueta a las respuestas más extremas, de manera de guiar al encuestado. Por ejemplo en la pregunta:

¿Cuál es tu posición política?  
Muy de izquierda 1 2 3 4 5 6 7 8 9 10 muy de derecha

En este caso, se presupone una estructura unidimensional de las posiciones políticas, cuyos polos están constituidos por las etiquetas asignadas por el investigador.

Este tipo de preguntas tienden a utilizarse cuando es sostenible esta estructura unidimensional y cuando se puede suponer que las personas son capaces de situarse en este continuo (en general esta capacidad requiere niveles educativos más o menos altos). Además, esta forma de preguntar funciona mejor en aplicaciones autoadministradas o con un encuestador provisto de tarjetas, ya que comprender la pregunta usualmente requiere observar el continuo y sus polos.

La ventaja de este tipo de preguntas es que permite una amplia variabilidad de respuestas con sólo una pregunta (por tanto es más sensible a las actitudes, deseos o conductas del sujeto, al contrario que las preguntas dicotómicas), por lo que tiende a utilizársela cuando queremos medir un concepto con una sola pregunta, pero al mismo tiempo necesitamos que esta sola pregunta sea lo más intervalar posible.

El principal defecto de las preguntas intervalares es que salvo los extremos, el resto del continuo no tiene un sentido asignado, por lo que es el sujeto que responde el encargado de darles sentido. Esta

indeterminación de la escala utilizada implica que es muy posible que dos personas utilicen de manera diferente la escala, asignando números diferentes a iguales actitudes o conductas sólo porque para uno un número determinado es muy bajo, mientras que para otro es intermedio. En términos más concretos, para una persona marcar un 8 indica ser moderadamente de derecha, mientras que otro con la misma actitud política marcaría un 6.

Esperamos que esta discusión sobre los diferentes tipos de preguntas a incluir en un cuestionario hayan servido tanto para conocer las diversas posibilidades de preguntar en estos instrumentos, como para familiarizarse con las limitaciones y desafíos que impone cada tipo de pregunta.

A continuación trataremos de resumir algunos consejos generales para la redacción de las preguntas que conviene tener en cuenta al diseñar cuestionarios. Creemos que este recetario puede servir como lista de chequeo para quienes se encuentren en dura tarea de redactar las preguntas de un instrumento.

### Recomendaciones prácticas para redactar preguntas y respuestas:

<b>Recomendaciones generales</b>	
<ul style="list-style-type: none"> <li>• Trabaje siempre en grupo. Mientras más puntos de vista existan al momento de formular preguntas, mejor saldrán.</li> </ul>	
<ul style="list-style-type: none"> <li>• Conozca lo mejor posible la cultura y lenguajes de las personas a las que se va a aplicar el cuestionario. Trate que éste sea comprensible por el subgrupo menos educado de ellos.</li> </ul>	
<ul style="list-style-type: none"> <li>• Adapte el lenguaje a las personas a las que va a aplicar el instrumento (por ejemplo, si el cuestionario es para jóvenes prefiera tutear que tratar de usted). No obstante no exagere en la utilización de modismos, pues estos pueden dificultar la interpretación de los resultados.</li> </ul>	
<ul style="list-style-type: none"> <li>• Recuerde que usualmente el dominio de lenguaje del investigador es mucho mejor que el de los sujetos investigados, por lo que trate de sólo utilizar palabras usuales en una conversación cotidiana (incluso aquellos conceptos que usted puede creer conocidos, no son dominados por muchas personas).</li> </ul>	
<ul style="list-style-type: none"> <li>• Someta el cuestionario final a jueces que evalúen al menos a) la pertinencia de su lenguaje, b) su longitud, c) el interés que suscitará en los encuestados.</li> </ul>	
<b>Recomendaciones específicas</b>	<b>Para formular preguntas</b>
	<ul style="list-style-type: none"> <li>• Formule preguntas o frases que permitan que las personas se distribuyan en las diferentes opciones de respuesta. Recuerde que una de las capacidades que debe tener una pregunta es "discriminar" entre personas que tengan distintas conductas, actitudes u opiniones, por lo que si usted formula preguntas donde el 90% de las personas escogerá una alternativa, posiblemente sólo habrá descubierto una homogeneidad cultural, pero no podrá utilizar esta pregunta como una variable (mal que mal, será prácticamente una constante). No es que descubrir estas homogeneidades sea inútil, pero usualmente resulta más interesante trabajar con variables.</li> </ul>

	<ul style="list-style-type: none"> <li>• Evite formulaciones "gratuitas" cuando pregunte por intenciones de conducta muy deseables, pues posiblemente todo el mundo estará de acuerdo con lo planteado. Por ejemplo la pregunta: "¿le gustaría a usted volver a estudiar si tuviera la oportunidad?", será respondida afirmativamente por casi toda la población. Trate siempre de poner un contrapeso del tipo "¿le gustaría a usted volver a estudiar, aunque tuviera que hacerlo por las noches o el fin de semana?".</li> <li>• Si está preguntando un tema muy complejo moralmente, formule la pregunta utilizando el "yo auxiliar", es decir, refiera la pregunta a "la gente como usted". A pesar que esta formulación hará más compleja la interpretación de las respuestas puede valer la pena para estudiar temas éticamente complicados.</li> <li>• No introduzca dobles negaciones pues confunden al respondiente (por ejemplo evite frases del tipo: "¿Cuán en desacuerdo está usted con la frase: no me gusta esta democracia?").</li> <li>• Intente que en cada formulación exista sólo una idea, eso facilitará interpretar posteriormente las respuestas.</li> <li>• Frente a temas complicados trate de legitimar en la formulación todas las posibles respuestas. Esto se puede lograr colocando frases del tipo: "Frente a este tema algunas personas piensan "x", mientras que otras piensan "y", ¿con quienes está de acuerdo usted?".</li> <li>• Evite palabras cargadas emocionalmente, tanto positiva como negativamente (por ejemplo, evite la palabra "aborto", prefiriendo "interrupción de embarazo"). Recuerde que las personas reaccionan tanto a la carga emocional de las palabras utilizadas, como a la pregunta realmente efectuada.</li> <li>• Trate que la redacción de sus preguntas nunca tenga más que 20 palabras. Es difícil que las personas retengan una pregunta más larga.</li> </ul>
<b>Recomendaciones específicas</b>	<b>Para formular alternativas de respuesta</b>
	<ul style="list-style-type: none"> <li>• No formule demasiadas alternativas de respuesta (no más de 5 ó 6), salvo que sea estrictamente necesario y que se disponga de un sistema de tarjetas o la encuesta sea autoadministrada. Las personas no retendrán demasiadas alternativas de respuesta.</li> <li>• Prefiera una redacción simple (una sola idea) y breve. Las alternativas deben tener todas más o menos la misma longitud, ya que las respuestas más extensas tienden a funcionar diferente que las más breves.</li> <li>• Evite palabras cargadas emocionalmente, tanto positiva como negativamente (por ejemplo, frases que incluyan palabras como: "democracia" o "libertad" tienden a ser más marcadas que otras que no tengan este sesgo emocional).</li> </ul>
	Regule el número de alternativas de respuesta de acuerdo a: el nivel educativo de los sujetos a quienes va dirigido el instrumento (menos alternativas cuanto menor sea el nivel educativo de las personas), y de acuerdo a la forma de aplicación prevista (menos alternativas en entrevista telefónica, más en aplicaciones autoadministradas o con encuestador provisto de tarjetas).

<b>Recomendaciones específicas</b>	<b>Para preguntas abiertas</b>
	<ul style="list-style-type: none"> <li>No incluya demasiadas preguntas abiertas, límitelas sólo a lo indispensable. El trabajo de codificación posterior es muy pesado y toma mucho tiempo.</li> </ul>
	<ul style="list-style-type: none"> <li>No deje mucho espacio para que las personas escriban sus respuestas, eso sólo puede incentivar respuestas demasiado largas. Las personas que requieren más espacio de todas maneras lo consiguen escribiendo en los márgenes del instrumento.</li> </ul>
<b>Recomendaciones específicas</b>	<b>Para preguntas cerradas de respuesta simple</b>
	<ul style="list-style-type: none"> <li>Trate siempre de incluir la alternativa "otra", seguida de una línea en que el sujeto pueda escribir su respuesta. Si esa alternativa recibe un número desusado de respuestas, elimine del análisis la pregunta y trate de mejorarla para estudios posteriores utilizando lo redactado en el espacio.</li> </ul>
<b>Recomendaciones específicas</b>	<b>Para preguntas de respuesta intervalar</b>
	<ul style="list-style-type: none"> <li>Siempre que sea posible trate de escribir palabras bajo cada uno de los números que indiquen a las personas la interpretación que deben dar a cada uno de ellos (por ejemplo utilizando palabras como "bastante de derecha", "derecha", "centro derecha"). Esto ayudará a la interpretación de los resultados y disminuirá los errores producto de diferentes formas de utilización de la escala numérica.</li> </ul>

## 5.- La construcción de índices y escalas: la utilización de múltiples preguntas.

*"Los átomos son, pues, sólidos y simples,  
formando un todo coherente de partes mínimas"*  
Plutarco

El tema de cómo se construyen y en que consisten los índices y las escalas tiene una relación directa con la respuesta a la interrogante: ¿cuántas preguntas son necesarias para medir un concepto latente?.

A continuación desarrollaremos este tema. No obstante, dado el carácter general de este capítulo, sólo trataremos de la construcción de índices y escalas de manera sumaria, por lo que quienes requieran información más precisa deberán acudir a la literatura especializada (ver por ejemplo: Martínez Arias, 1995; Muñiz 1998; Summers 1982).

Para responder a la interrogante sobre cuantas preguntas se deben utilizar para medir un concepto, debemos primero estudiar los tipos de errores de medición que podemos cometer cuando utilizamos sólo una pregunta para medir una variable.

¿Podemos utilizar sólo una pregunta para medir un concepto?. Claro que sí. Tal como explicamos anteriormente, cuando medimos un concepto muy simple, podemos formular una sola pregunta para hacerlo.

No obstante, confiar en una sola pregunta tiene al menos dos debilidades: por un lado, es perfectamente posible que las respuestas de las personas sean influidas por el fraseo específico que hemos utilizado para formular la pregunta o las alternativas de respuesta. Por ejemplo, no será igual preguntar: "¿estaría usted dispuesto a votar por el candidato presidencial que proponga la Concertación?", que: "¿estaría usted dispuesto a votar por el candidato presidencial que proponga el gobierno?", o "¿estaría usted dispuesto a votar por el candidato presidencial que propongan los partidos políticos gobernantes?". En cada uno de estos casos, la connotación de las palabras "Concertación", "gobierno" o "partidos políticos gobernantes", influye en las respuestas.

En segundo lugar, resulta evidente que mientras más complejo es el concepto que necesitamos medir, más complicado se volverá confiar en una sola pregunta. El sólo hecho de que hablemos de un concepto complejo implica la necesidad de dividir este en varios sub conceptos y preguntar por cada uno de ellos, por lo que al menos necesitaremos una pregunta por cada sub concepto.

No obstante, incluso en el caso de medir un subconcepto se aplica lo señalado anteriormente: confiar en una sola pregunta implica estar expuesto a las influencias del fraseo utilizadas. ¡Incluso el orden en que se colocan las alternativas de respuesta a una pregunta influye en los porcentajes de respuesta que cada alternativa obtiene!.

¿Cómo evitar este problema?. Evidentemente, utilizando varias preguntas para medir un solo concepto o subconcepto.

La idea es que si bien cada una de las preguntas está influida en algún sentido por las características específicas de su redacción, algunas de ellas distorsionarán la opinión de las personas hacia un sentido, mientras otras lo harán en la dirección opuesta, por lo que una medición conjunta (un promedio, por ejemplo), será una mucho mejor estimación de la opinión, actitud, conocimiento o actitud de las personas.

En un ejemplo sencillo: si en una prueba de matemáticas de 60 preguntas un alumno muy capaz se confunde por la redacción de la pregunta y falla accidentalmente una respuesta, eso será controlado por su rendimiento en las otras preguntas, en alguna de las cuales quizás acertado a algún tema que no conocía en detalle también por accidente. En definitiva, el puntaje que obtenga esta persona en los 60 ítems será mejor estimador de su conocimiento global que sus resultados en cada una de las respuestas por separado.

En realidad, dados los múltiples errores que se pueden cometer al medir un concepto con una sola pregunta (desde errores de digitación, hasta de comprensión de la pregunta), el utilizar múltiples preguntas para medir un concepto debiera ser un deber de primer orden al construir un cuestionario.

No obstante, opuesto a este óptimo encontramos criterios de factibilidad. Muchas veces queremos medir tantas cosas en un solo cuestionario (mal que mal cuesta obtener los recursos para realizar una investigación de este tipo), que para no cansar en exceso a los encuestados, debemos utilizar muy pocas preguntas incluso para medir conceptos complejos.

¿Cuál es el mínimo y el óptimo entonces?. El mínimo absoluto es realizar al menos una o dos preguntas por cada subconcepto (de un concepto complejo) o una o dos preguntas por cada concepto simple. El

óptimo es tener claro de antemano cuales son los conceptos centrales del estudio, aquellos que es necesario medir con mayor precisión, y en esos "invertir" el máximo número de preguntas posibles, dejando que otros conceptos menos relevantes (variables asociadas, variables de identificación, conceptos muy simples), sean medidos con sólo una o dos preguntas.

Entonces, ¿qué es un índice y en que se diferencian éstos de las escalas?.

Un índice, tal como hemos señalado anteriormente, es una formula que combina en forma más o menos arbitraria un conjunto de preguntas con el objeto de producir una sola puntuación que indicará el grado en que los sujetos de estudio poseen un concepto latente.

¿Porqué decimos que es una formula más o menos arbitraria?, porque para la construcción de índices no existen reglas rígidas, quedando las decisiones a tomar en manos del investigador.

Por ello, si dispongo de tres preguntas de respuesta verdadero/falso para medir el nivel de conocimientos o actitud de un sujeto, perfectamente puedo construir un índice que asigne un punto a acertar o responder en un sentido la pregunta 1, dos puntos a acertar o responder en el mismo sentido la pregunta 2 y tres puntos a responder acertadamente o en ese mismo sentido la pregunta 3. En este caso, la persona que tenga en alto grado la actitud medida, o acierte a las tres preguntas obtendría un máximo de 6 puntos. Por otro lado, también el investigador podría haber decidido asignar homogéneamente un punto a cada pregunta. En este segundo caso quien acierte a las tres preguntas tendría un máximo de tres puntos.

En síntesis, la decisión de la forma concreta que tomará un índice es atribución del investigador, pero... ¿existen reglas o normas que regulen sus elecciones?.

Existen dos orientaciones generales que tienden a guiar las decisiones del investigador:

- a) Con relación al peso que se debe asignar a cada pregunta, se asume que si no hay ninguna razón teórica o empírica para asignar más peso a una pregunta o a otra, lo mejor es que todas pesen igual (por un asunto de prudencia mínima, si no hay más información, es mejor abstenerse de cálculos muy originales).

Por el contrario, si existe reflexión teórica que indica que un tema es más relevante que otro en la conformación del concepto a medir (digamos, una pregunta tienen mayor grado de dificultad teórica o la calidad de la vivienda depende más de su materialidad y servicios básicos que de los servicios comunitarios en los alrededores), se debe dar más peso a las preguntas que miden esos temas cruciales que al resto de las preguntas.

De la misma manera, si existe información empírica que muestre que un tema es más relevante para la conformación de un concepto para la población, también se puede dar más peso a esas preguntas. Por ejemplo, para medir el nivel de estrés a que han estado sometidas las personas en un lapso de tiempo, se pregunta por la ocurrencia o no de una serie de eventos vitales que se sabe que son causantes de estrés (pérdida de trabajo, cambio de casa, muerte de algún familiar, gastos excesivos en las vacaciones, etc.). Para poder ponderar, es decir, determinar cuantos puntos de estrés dar a cada uno de estos eventos, se han realizado encuestas en que las personas califican estos hechos de

acuerdo a su grado de importancia. Por ello, en estos índices se asigna más puntuación a la muerte de un familiar que al cambio de casa, por ejemplo.

También existen procedimientos de análisis estadístico multivariado, como la regresión múltiple, análisis discriminante o análisis factorial, que ayudan a determinar el peso que tiene cada pregunta en un concepto global.

En cualquier caso, si el investigador desea ponderar diferencialmente sus preguntas, debe justificar teórica o empíricamente su decisión de manera que sus colegas puedan juzgar la calidad de sus mediciones.

- b) Un criterio relevante a la hora de decidir que preguntas incluir en un índice (si es que tenemos muchas posibles), es la consistencia interna de las respuestas de las personas. En términos coloquiales: si estamos midiendo un solo concepto, se espera que las preguntas que lo midan se encuentren relacionadas entre sí (mal que mal todas están midiendo distintas partes de lo mismo). Por ello, una forma de decidir que preguntas no se deben incluir en un índice usualmente consiste en realizar un análisis del grado de asociación que tienen las preguntas entre sí.

Esto se puede realizar por medio de análisis factoriales, análisis de matrices de intercorrelación, análisis de grados de asociación, etc. En todos estos casos la idea de fondo es siempre la misma: si una pregunta cualquiera no se encuentra suficientemente relacionada con las demás, debe ser eliminada del índice ya que hay fuertes sospechas de que este midiendo otro concepto y no el buscado.

Este procedimiento se basa en el supuesto de que la mayor parte de las preguntas efectivamente miden el concepto buscado y que son las diferentes las que no lo hacen. Si ocurriese que la mayor parte de las preguntas midieran otro concepto y fueran las menos las que acertaran a lo que queremos medir, ¡serían esas las eliminadas!

Otro tema de la construcción de índices tiene que ver con que al igual que es posible construir puntuaciones globales para los conceptos complejos, también pueden construirse puntuaciones intermedias para cada subconcepto (si se dispone de varias preguntas para medir cada uno de éstos), generándose así "subíndices", que es posible combinar para construir el índice global.

Finalmente, ¿en que se diferencian los índices y las escalas?.

En términos generales, las escalas son índices, sólo que las primeras se rigen por reglas de construcción fuertemente estructuradas, que determinan tanto el tipo de preguntas que se pueden formular, el tipo de alternativas de respuesta que se puede incluir y el análisis estadístico que hay que realizar para su construcción definitiva.

Estamos en presencia entonces de "índices formalizados".

Esto implica que hay tantos tipos de escalas, como procedimientos de construcción formalizados se han desarrollado en la literatura y en la práctica de la investigación social.

El más conocido y difundido de los procedimientos de construcción de escalas es el llamado "escala sumativa" o también "Escala Likert", ya que fue diseñado por Rensis Likert en 1932.

Esta escala se construye redactando una serie más o menos larga -entre 30 a 60 usualmente- de afirmaciones (no interrogativas), frente a cada una de las cuales el sujeto debe señalar su grado de cercanía.

Si bien se utilizan varios formatos de respuesta (verdadero/no sé/falso, siempre / a veces / casi nunca / nunca, por ejemplo), el más socorrido es el que establece 5 grados de acuerdo con las afirmaciones: a) muy de acuerdo, b) de acuerdo, c) indeciso, d) en desacuerdo y, e) muy en desacuerdo.

Estas distintas afirmaciones se puntúan del 1 al 5, respetando siempre una misma dirección. Por ejemplo, si para una frase que indica un alto nivel de "intolerancia ante la homosexualidad" (como por ejemplo: "debería prohibirse que las personas homosexuales aparecieran en televisión, porque pueden confundir a los niños") asignamos 5 puntos para la alternativa "muy de acuerdo", 4 a la alternativa "de acuerdo", 3 a "indeciso", 2 a "en desacuerdo" y 1 a "muy en desacuerdo", frente a una frase que indicara un bajo nivel de intolerancia (como por ejemplo: "las personas homosexuales son iguales a cualquier otra persona"), deberíamos asignar las puntuaciones exactamente al revés.

De esta forma, un mayor puntaje indicaría mayor nivel de intolerancia y un menor puntaje, menor nivel de esa variable. Por esa razón, se utiliza como indicador del nivel de intolerancia de las personas la suma de puntajes obtenidos en todas las preguntas.

Nótese que para que el sistema funcione, sólo se pueden incluir en el instrumento afirmaciones polares, es decir, que indiquen alta intolerancia o baja intolerancia. Las afirmaciones neutras o ambiguas no sirven, ya que estar muy en desacuerdo con ellas no permite saber en que polo se ubica el sujeto (si yo formulase la afirmación "la homosexualidad tiene cosas buenas y malas", estar muy en desacuerdo con esa frase podría indicar que soy muy intolerante con la homosexualidad, como que soy muy tolerante).

No obstante esta necesidad de frases polares, debemos tener en cuenta al redactar las frases una recomendación que señalamos anteriormente: las frases no deben ser tan extremas como para que sea imposible (o extremadamente raro) estar totalmente a favor con ella: las frases deben permitir discriminar a las personas y, por tanto, deben ser capaces de separarlas en grupos (por ejemplo, no incluya frases del tipo "odio tanto a los homosexuales que creo que deberían ser públicamente azotados". Si bien esa frase es polar, esperamos que sea demasiado extrema como para que alguien este a favor de ella).

Existen dos razones para comenzar con un número tan grande de ítems al construir una escala Likert: a) medir adecuadamente los varios subconceptos que puede tener un concepto complejo, b) eliminar de la sumatoria global (y de la escala definitiva) aquellos ítems que no funcionan demasiado bien.

Esto implica que parte de los procesos de construcción de una escala likert está constituido por procedimientos de eliminación de ítems inadecuados. Básicamente se eliminan:

- a) Aquellos ítems que poseen una consistencia interna demasiado baja, es decir, tal como ya hemos visto, que tienen bajas correlaciones con los otros ítems, medido esto usualmente a través de la correlación entre el ítem y la puntuación global.

- b) Aquellos ítems que no discriminan suficientemente entre los sujetos, es decir, que concentran inadecuadamente las respuestas en pocas alternativas. Esto se mide usualmente comparando la media de respuestas de los sujetos al ítem con la desviación estándar de las respuestas.
- c) Aquellos ítems que de ser incluidos en la escala final disminuirían la fiabilidad del instrumento (explicaremos esto más adelante, pero por ahora podemos adelantar que esta propiedad se mide con un índice especializado: el Alfa de Cronbach).

Entonces, luego de este proceso de eliminación de ítems, los ítems sobrevivientes constituyen la escala final, y es sólo la sumatoria de éstos la que constituye la puntuación de los sujetos en el concepto medido. Usualmente estos procesos de eliminación excluyen entre el 20% y el 40% de los ítems originales.

Si bien actualmente están apareciendo otros procedimientos de la Teoría de Respuesta al Ítem (TRI) para construir las puntuaciones globales de las Escalas Likert que no se basan en la simple sumatoria de las puntuaciones obtenidas en cada uno de los ítems, sigue siendo mayoritario aún el procedimiento tradicional.

Por otro lado, ya hemos dicho que la escala Likert sólo es uno de los procedimientos de construcción de escalas disponibles. Otros procedimientos utilizados con cierta frecuencia (que no desarrollaremos en detalle) son:

- **Las Escalas de Intervalos Aparentemente Iguales o Escalas Thurstone:** Estan constituídas por alrededor de 22 afirmaciones, de entre las cuales el sujeto debe escoger las que más se aproximan a su opinión.

La principal diferencia entre esta escala y la Likert es que acá las personas deben elegir las afirmaciones de su preferencia, no reaccionar a cada una de ellas con su grado de cercanía. Por ello, la actitud de las personas es definida por las frases que selecciona, no una sumatoria de respuestas.

Además, este procedimiento requiere un previo estudio (con sujetos jueces) que asigne a cada afirmación una puntuación actitudinal, lo que lo hace mucho más laborioso que el procedimiento Likert.

Si bien diversos estudios han afirmado que la validez de las escalas Thurstone es mayor que la de las escalas Likert, estas últimas son lo suficientemente buenas como para justificar su utilización dados los menores costos de tiempo y recursos que exigen (Summers, 1982).

- **Las Escalas de Distancia Social:** Son instrumentos diseñados para medir la distancia o cercanía de interacciones que una persona desea tener con un determinado grupo social (digamos los inmigrantes peruanos). Estas escalas están compuestas por unas 6 ó 7 preguntas, cada una de las cuales debe ser respondida por el sujeto afirmativa o negativamente (Goode y Hatt, 1967).

Cada una éstas interroga al encuestado por la posibilidad de tener un determinado contacto social con el grupo objeto de estudio. Además, las preguntas están diseñadas de manera crecientemente cercana de manera que si bien la primera casi no implica contacto social con el grupo en cuestión

(por ejemplo la pregunta: "¿aceptarías que inmigrantes peruanos pudieran vivir en Chile?"), paulatinamente el encuestado se acerca a preguntas que implican mayor nivel de contacto (por ejemplo: "¿aceptarías tener como vecinos a inmigrantes peruanos?" O incluso: "¿dejarías que tu hijo o hija se casara con un inmigrante peruano?").

Evidentemente si una persona responde afirmativamente a todas las preguntas es que desea mayor nivel de contacto con el grupo objetivo que si responde negativamente, por lo que la sumatoria de respuestas "no" constituye un indicador de la "distancia social" del sujeto hacia el grupo.

- **Los Diferenciales Semánticos:** Son instrumentos diseñados para medir el significado o contenido emocional que le genera a las personas un determinado concepto (Summers, 1982).

Están constituidos por alrededor de 7 pares de adjetivos polares, entre los cuales se incorpora un continuo numérico. Se espera que el sujeto lea el concepto y responda con un número a cada par de adjetivos, de acuerdo a cuan cerca se siente él a esos adjetivos, al pensar en el concepto.

Por ejemplo, si quisiéramos medir el contenido emocional que tiene para las mujeres de entre 30 y 40 años el fútbol chileno podríamos construir el siguiente (breve) diferencial semántico:

		Fútbol Chileno								
Entretenido	1	2	3	4	5	6	7		Aburrido	
Activo	1	2	3	4	5	6	7		Pasivo	
Feo	1	2	3	4	5	6	7		Bonito	
Sucio	1	2	3	4	5	6	7		Limpio	
Ordenado	1	2	3	4	5	6	7		Desordenado	

Como hemos dicho, se espera que cada encuestada circule un número en cada par de adjetivos de acuerdo a cuan cerca está su percepción del fútbol chileno de cada uno de los polos.

Este tipo de escala permite análisis estadísticos muy interesantes (perfiles, gráficos tridimensionales, etc.), pero usualmente sólo funciona bien en instrumentos autoadministrados o aplicados en sujetos con un nivel educativo alto.

## 6.- El cuestionario como conjunto:

*Hay que desconfiar de los ingenieros.  
Se empieza con la máquina de coser  
y se termina con la bomba atómica"*  
Oppenheimer

Luego de la formulación de las preguntas y respuestas y determinar cuales de ellas se constituirán en índices y escalas, se debe enfrentar el tema de como construir el instrumento completo.

Evidentemente, un cuestionario es mucho más que la suma de las preguntas que lo componen, por lo que la forma en que ellas están combinadas influirá en la calidad de los resultados obtenidos.

Lo primero que debemos tener claro es el orden en que se deben colocar las preguntas en el cuestionario. Para esto debemos considerar que un cuestionario puede dividirse en las siguientes secciones:

- a) **Presentación:** en esta parte o portada, se explica sumariamente a los encuestados el sentido y organización que realiza el estudio (el que no tiene porqué ser muy explícito si no es bueno que las personas conozcan exactamente el tema a estudiar), se recalca la necesidad de que las personas respondan seriamente, se enfatiza que no hay respuestas correctas ni incorrectas, por lo que todas las respuestas son igualmente válidas (a excepción de los cuestionarios de aptitud o conocimientos), se explica a las personas el tiempo promedio que tomará la encuesta y se pide su colaboración.

Si en esta primera parte se obtiene la confianza del encuestado, la validez de las respuestas aumentará.

- b) **Preguntas de identificación o demográficas:** casi siempre conviene comenzar la aplicación del cuestionario con preguntas simples y fáciles de responder, por lo que preguntas como edad, lugar de residencia, trabajo, sexo, nivel educativo, tienden a ser incluidas en esta parte.

Las únicas preguntas que no se deben incluir acá son las preguntas de identificación que pudieran generar resistencias por parte del sujeto, por ejemplo, las que dicen relación con el nivel socioeconómico, sobre todo si se pregunta por ingresos (pregunta usualmente incómoda para las personas), o por bienes del hogar (pregunta que genera inseguridad por posibles robos). Estas preguntas deben ser incluidas al final del instrumento.

- c) **Preguntas temáticas:** Acá se deben incluir las preguntas sobre el tema que se pretende estudiar. El orden en que estás preguntas se pueden incluir debe facilitar la tarea de responder al cuestionario, por ello, se debe tratar de crear una secuencia lógica de temas que, dentro de lo posible, los articulen desde lo más simple o menos implicador, hacia los temas más complicados de responder por sus connotaciones éticas o afectivas.

También es importante disminuir en lo posible los cambios de estilos de preguntas, agrupando en un solo lugar del cuestionario las preguntas similares (por ejemplo, se debe tratar de colocar juntas todas las preguntas que se responden con verdadero o falso).

Otro tema del que nos debemos preocupar es el llamado "efecto de halo". Con este concepto denominamos el impacto que produce una pregunta (y tener que responder a ella), sobre las respuestas a las siguientes preguntas.

Por ello, se debe tratar que las preguntas que pudieran influir en las restantes sean efectuadas hacia el final del cuestionario, de manera de disminuir su influencia. Nótese también que el efecto de halo disminuye con el correr de las preguntas, por lo que otra manera de atenuar su influencia es preguntar sobre dos temas complejos en lugares muy distintos del cuestionario.

Finalmente, se debe tomar en cuenta que la calidad de las respuestas de los encuestados está directamente relacionada con el nivel de cansancio que ellos llevan al responder, por lo que las preguntas que están hacia el final del cuestionario (si este es muy largo) tendrán más error que las primeras. Por ello, no se deben dejar las mediciones en que se requiere más precisión o las pregunta

más difíciles de responder para el final. En otras palabras, el máximo rendimiento de las personas se produce en el centro de la encuesta, por lo que allí se deben incluir las preguntas cruciales del estudio.

- d) **Preguntas de identificación intrusivas:** con esta etiqueta denominamos a todas aquellas preguntas de identificación que generan más inseguridad en los encuestados (usualmente las referidas a su situación socio económica), por lo que este es el espacio en que se tienden a preguntar.

Se espera que luego de un cierto tiempo de entrevista, ya exista cierta confianza entre el entrevistador y el entrevistado, de manera que se obtengan respuestas más válidas. Además, si el sujeto se molesta y no desea seguir con la encuesta, al menos ya se ha obtenido gran parte de la información buscada, así que no se ha perdido todo el trabajo.

- e) **Cierre:** al finalizar la encuesta debemos agradecer a las personas su participación y, si es que el tema estudiado es complicado para los sujetos, permitir que ellas agreguen libremente lo que deseen a lo ya preguntado (esto es muy pertinente, por ejemplo, cuando se investiga sobre hechos traumáticos o estresantes).

Además del orden de las preguntas, otro tema que debemos tener siempre en mente cuando construimos un cuestionario (y cuando redactamos las preguntas y respuestas), es la forma en que esperamos sea aplicado el cuestionario.

Por ejemplo, no es lo mismo construir un cuestionario para ser aplicado en encuestas telefónicas (en donde tendremos que preocuparnos especialmente de que las preguntas sean breves, el cuestionario corto, y que existan pocas alternativas de respuesta a cada pregunta), que otro diseñado para ser auto administrado (en donde lo crucial será redactar explicaciones lo más claras posibles y evitar en lo posible los saltos y preguntas filtro -preguntas que si son respondidas en cierta dirección, hacen que otras preguntas no sean necesarias o pertinentes, por lo que exigen que el sujeto se "salte" dichas preguntas-).

Por otro lado, en las encuestas diseñadas para aplicación cara a cara con un encuestador entrenado y provisto de tarjetas, podremos incluir preguntas más complejas, saltos y más alternativas de respuesta, pues la presencia del encuestador nos asegura un nivel más alto de comprensión del instrumento (aunque a mayores costos por cada cuestionario aplicado, claro).

También es importante considerar globalmente cuando se está construyendo el cuestionario el nivel educativo de las personas a las que se pretende aplicar éste. Por ejemplo, si las personas a las que va dirigido el instrumento son estudiantes universitarios se puede incluir preguntas más complejas (como diferenciales semánticos por ejemplo), mientras que si estamos entrevistando a personas de sectores populares, las preguntas deben ser más simples y con menos opciones de respuesta.

Lo mismo se puede decir con relación a la longitud global del cuestionario. Si bien ningún instrumento debiera tomar más 45 minutos para su aplicación, el tiempo disponible varía fuertemente de acuerdo al nivel educativo de los respondientes. Por ejemplo, si aplicamos el cuestionario a estudiantes, ellos estarán tan acostumbrados a este tipo de experiencias (las pruebas son cuestionarios de conocimientos) que no será problema exceder un poco el límite máximo marcado arriba. Lo contrario ocurre al encuestar a una población que ha dejado de estudiar hace mucho tiempo o ha tenido breves

experiencias educativas formales. Si un cuestionario toma mucho tiempo de responder o resulta muy cansador para esta población, la calidad de las respuestas disminuirá y podremos confiar mucho menos en las mediciones que hayamos realizado.

Por ello, cuando nuestro instrumento esté diseñado para ser utilizado sobre un universo de personas heterogéneo o en diferentes formas de aplicación, debemos pensar siempre en que éste sea posible de aplicar con comodidad en la forma de aplicación más simple y para el encuestado de menor nivel educativo.

En términos generales, cada cuestionario debe ser diseñado para su aplicación en determinado contexto y por ello es razonable que el constructor del instrumento explicité estas condiciones en una ficha técnica que acompañe al cuestionario, de manera que quien desee aplicarlo en otras situaciones o poblaciones tenga claro que deberá introducirle algunas modificaciones.

Finalmente, debemos saber que la calidad de las respuestas a un cuestionario depende también del interés que el tema despierte en las personas. Si los entrevistados se sienten entretenidos o motivados respondiendo el cuestionario podremos confiar mucho más en sus respuestas. El desafío es entonces también como construir instrumentos y preguntas que sean entretenidas y pertinentes para nuestros encuestados.

## **7.- La Validez y fiabilidad de la investigación con cuestionarios**

*"Cierta grado de desilusión ayuda  
a los hombres a ser prudentes".  
Macaulay Trevelyan*

Una pregunta que muy frecuentemente formulan las personas que están comenzando a aprender metodología cuantitativa o que son escépticas frente a esta estratégica de investigación es: ¿pero... miden realmente los cuestionarios conceptos importantes para las ciencias sociales?.

Responder a esa pregunta nos remite a dos conceptos clásicos y centrales para esta línea de investigación: los conceptos de validez y fiabilidad.

Como muchas cosas en la vida, ambos conceptos pueden ser definidos desde distintos puntos de vista que destacan diferentes aspectos de ellos. Por ejemplo, la fiabilidad puede ser entendida como una propiedad de los instrumentos de medición consistente en que éstos, aplicados dos veces a los mismos sujetos (quienes no han cambiado en nada entre ambas aplicaciones), deben obtener el mismo valor. Esto implica definir la fiabilidad como la estabilidad de nuestras mediciones. En un ejemplo extraído de otras ciencias, una regla de medir fiable sería aquella que aplicada a medir dos veces el mismo objeto, nos entregara siempre el mismo valor como resultado. ¡Evidentemente no confiaríamos en una regla que primero nos dijera que un libro mide 15 cm y después de cinco minutos nos diera 20cm!.

A pesar que esta propiedad parece obvia, toda vez que las ciencias han comenzado a medir con cierto nivel de exactitud, han tenido que reconocer que toda medida (incluso la realizada con los instrumentos más sofisticados) tiene un cierto grado de falta de fiabilidad. En otras palabras, si una persona se pesa dos veces en la misma pesa de precisión, posiblemente el valor obtenido en ambas mediciones, si bien será gruesamente el mismo, tendrá ligeras diferencias.

Por ello, también es posible definir la fiabilidad como el porcentaje de error de medición que tiene todo instrumento aplicado en una determinada situación.

Por otro lado, con relación a la validez, la definición clásica dice que un instrumento es válido si efectivamente mide lo que dice que mide (en otras palabras, si yo digo que estoy midiendo autoritarismo, mida efectivamente eso y no otro concepto).

No obstante, también es posible pensar un poco más ampliamente la validez como una propiedad no sólo de los instrumentos, sino más bien de las interpretaciones que hacemos de sus resultados. Desde ese punto de vista podemos entender la validez como el grado en que son legítimas las interpretaciones que deducimos de un instrumento (Navas Ara, 2001).

En otras palabras, la validez estaría también relacionada con la interpretación de los datos, de manera que una investigación será más válida si las interpretaciones que hagamos de ella se sustenten en lo que efectivamente medimos.

Al igual que en el caso de la fiabilidad, la validez de un estudio es siempre parcial. No importa cuanta voluntad y esfuerzo pongamos al realizar una investigación, siempre estaremos midiendo parcialmente otros conceptos distintos a los buscados y/o haciendo interpretaciones leve o parcialmente ilegítimas.

El objetivo entonces es ser capaz de producir investigaciones lo más fiables y válidas posibles. Desde éste punto de vista resulta productivo conocer cuáles son las principales limitaciones a la validez que enfrentamos al medir utilizando cuestionarios y como controlarlas.

Nos parece que las limitaciones se pueden agrupar en los siguientes tipos:

- a) **Confiar en Autoreportes:** Nunca debemos perder de vista que cuanto utilizamos cuestionarios para medir conceptos, la mayor parte de las veces no registramos lo realmente sucedido o deseado por el sujeto, sino lo que el nos quiere reportar.

Esta limitación implica que nuestras mediciones están sujetas a un porcentaje de error producto de esta mediación. No sólo se trata de que el sujeto voluntariamente nos mienta (que igual puede ser un problema), sino que, como ya se ha demostrado, el recuerdo es socialmente determinado, por lo que lo reportado por el sujeto, independiente de su voluntad de decirnos la verdad, estará distorsionado por su situación y perspectiva actual.

Por otro lado, cuando las personas responden a un cuestionario también lo hacen buscando quedar bien ante sus interlocutores (encuestador, investigador, etc), por lo que distorsionan sus respuestas en este sentido. Este problema afecta especialmente a los conceptos que implican valores morales en donde existe un "deber ser" socialmente legitimado (por ejemplo, cuando medimos discriminación hacia la mujer debemos tomar en cuenta que en nuestra sociedad está mal vista esta actitud, independiente de cuan discriminadores y machistas seamos realmente).

Esta limitación afecta especialmente a las mediciones de fenómenos subjetivos, pero incluso en la medición de hechos (que están "anclados" en un recuerdo) las personas pueden distorsionar la realidad. Por ejemplo, en un estudio del Centro de Investigaciones Sociológicas de España se

preguntó a los sujetos por quien habían votado en las últimas elecciones, encontrándose con que casi el 65% de la muestra declaró haber votado por el candidato finalmente ganador, respuesta que no coincidía con el hecho que este candidato sólo obtuvo alrededor del 50% de los votos.

- b) **Influencia de las características particulares de las preguntas:** Cuando redactamos una pregunta, no importa cuanto nos esforcemos por redactar preguntas no sesgadas valóricamente por las palabras que utilicemos: igual el fraseo de la pregunta, influirá en las respuestas.

Incluso el orden en que se colocan las respuestas en una pregunta cerrada influye en las respuestas que dan los sujetos.

- c) **Adaptación de los sujetos a las respuestas:** En muchas preguntas las personas que responden no encontrarán entre las alternativas de respuesta que les presentamos, las alternativas que ellos desearían haber marcado.

Por ejemplo, puede que una parte importante de las personas sienta que las alternativas "si" o "no" no son suficientes para reflejar su verdadera conducta teniendo que adaptar lo que ocurrió realmente a esas opciones polares. Esta adaptación constituye un error de medición, pues no se recoge lo que ocurrió, sino sólo una aproximación.

- d) **Cristalización:** Frente a algunos temas las personas encuestadas no tienen realmente una opinión. Quizás incluso nunca se habían planteado el tema. No obstante, si se les pregunta sobre ello, las personas mayoritariamente construirán una respuesta.

El problema es que el significado de estas respuestas sobre temas poco pertinentes para las personas no es el mismo que tienen las respuestas que dan las personas a un tema largamente debatido en la opinión pública. Usualmente esta "intensidad" de la opinión no es realmente conocida por el investigador, el que tiende a interpretar de igual manera las respuestas a una pregunta poco pertinente, como las respuestas a preguntas altamente implicadoras.

- e) **Adopción de papeles:** El proceso de aplicar un cuestionario, aún en el caso que éste sea autoadministrado, constituye una relación social y como tal está sujeto a las interpretaciones que hagan de él las personas que participan (König, 1973).

En este sentido, el papel en que se ubique el sujeto al responder influirá en sus respuestas. Por ejemplo, si la persona sume el rol de asistido frente a un posible benefactor, tenderá a acentuar sus carencias, con el fin de maximizar sus posibilidades de recibir recompensas. Si por el contrario, el sujeto asume el rol de ciudadano, acentuará su toma de posición en muchos temas que usualmente no le interesan demasiado.

Lo que queremos decir es que los sujetos asumen roles ante el proceso de investigación y esos roles asumidos influirán en las respuestas. Esta influencia quizás no será tanta como para modificar las opiniones de las personas, pero si para acentuarlas en alguna dirección.

- f) **Ambigüedades en la valoración:** Incluso cuando disponemos de un instrumento complejo que nos entrega un resultado altamente desagregado de las puntuaciones de los sujetos frente a un determinado concepto (por ejemplo, nos dice que el nivel de autoritarismo de una persona es 24), lo

que busca el investigador y lo que espera el público en general es que nuestras investigaciones sean capaces de valorar correctamente a las personas y grupos. En otras palabras, lo que estamos buscando no es el número asignado a un sujeto, sino una etiqueta: "esta persona es muy autoritaria" o "esta persona no es autoritaria" y eso es muy difícil de especificar solamente a partir de la puntuación obtenida.

Algo muy similar ocurre cuando queremos interpretar las respuestas de un grupo hacia una pregunta: por ejemplo, si el 35% de los miembros de la elite chilena pensarán que su aporte al país está muy poco reconocido, ¿significaría eso que en la elite chilena existirían importantes grados de frustración?, ¿Estaríamos frente a una crisis de liderazgo?, o por el contrario, ya que el 65% de dicha muestra no señalaría esta respuesta ¿deberíamos considerar que la elite chilena esta satisfecha de la forma en que las personas reconocen su actividad?.

En otras palabras, aunque nuestras mediciones sean muy precisas sobre un tema, la verdad es que usualmente cuesta mucho interpretar tanto una respuesta particular como una puntuación por si misma.

Esta ambigüedad en la valoración de los resultados usualmente conduce a que iguales resultados numéricos sean interpretados divergentemente por diversos investigadores.

- g) **Omisión de preguntas claves:** Es muy habitual que una vez que hemos comenzado el análisis de datos, nos damos cuenta que una pregunta muy relevante para dilucidar interpretaciones divergentes o arrojar luz sobre un tema que aparece como muy crucial, no ha sido preguntada.

Realmente es muy difícil que el investigador sea capaz de predecir que preguntas adquirirán relevancia en el análisis de datos, y por tanto, también le será muy difícil cubrir todos los temas posiblemente iluminadores. Hay que convivir con un cierto margen de preguntas cruciales no formuladas.

- h) **Incomprensión de las preguntas:** Uno de los problemas más frecuentes que disminuyen la validez de las respuestas de las personas ante los cuestionarios ocurre cuando un porcentaje importante de las personas no comprenden la pregunta, ya sea por su estructura o léxico.

Contrariamente a lo que uno podría pensar, es muy usual que estas personas no consulten al encuestador ni omitan la respuesta, sino que efectivamente escojan alguna respuesta (más o menos al azar), lo que puede llevarnos a interpretaciones equivocadas. Por ello, resulta crucial redactar las preguntas y sus respuestas de la forma más simple posible.

Además de los problemas anteriores, que disminuyen la validez de las respuestas dadas a un cuestionario, también debemos enfrentarnos a problemas específicos relacionados con la fiabilidad del instrumento, es decir, con su capacidad de obtener siempre el mismo resultado, cuando aplicamos el instrumento a un sujeto que no ha cambiado en la propiedad que deseamos medir.

Nos parece que a este respecto nos enfrentamos especialmente con dos dificultades:

- a) **Impacto de variables externas sobre las respuesta:** En muchas ocasiones las respuestas que los sujetos dan a un cuestionario son afectadas por otras variables (externas al concepto que deseamos medir) y coyunturales, como el estado de ánimo de la persona, sus experiencias recientes, etc.

Si bien el impacto de estas variables es relativamente inevitable y debemos suponer que incide en un cierto porcentaje de todo tipo de preguntas, hay algunas formas de preguntar que se ven más distorsionadas que otras.

Según investigaciones anteriores, son las preguntas que miden fenómenos afectivos o emocionales y las preguntas que tienen poco especificadas las alternativas de respuesta, las más impactadas. Esto hace que, en general, resulten menos fiables las respuestas a las escalas intervalares como el diferencial semántico que las respuestas a preguntas de escalas Likert o relativas a hechos.

- b) **El componente de azar que incluye la selección de respuestas:** Cuando una persona responde una pregunta de un cuestionario no sólo lo hace basándose en un cierto hecho o en el grado en que él está de acuerdo con una opinión o afirmación, sino también en función de un cierto grado de azar.

En otras palabras, si suponemos que el sujeto no encuentra exactamente una respuesta que sea perfectamente adecuada a su opinión, recuerdo o percepción, él intentará encontrar rápidamente otra respuesta cercana a su opción ideal. Nuestro supuesto es que el realizar este proceso en forma rápida implica que el sujeto tiende a seleccionar en forma relativamente aleatoria entre alternativas que se encuentran aproximadamente equidistantes de su opinión verdadera.

Esto implica que si esta misma pregunta se la presentamos a este mismo sujeto en otro momento en el tiempo (ojalá cuando ya haya olvidado su respuesta anterior), el podría marcar otra respuesta (aproximadamente tan cercana a su opinión como la primera), sólo por azar.

Este tipo de error ocurre especialmente en aquellas preguntas en que sea más probable que el sujeto no encuentre su verdadera opinión entre las alternativas de respuesta y en aquellas en que el tema le resulte poco pertinente (y por tanto, le dé más o menos lo mismo la respuesta escogida).

Estas son las principales amenazas a la validez y fiabilidad que encontramos en un cuestionario. La pregunta que trataremos de responder a continuación es: ¿Cuáles son los principales síntomas que pueden indicarnos la existencia de problemas en una pregunta?.

Este tema es muy importante, ya que si contamos con medios para detectar los problemas de nuestras preguntas, podremos eliminar o modificar aquellas que funcionen deficientemente.

Fuera del análisis teórico por parte de jueces expertos que podrían detectar algunas preguntas con problemas, nos parece que existen tres síntomas empíricos que deben hacer que pongamos en sospecha a algunas de ellas:

- a) **La pregunta no discrimina:** Si más del 85% de las respuestas a una pregunta se concentran en una sola de sus alternativas de respuesta, es posible que estemos en presencia de problemas.

Usualmente esa concentración indica que existe algún elemento en la redacción de la pregunta o de sus alternativas de respuesta que ha causado problemas. Había que ver si alguna palabra o formulación ha hecho demasiado deseable o indeseable sus respuestas, o la pregunta es demasiado gratuita, por ejemplo.

Si no creemos que sea eso lo que está ocurriendo, quizás de todas maneras no resulta deseable mantener la pregunta en un índice o escala, ya que estos instrumentos funcionan mejor si todos sus ítems discriminan. No obstante la pregunta puede ser utilizada igualmente en solitario para hablar de los temas en que el universo estudiado es casi totalmente homogéneo, objetivo que podría ser útil en algunos estudios.

- b) **Las respuestas a una pregunta no se relacionan coherentemente con otras preguntas u otra información preexistente:** Si existe una contradicción lógica entre una pregunta y otras del cuestionario, u otra información relevante y que se considera válida (que podría ser de carácter teórico o empírico), podremos poner en duda la validez de las respuestas a la pregunta.

El punto central en esta situación es que realmente confiemos en la información de las otras preguntas o de estudios previos, pues también podría ocurrir que sea esa información la incorrecta.

Además, podrían ser nuestra propia y particular perspectiva la que sostiene que existe incoherencia, pudiendo otra persona encontrar un discurso coherente que justifica ambos resultados (como afirman Berger y Luckman, la coherencia de un sistema está más en los ojos del observador que en las características del sistema).

Por tanto, hay que tener mucho cuidado al descartar información por esta razón. Hay que tener fuertes sospechas en contra de las respuestas a la pregunta.

- c) **Aumento fuerte de las no respuestas o de las respuestas “otra”:** Si bien la mayor parte de las personas prefiere dar cualquier respuesta a omitir la pregunta, o escoger una alternativa de entre las presentadas que escoger la alternativa “otra”, de todas maneras hay un cierto porcentaje de población que se atreve a no marcar ninguna respuesta o acude a esa alternativa.

Por tanto, cuando observemos una cantidad mucho más alta de estas respuestas, podremos sospechar que algo extraño ocurre con la pregunta o sus alternativas de respuesta. Posiblemente hemos omitido alguna opción de respuesta relevante y pertinente para las personas.

No obstante, esta sospecha debe descartar primero otras posibles explicaciones del fenómeno, por ejemplo, también se observa un relevante aumento de las no respuestas en las preguntas relacionadas con temas valóricamente complejos (por ejemplo, la tolerancia hacia la homosexualidad).

En términos generales, siempre hay preguntas que debemos descartar del análisis definitivo de un cuestionario y que hemos detectado por algunos de los procedimientos anteriores. Puestos en una posición pesimista (y realista), esto incluye hasta un 10% de las preguntas de un instrumento complejo.

Ello es lo que justifica formular varias preguntas para medir cada concepto: si alguna debe ser descartada, siempre tendremos otras disponibles para estimar la variable buscada.

No obstante, el ideal es que el cuestionario tenga un máximo de preguntas válidas y fiables. Por ello vale la pena preguntarnos, ¿cómo podemos aumentar la probabilidad de que nuestras mediciones sean válidas y fiables?.

- a) En primer lugar, creemos que seguir las recomendaciones señaladas con anterioridad para redactar las preguntas mejorará la calidad de la medición.
- b) En segundo lugar, volvemos a recomendar el redactar múltiples preguntas para medir cada concepto.
- c) En tercer lugar, un encuestador bien capacitado usualmente disminuye mucho el error de medición al dar al sujeto indicaciones claras y pertinentes cuando éste tiene dudas sobre una pregunta o sus alternativas de respuesta.
- d) También ayudará construir una buena introducción que ayude a generar confianza y responsabilidad en la persona que responda. Si en esta introducción logramos convencerlo que sus respuestas son relevantes, que su anonimato está asegurado y que merece la pena tomarse el tiempo para responder, sin duda mejoraremos la calidad de nuestras mediciones.
- e) Utilizar un pre test cualitativo podría ayudar a mejorar la comprensibilidad general del instrumento. Entendemos por pre test cualitativo uno en que las personas no sólo respondan al instrumento, sino también puedan expresar en forma hablada las dudas e impresiones que tuvieron de él.
- f) Someter el instrumento a jueces también ayudará a que este mejore sus cualidades, siempre una mirada externa podrá resultar aportadora.
- g) No inventar todo el cuestionario, concentrarse sólo en aquellos temas que no han sido medidos con anterioridad y aplicar o adaptar instrumentos pre existentes para medir aquellos temas ya estudiados por otros investigadores. Esto permitirá concentrar los esfuerzos creativos en sólo parte de la tarea y además brindará la posibilidad de comparar nuestros resultados con los de otras investigaciones, lo que casi siempre es muy útil a la hora de interpretar y valorar lo obtenido.
- h) Asegurar la confidencialidad es muy relevante para mejorar la validez de aquellos cuestionarios que miden temas complicados para el sujeto (generalmente asociados a valores o a relaciones sociales cercanas, como los cuestionarios de clima organizacional).

Este aumento de la percepción de confidencialidad se puede lograr disminuyendo las preguntas de identificación al mínimo indispensable (el sujeto puede sospechar de la declaración de anonimato si le hacemos demasiadas preguntas que podrían detectar quien responde), dejando que parte del instrumento lo responda secretamente (sin mediación del encuestador) o creando un diseño de entrevista en que el sujeto entregue anónimamente el cuestionario respondido (con un sistema de urnas, por ejemplo).

- i) Tratar de hacer preguntas que le sean pertinentes e interesantes al sujeto que responde. De esta manera su concentración será mayor y, por tanto, habrá más fiabilidad en las mediciones que obtengamos.

- j) Para asegurarnos que los análisis que realicemos sean efectuados sobre preguntas de buena calidad métrica, deberíamos hacer un chequeo previo de las preguntas, siguiendo los criterios señalados anteriormente.
- k) Finalmente, para mejorar nuestra capacidad de interpretar correctamente y valorar las respuestas de los sujetos, se pueden sugerir los siguientes procedimientos:
- Incluir en el diseño muestral otros grupos originalmente fuera de nuestro interés, con el fin de permitirnos una interpretación comparada (por ejemplo, incluir una pequeña muestra de hombres, a pesar que nuestra investigación sea sobre mujeres).
  - Tratar de comparar con información similar del pasado, de manera de cimentar la interpretación en los cambios en el tiempo.
  - Preguntar sobre temas similares al de interés, con el objeto de comparar entre diferentes temas.
  - Utilizar el mismo cuestionario varias veces, de manera que el uso reiterado le terminen confirmando sentido a las respuestas (como ocurre con las pruebas de selección universitaria, en donde las personas ya saben que significa cada puntaje por sus consecuencias prácticas, aunque no entiendan nada de los procedimientos de transformación por los cuales esas puntuaciones se obtienen).
  - Investigar posteriormente la relación de las respuestas de un cuestionario con las conductas reales de los sujetos o con el significado que quisieron darle las personas a sus respuestas. Lamentablemente esta recomendación implica a veces realizar otra investigación, generalmente de carácter cualitativo.

Fuera de estas recomendaciones de carácter general, se sugiere además chequear la fiabilidad y la validez de nuestras principales escalas e índices siguiendo alguno o algunos de los siguientes procedimientos. Todos estos procedimientos se deben utilizar en un pre test diseñado especialmente con el fin de obtener indicadores empíricos de la fiabilidad y validez de los instrumentos. Dado el alto costo que implican, estas pruebas suelen reservarse sólo para la primera aplicación de escalas que se espera utilizar reiteradamente.

1. Procedimientos para chequear **la fiabilidad** de una medición (índice o escala):

- a) **Repetición de la prueba:** Consiste en aplicar dos veces el mismo instrumento a las mismas personas, dejando pasar un lapso pequeño de tiempo (lo suficiente para que las personas olviden sus respuestas anteriores, pero no tanto como para que les sucedan cambios en el concepto medido).

Usualmente se deja un tiempo de entre 2 semanas y un mes entre ambas mediciones, pero esto depende de cada concepto.

El supuesto es que si hay relación entre las respuestas dadas en la primera y la segunda oportunidad, el instrumento ha demostrado un grado apreciable de fiabilidad. Esta asociación usualmente se mide

a través de índices de correlación (Pearson con frecuencia) y se considera aceptables correlaciones mayores que 0,8.

El problema de esta forma de medir la fiabilidad es que una alta correlación entre medidas también podría deberse a efectos de la memoria o, por el contrario, una baja correlación podría deberse a cambios ocurridos en los sujetos en el intervalo entre ambas mediciones.

- b) **Diseño de formas paralelas:** Consiste en generar dos instrumentos muy parecidos, que midan el mismo concepto. Ambos instrumentos (que no son exactamente iguales, al menos en su redacción) deben ser aplicados en un solo momento al grupo.

El supuesto detrás de este procedimiento es que si las personas responden lo mismo en ambas formas, esto ha sido equivalente a aplicar dos veces el mismo instrumento, por lo que la correlación entre ambas formas nos dará una aproximación a la fiabilidad de cada una de ellas.

Si bien este procedimiento tiene la ventaja de disminuir el impacto de la memoria sobre la prueba de fiabilidad, introduce el problema de la diferencia entre las dos formas, por lo que la falta de relación podría deberse sólo a que ambas formas son muy diferentes y no a falta de fiabilidad de cada instrumento.

Por otro lado, si tratamos de evitar este problema creando dos formas muy parecidas, puede suceder que las preguntas sean tan similares que la correlación entre las dos formas sólo se explique por esta similitud, y no sea adjudicable a la fiabilidad de cada instrumento.

- c) **Consistencia interna:** Otra dificultad que tienen los procedimientos anteriores es que requieren un alto esfuerzo consistente en lograr que las personas respondan dos veces al mismo instrumento, o en crear dos instrumentos cuando sólo necesitamos uno.

Por ello, una solución posible es aplicar un solo instrumento una sola vez, a un solo conjunto de sujetos, pero analizar las respuestas como si correspondieran a la aplicación de dos cuestionarios paralelos. Esto se logra dividiendo el instrumento en dos, calculando una puntuación para cada persona en cada mitad del instrumento y correlacionando ambas mitades. Luego esta correlación se tiende a corregir para evitar la pérdida de asociación producida por hacer el cálculo sólo con la mitad de los ítems.

Si la correlación entre ambas mitades es lo suficientemente alta (0,8 usualmente), se asume que el instrumento es fiable, pues esta correlación nos indica que habría pasado si aplicamos el cuestionario dos veces a un solo grupo.

Producto de su bajo costo (sólo requiere procedimientos de análisis estadístico, no aplicaciones repetidas o formas paralelas), esta forma de estimar la fiabilidad es la más utilizada, lo que ha generado una utilización mayor aún de la escala Likert, ya que ella se presta particularmente bien a su división en mitades.

2. Procedimientos para chequear **la validez** de una medición (índice o escala):

a) **Análisis de Convergencia:** Si la puntuación obtenida por las personas en el instrumento que estamos evaluando correlaciona fuertemente con otras medidas del mismo concepto o conceptos similares, se considera que eso es un indicador de la validez de la medición realizada.

Lo ideal es que las otras mediciones del concepto sean lo más independientes posibles de nuestra propia medición y que se considere (por estudios previos por ejemplo), que ellas son a su vez mediciones válidas.

Estas otras mediciones pueden ser de la más diversa naturaleza: puntuaciones de cuestionarios, información secundaria, registros de hechos, etc.

b) **Análisis de Divergencia:** Al contrario que en el procedimiento anterior, otra forma de validar una medición es mostrar que ella se diferencia de conceptos de los cuales debe diferenciarse. Por ello, si nuestra medición no está correlacionada con variables con las cuales no debe correlacionar, esa información también aportará a la validación del instrumento.

c) **Análisis de Discriminación:** Por otro lado, también se puede aportar evidencias a la validez de un instrumento demostrando que éste es capaz de separar con pocos errores a dos grupos que son realmente diferentes.

Por ejemplo, si aplicamos un instrumento para medir autoritarismo a dos grupos: uno realmente autoritario y otro realmente no autoritario, nuestro instrumento, si es válido, debiera ser capaz de separar efectivamente entre ambos grupos, asignando puntuaciones muy diferentes a ambos.

d) **Análisis de jueces:** Una forma no empírica de afirmar la validez de un instrumento, es someter a éste al juicio de expertos, quienes respaldan, a partir de sus conocimientos previos, que el instrumento es adecuado para medir lo que se desea medir.

e) **Análisis teórico:** Finalmente, también se puede respaldar la validez de un estudio realizando un análisis teórico que justifique porque esas preguntas pueden informar adecuadamente del concepto a medir. Este procedimiento se utiliza especialmente en aquellos instrumentos en que la relación pregunta / concepto, no es tan clara porque se ha optado por utilizar preguntas indirectas para construirlo.

Por ejemplo, la Escala de Autoritarismo de Adorno, incluye algunas preguntas relacionadas con la relación entre el sujeto y sus padres, especialmente en el campo de la obediencia a las normas familiares. Se justifica esta estrategia a partir de la teoría psicoanalítica (Pastor, 1986).

En conjunto, debemos entender que todas estas pruebas nunca demuestran absolutamente la validez de las medidas que tomamos, ya que tanto la validez como la fiabilidad son siempre relativas. Por ello estas pruebas deben considerarse como evidencias de validez y fiabilidad. Si un instrumento acumula muchas de estas evidencias, sin duda tendremos más confianza en las puntuaciones que obtengamos con él.

No obstante, no debemos olvidar que a pesar de obtener buenas evidencias de la validez o fiabilidad de alguna medida y que hayamos tomado todas las providencias para medir de la mejor manera posible ese concepto, igualmente es altamente probable que parte de lo que hayamos medido sea sólo error.

Por ejemplo, posiblemente un cierto porcentaje de las personas que respondieron lo hicieron de mala gana o afectadas por algún problema personal, otro grupo posiblemente no entendió demasiado bien la pregunta, algunos no se sintieron representados por ninguna de las alternativas propuestas, un pequeño porcentaje no quiso responder su verdadera opinión, etc. En definitiva, medir conceptos de ciencias sociales siempre incorpora un grado variable, pero frecuentemente importante de error. Equivale a tratar de comprender una conversación escuchada en una radio con mucha estática y ruido de fondo.

Sin embargo, a pesar de todo ello, si construimos nuestros instrumentos con cuidado posiblemente obtendremos asociaciones y resultados coherentes e iluminadores del alguna área de la realidad social. Claro, dado el error con que medimos, quizás nunca nuestras correlaciones entre variables serán todo lo fuertes que nos gustaría que fueran, pero al menos la mayor parte de las veces alcanzarán los niveles necesarios como para que podamos interpretar sus resultados.

Además, si los resultados obtenidos en un estudio se replican con los obtenidos en otros, eso nos servirá para confirmar más aún la validez de nuestra información y conclusiones.

## **Bibliografía**

Azofra, M. J. (1999). *Cuestionarios*. Cuadernos Metodológicos N° 26. Madrid: Centro de investigaciones sociológicas.

Cea, M.A. (1998). *Metodología cuantitativa: estrategias y técnicas de investigación social*. Madrid: Síntesis.

Goode W. y Hatt, P. (1967). *Métodos de Investigación Social*. Ciudad de México: Trillas.

König, R. (1973). *Tratado de sociología empírica*. Madrid: Tecnos.

Martinez Arias, R. (1996). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.

Muñiz, J. (1996). *Psicometría*. Madrid: Editorial Universitas.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.

Muñiz, J. (1998). *Teoría Clásica de los tests*. Madrid: Pirámide.

Muñoz, V. (2004). Imágenes y estudios cuantitativos en la construcción social de la juventud chilena. *Última Década*, 20, 71-94.

Navas Ara, M.J. (2001). *Métodos, diseños y técnicas de investigación psicológica*. Madrid: UNED.

Pastor, G. (1986). *Ideologías. Su medición psicosocial*. Barcelona: Herder.

PNUD. (2002). *Desarrollo humano en Chile: nosotros los chilenos, un desafío cultural*. Santiago de Chile: PNUD.

Sierra, Restituto (1979). *Técnicas de investigación social*. Madrid: Paraninfo.

Summers, G.F. (1982). *Medición de Actitudes*. Ciudad de México: Trillas.