

Desarrollo de escalas de medición para investigación en las ciencias administrativas

Celestino Robles Estrada

Profesor e investigador, Departamento de Mercadotecnia y Negocios Internacionales, Universidad de Guadalajara.

Resumen

El presente artículo analiza el papel de la medición en las ciencias administrativas, haciendo énfasis en mercadotecnia, y la importancia de la teoría en el desarrollo de escalas de medición. Ofrece sugerencias para generar y evaluar instrumentos de medición, y recomienda un procedimiento para el diseño y conducción de estudios para el desarrollo de escalas.

Abstract

The present article discusses the importance of measurement in the management sciences focusing in marketing research, and analyzes the importance of theory in scale development. Establishes suggestions for generating and judging measuring items and recommends a procedure for designing and conducting studies to develop scales.

INTRODUCCIÓN

La medición es el corazón de virtualmente todos los esfuerzos científicos, y las investigaciones de mercadotecnia no son la excepción. La medición efectiva es una piedra angular de la investigación científica de negocios. Sin embargo, y a pesar de ello, muchos investigadores de las ciencias sociales y profesionales de la mercadotecnia y los negocios carecen de las herramientas necesarias para desarrollar una adecuada evaluación de los instrumentos usados para la medición de los constructos que desarrollan.

Existen diferentes estrategias para desarrollar y refinar medidas y la relevancia de estas estrategias depende en buena medida del fenómeno científico que está siendo medido. En las ciencias sociales y administrativas, por su propia naturaleza, a menudo estos constructos representan abstracciones que sólo pueden ser evaluadas de forma

indirecta o latente. Esta evaluación indirecta toma la forma de reportes de mediciones tabuladas con lápiz y papel, en los cuales una buena cantidad de indicadores se usan para medir el constructo, es decir, para "escalarlo".

Más aún, los constructos a que se hace referencia en el artículo son preceptuales; los encuestados se tasan a sí mismos o a otros en constructos que son subjetivos y basados en opiniones.

DESARROLLO. Perspectivas sobre medición en las ciencias administrativas y de negocios

¿Qué es medición?

Fundamentalmente, la medición consiste en reglas para asignar símbolos a objetos para representar numéricamente cantidades de atributos. La medición incluye evaluar los números de tal forma que refle-

jen los diferentes grados del atributo que está siendo evaluado (DeVellis, 1991; Nunnally y Berstein, 1994).

En las ciencias sociales, y en las económico-administrativas, la mayor parte del tiempo los "objetos" son personas, "reglas" se refiere a la explicitud con la que se asignan los números, y "atributos" son características particulares de los objetos bajo medición. Como tales, es importante notar que los objetos (v. gr., personas), no son medidos; son medidos sus atributos (v. gr., autoestima).

Algunas reglas son obvias y universales, tales como medir el peso en toneladas o kilogramos, pero las reglas para medir los constructos socio-psicológicos no son tan obvias. Por ejemplo, ¿cuáles son las reglas apropiadas para medir constructos tales como autoestima, satisfacción en el trabajo, auto-confianza del consumidor o excelencia académica? A pesar de que no existen reglas "universales" para medir tales constructos,

desarrollar reglas que eventualmente sean aceptadas es importante para la estandarización y el establecimiento de normas. Una medida se dice estandarizada cuando: a) las reglas de medición son claras, b) es de aplicación práctica (no es incómoda para el administrador del estudio o el encuestado), c) los resultados no dependen del administrador del estudio (Nunnally y Berstein, 1994), y d) tal medición genera resultados similares a través de las diferentes aplicaciones (es decir, la medida es confiable) y ofrece conteos de fácil interpretación.

En muchas ocasiones, en los estudios administrativos y en los mercadológicos en especial, como ya se señaló anteriormente, normalmente no se mide a las personas per se, se miden sus atributos. La distinción es importante porque enfatiza la naturaleza abstracta de la medición. Es decir, se debe "abstraer" al atributo de la persona. Muchos estudios en mercadotecnia y negocios intentan determinar la relación entre dos atributos (v. gr., autoestima y propensión al consumo). Para evitar confusiones entre atributos relacionados, debe determinarse y especificarse la naturaleza exacta del atributo. Más aún, debe evaluarse si el atributo es realmente factible de medirse.

Utilidad de las escalas de medición en los negocios y la mercadotecnia

¿Qué propiedades constituyen la utilidad de una medida? Como ya se mencionó anteriormente, existen múltiples criterios que son usados para evaluar las medidas. Los criterios con mayor relevancia dependen de las metas de la evaluación o del esfuerzo científico que se lleve a cabo. Aunque existen dife-

rentes opiniones, un punto de vista preponderante entre la comunidad científica señala que los resultados deben ser *repetibles* y que la medida en sí misma debe estar *estandarizada*. *Repetibilidad y estandarización* son conceptos relacionados. Bajo circunstancias similares, un hallazgo de investigación basado en la misma medida debe repetirse. Este es el concepto básico de repetibilidad —que la medida es confiable bajo condiciones de prueba similares—. Los procedimientos de investigación mercadológica sólidos incluyen el establecimiento de normas. Cuando estas normas pueden ser interpretadas para describir a una persona como baja, media o alta en algún atributo, se dice que la medida está estandarizada.

La estandarización tiene varias ventajas:

Primero, a pesar de que lo que se mide son percepciones que por su naturaleza son subjetivas, una medición estandarizada mejora la objetividad de las ciencias administrativas. Cuando un investigador puede independientemente verificar la relación entre dos constructos que fue encontrada por otro investigador, se mejora la objetividad.

Segundo, la estandarización produce resultados numéricos cuantificables. Esta cuantificación permite la creación de categorías para análisis matemáticos y estadísticos. La cuantificación también mejora la comunicación y la generalización de los resultados. El conocimiento se acumula en las ciencias administrativas cuando los investigadores comparan sus resultados con los resultados de estudios previos. Cuando las mismas medidas estandarizadas, se usan a través de aplicaciones científicas, resultados nom-

brados como "bajo" o "alto" nivel de satisfacción tienen un significado común entre los investigadores.

Tercero, el desarrollo de escalas de medición es una tarea que consume tiempo. Si una medida ha sido desarrollada adecuadamente, ese tiempo ha sido bien usado, ya que la medida se vuelve disponible para su uso inmediato debido a que fue acordada sobre estándares. En el corazón de la repetibilidad y la estandarización se encuentran las propiedades de medida denominadas *confiabilidad y validez*.

Desarrollo de escalas para constructos latentes con múltiples ítems

Como ya se señaló, existen ocasiones en que los científicos y los profesionales de la administración necesitan medir atributos u objetos que por su propia naturaleza tienden a ser abstractos. Tales abstracciones son latentes por naturaleza. Los constructos latentes, por lo tanto, no son observables o cuantificables de forma directa. Un constructo latente es además, variable, es decir: la fuerza de la magnitud de las calificaciones para un constructo latente puede cambiar en el tiempo. Por ejemplo, la necesidad de logro constituye un constructo latente (v. gr., un atributo personal). No puede ser observado directamente por un investigador y por lo tanto requiere una escala para determinar su magnitud real en un momento dado en el tiempo. Más aún, el nivel de necesidad de logro de un individuo puede cambiar en el tiempo.

Este artículo se enfoca a constructos socio-psicológicos y mercadológicos de naturaleza teórica. Esto es, a aquellos constructos que bajo medición encajan en un marco

teórico. Otros constructos pueden ser de naturaleza empírica y sin sustento teórico (v. gr., encuestas de opinión). Con esto no se pretende sugerir que tales constructos empíricos o ateoréticos tengan poco valor, de hecho, algunos de éstos han probado ser útiles para el desarrollo de teorías.

Es una idea generalmente aceptada que la medición de constructos teóricos latentes requiere múltiples ítems o declaraciones para revelar con precisión los niveles de variación de los constructos; esto es, son escalables (DeVellis, 1991; Nunnally y Bernstein, 1994). Debido a que el nivel del atributo en cuestión en la persona (u objeto) es latente y abstracto, no puede ser medido de forma directa, y por lo tanto, debe construirse una escala. A pesar de que en algunas ocasiones es posible inferir un nivel de un constructo psicológico latente a través del comportamiento (v. gr., a través de repetición de compra de una marca específica, se puede inferir que el individuo considera que la compra de la marca determinada representa un buen valor por su dinero [constructo latente]); en algunas otras ocasiones, un comportamiento puede no ser indicativo de un constructo latente. En tales casos es necesaria una escala multi-ítem del constructo bien construida y validada.

Finalmente, es importante recalcar que el artículo se enfoca al desarrollo de escalas, no de índices. En las escalas, las calificaciones de los ítems son impulsadas teóricamente por el constructo latente. En los índices, las calificaciones en los ítems (indicadores) señalan la calificación total del índice; es decir, los ítems/indicadores "forman" la calificación del índice construido. A pesar de

que son latentes en muchos aspectos, los ítems/indicadores no son considerados escalas debido a que sus calificaciones no son reflejadas necesariamente por el constructo latente. Un ejemplo usado a menudo de ítems indicadores que resultan en un índice es el nivel socioeconómico (NSE). Los ítems o indicadores pueden incluir los ingresos, nivel de educación, ocupación y tipo de vivienda. A pesar de que algunos de estos indicadores tienen la propiedad latente de no ser observables directamente, se considera que sus calificaciones forman el índice de NSE y no a la inversa. Para una discusión más detallada de medidas "formativas" vs. "reflectivas" se puede consultar a Diamantopoulos y Winklhofer (2001).

Es importante, además, señalar que el proceso de desarrollo de escalas en estudios de negocios y de mercado comienza con una exhaustiva revisión bibliográfica de la literatura, en la que se delinee y describa una sólida y clara definición del constructo y su dominio. Esta definición debe señalar qué se incluye en el dominio del constructo, y qué está excluido de su dominio, y la dimensionalidad a priori del dominio del constructo. En esencia, la definición del constructo y el dominio del contenido determinan la dimensionalidad teórica.

Tendencias recientes en el desarrollo de escalas para constructos psicológicos y administrativos latentes

Textos recientes han compilado escalas usadas a menudo en mercadotecnia y comportamiento del consumidor (Bearden y Netemeyer, 1998; Bruner y Hensel, 1997). Este reno-

vado interés obedece a dos factores.

Primero, conforme las teorías en las ciencias sociales y administrativas se desarrollan y evolucionan, se desarrolla a su vez la necesidad de validarlas objetivamente. Estas teorías requieren la operacionalización de los constructos de interés. Cuando los constructos se miden bien (con confiabilidad y validez), se mejora la validación de la teoría. Más aún, a menudo ocurre que alguna escala usada con anterioridad necesita actualizarse o refinarse para reflejar más adecuadamente un constructo de interés.

Segundo, el avance de la computación y los programas de cómputo (*software*) ha mejorado la habilidad de los investigadores para desarrollar medidas. Los paquetes de cómputo estadísticos tales como el SPSS, SAS, LISREL, EQS, CALIS y AMOS, se han vuelto más amigables hacia el usuario, con lo que han hecho más fácil y rápido el llevar a cabo la mayoría de los análisis básicos y avanzados recomendados en el desarrollo de escalas de medición. Es importante sin embargo, advertir algo: los avances recientes en computación y *software* estadístico pueden permitir el desarrollo de escalas de forma más rápida, pero no necesariamente el desarrollo de mejores escalas. Para desarrollar escalas de medición sólidas deben seguirse los procedimientos adecuados (DeVellis, 1991; Nunnally y Berstein, 1994).

No sólo existe una tendencia al desarrollo de más escalas, los procedimientos usados para validarlas también han evolucionado. Ayudados por los avances de los paquetes estadísticos señalados, se han vuelto aparentes pruebas más completas y elaboradas para validar la dimensionalidad de la escala y los efectos del

método (Bearden y Netemeyer, 1998). A pesar de que existe desde hace algún tiempo, la teoría de la generalización (Teoría G) está siendo aplicada más frecuentemente al desarrollo y validación de escalas de medición (Marcoulides, 1998).

Constructos latentes

Como ya se comentó, los constructos latentes no son observables o cuantificables directamente, y las calificaciones de las mediciones de los constructos latentes pueden cambiar en el tiempo. Los constructos latentes requieren una elaboración cuidadosa en función de su nivel de abstracción y especificidad.

Teoría y validez

No se puede menospreciar la importancia de la teoría al desarrollar medidas de constructos latentes. En su trabajo clásico sobre medición y validez, Cronbach y Meehl (1955) plantean elocuentemente la importancia de la teoría en la medición. Para que las medidas de un constructo latente sean relevantes a las ciencias sociales y administrativas, deben tener sólidas raíces en un marco teórico. En otras palabras, ¿qué predice el constructo latente en cuestión? Estas relaciones han sido llamadas “red nomológica del constructo latente” (Cronbach y Meehl, 1955).

La teoría no tiene que ver sólo con el constructo latente de interés, sino también con la validez de la medición del constructo. La relevancia del constructo latente depende de su “validez de constructo”. En pocas palabras, la validez del constructo es una evaluación del grado en el cual una medida mide realmente al constructo latente que intenta medir.

Cronbach y Meehl (1955) plantearon que el demostrar la validez del constructo involucra al menos tres pasos: a) especificar un conjunto de constructos teóricos y sus relaciones (una teoría), b) desarrollar métodos para medir los constructos de la teoría, y c) probar empíricamente qué tan bien miden los indicadores (ítems) manifiestos (observables), los constructos en la teoría, y qué tan bien contrastan las relaciones hipotetizadas entre los constructos teóricos. Más aún, evaluar la validez de un constructo es un proceso continuo. No es suficiente un estudio que apoye la validez de un constructo para concluir que la medición ha sido validada. Se requieren múltiples pruebas y aplicaciones en el tiempo, y muchas veces esto puede requerir la refinación del constructo en sí mismo, así como su medición.

Importancia de la revisión bibliográfica

Una teoría sólida comienza con conceptualizaciones basadas en una revisión exhaustiva de la literatura. Esta revisión tiene dos propósitos. Primero, la revisión bibliográfica debe alertar al investigador de intentos previos de conceptualizar el constructo de interés y de teorías en las cuales el constructo puede ser útil como variable dependiente o independiente. La revisión bibliográfica también indicará intentos previos de medición del constructo y las fortalezas y debilidades de dichos intentos.

Segundo, dado que el proceso de desarrollo de escalas y su validación es una tarea que consume mucho tiempo y a veces es costosa, la revisión bibliográfica debe ayudar a responder la siguiente pregunta: ¿se necesita realmente la escala? Si

ya existen escalas de medición adecuadas del constructo, sólo es valioso desarrollar una nueva si muestra una ventaja teórica o empírica sobre las escalas de medición existentes para el constructo.

Dimensionalidad, confiabilidad y validez

La dimensionalidad, la confiabilidad y la validez son, todas ellas, propiedades de las medidas que están interrelacionadas entre sí. Estas propiedades permiten determinar en su conjunto la calidad de las escalas de medición.

Dimensionalidad

Establecer la dimensionalidad de los constructos es una parte importante del proceso de desarrollo de escalas. Es casi imposible desarrollar buenas medidas de un constructo sin conocer la dimensionalidad del constructo. Por ejemplo, considérese la tarea de desarrollar una escala para medir la inteligencia. ¿Tiene la inteligencia una faceta o dimensión única? (v. gr., es unidimensional), ¿o tiene múltiples facetas y dimensiones (v. gr., es multidimensional)? Si la inteligencia es un constructo unidimensional, entonces se requiere un simple número para medir la inteligencia de una persona. Si, sin embargo, se acepta la teoría de Howard Gardner (Gardner, 1993) que señala que la inteligencia se compone de ocho dimensiones—lingüística, lógica, musical, espacial, cenestésica, intrapersonal, interpersonal y naturalística— se necesitan entonces mediciones de las ocho dimensiones para medir la inteligencia de una persona.

La dimensionalidad es, pues,

una medida que tiene que ver con la homogeneidad de los ítems. Básicamente, una medida considerada unidimensional tiene propiedades estadísticas que demuestran que sus ítems son la base de un constructo simple o factor. Cuando una medida es multidimensional, los ítems contienen más de una dimensión o factor. El dominio de un constructo puede hipotetizarse como unidimensional, multidimensional, y/o como un factor de orden superior. Por lo tanto, la escala (o subescalas/factores) usados para operacionalizar el constructo deben reflejar la dimensionalidad hipotetizada. Dado que la unidimensionalidad de la escala (factor) se considera un prerrequisito de confiabilidad y validez, la evaluación de la unidimensionalidad es primordial (Gerbing y Anderson, 1988).

Se han utilizado una buena cantidad de procedimientos para revisar la dimensionalidad de una escala (v. gr., análisis de ítems, o análisis exploratorio de factores). Una técnica comúnmente aceptada es el análisis confirmatorio de factores, en el cual varios ítems multifactores (y las relaciones entre los factores) pueden ser especificados y evaluados con criterios usados para evaluar la dimensionalidad (v. gr., índices de ajuste, o presencia de errores de medición correlacionados) (Gerbing y Anderson, 1988; Kumar y Dillon, 1987).

Confiabilidad

La confiabilidad tiene que ver con la porción de la medición que se debe a efectos permanentes que persisten de muestra en muestra. Existen dos amplios tipos de confiabilidad a los que se hace referencia en la literatura econométrica y mercadométrica: a) Prueba-Reprueba (estabilidad

temporal) —la correlación entre las respuestas de una misma persona a un mismo juego de factores en dos puntos del tiempo—; y b) consistencia interna —la confiabilidad interna entre ítems o juegos de ítems en la escala.

La *confiabilidad prueba-reprueba* o “estabilidad” se determina a través de un coeficiente de estabilidad estimado mediante la magnitud de la correlación entre las mismas medidas (y mismas muestras) en diferentes ocasiones de evaluación. Si el coeficiente de estabilidad es bajo en magnitud —sin cambio en el constructo a través del tiempo—, la confiabilidad de la medición es dudosa. Debido a esto, la *confiabilidad prueba-reprueba* es útil ya que ofrece información del grado de confianza que se tiene de que la medición refleje el constructo y de que sea generalizable a otras ocasiones de evaluación.

La consistencia interna evalúa la interrelacionabilidad de los ítems. Los ítems que componen una escala (o sub-escala) deben mostrar altos niveles de consistencia interna. Algunos criterios usados comúnmente para evaluar la consistencia interna son la correlación promedio íter-ítems y el número de coeficientes de confiabilidad (Churchill, 1979). El coeficiente más ampliamente usado para determinar la confiabilidad interna es el coeficiente alfa de Cronbach (1951), aunque también se usan algunas “reglas de dedo” para determinar lo que constituye un nivel aceptable para el coeficiente alfa. Conforme el número de ítems aumenta, alfa tiende a incrementarse también, y debido a que la parsimonia también es una preocupación importante en la medición, una pregunta importante es ¿cuántos ítems se necesitan para medir un constructo? La respuesta a

esta pregunta depende parcialmente del dominio y las dimensiones del constructo. Naturalmente, un constructo con un amplio dominio y múltiples dimensiones requerirá más ítems para definir la relación dominio/dimensiones, que un constructo de estrecho dominio con pocas dimensiones. Dado que la mayoría de las escalas suelen ser auto-administradas y que debe tomarse en cuenta la fatiga del encuestado y/o su falta de cooperación, normalmente es ventajosa una escala breve (Churchill y Peter, 1994).

Con el advenimiento del modelado estructural de ecuaciones, se han vuelto disponibles otras pruebas de consistencia interna o de estructura interna/estabilidad. La *confiabilidad compuesta* (confiabilidad del constructo), la cual es similar al coeficiente alfa, se puede calcular directamente (Fornell y Larcker, 1981). Una prueba aún más rigurosa de la estructura interna/estabilidad involucra la evaluación de la cantidad de varianza capturada por la medición del constructo en relación a la cantidad de varianza debida al error de medición —la varianza promedio extraída (VPE, o AVE, por sus siglas en inglés)—. Al usar una combinación de los criterios señalados líneas arriba (v. gr., examinando la relación íter-ítems interna, coeficiente alfa, estabilidad compuesta y VPE), se pueden diseñar las escalas de una manera eficiente, sin sacrificar la consistencia interna.

Validez del constructo

La validez del constructo se refiere a qué tan bien una medida mide realmente el constructo que se intenta medir. La validez del constructo es la meta más importante al desarro-

llar un instrumento de medición y abarca toda la evidencia sostenida por una medición. No existe un acuerdo universal sobre la clasificación y los tipos de validez que tienen que ver con la validez de constructos. Sin embargo, muchos investigadores sostienen que la validación de un constructo se determina con la validez de traslación (forma y contenido), convergente, discriminante, predictiva, nomológica y de grupo conocido; estas formas de validación, en su conjunto, representan las formas más utilizadas de validez de constructo. Dado que toda la evidencia soportada por una medición contribuye a establecer la validez de un constructo, una medida debe, a priori, exhibir su dimensionalidad teórica y mostrar evidencia de confiabilidad para ser considerada válida. Como tal, la dimensionalidad y la variabilidad son condiciones necesarias pero insuficientes para validar el constructo. Se cuenta con un cierto número de procedimientos ya existentes para establecer la validez del constructo.

La *validez de traslación* tiene que ver con el contenido de los ítems; se tienen dos tipos de validez de traslación: validez de contenido y de forma. El término validez de contenido ha sido definido en muchas formas, pero la mayoría de las mismas hacen hincapié en que las medidas de los ítems son una muestra adecuada del dominio teórico del constructo (Messick, 1993). La mayoría de las definiciones concuerda en que la validez de contenido refleja el grado en el cual los elementos de un instrumento de medición son relevantes y representativos del constructo propuesto, en el grado en que el dominio total del constructo propuesto ha sido muestreado para un pro-

pósito particular de evaluación. Por "elementos" se refieren al contenido de los ítems individuales, formatos de respuesta e instrucciones a los encuestados, y por "representatividad" se entiende el grado en el cual los elementos son proporcionales a los dominios del constructo propuesto y al grado en que ha sido muestreado el dominio total del constructo propuesto. Esto es, que los ítems deben parecer consistentes con el dominio teórico del constructo en todos los aspectos, incluyendo los formatos de respuesta e instrucciones. Al desarrollar escalas válidas en contenido, se recomienda generalmente que se genere un buen número de ítems para "cubrir el dominio del constructo", que los ítems sean tamizados por expertos en el tema, y que se corran pruebas piloto en poblaciones relevantes para ajustar y refinar el conjunto de ítems (Robinson et al., 1991).

A la *validez de forma* se le ha llamado "la mera apariencia de que la medición tiene validez" y se sostiene que debe distinguirse de la validez de contenido (Nevo, 1985). Otros van un poco más allá en la diferenciación de la validez de forma y de contenido en términos de investigador y encuestado. Un instrumento con alta validez de forma mejora su uso en situaciones prácticas al inducir la cooperación de los encuestados gracias a (entre otras cosas) su facilidad de uso, nivel de lectura adecuado, claridad, instrucciones sencillas y amenas, y formatos de respuesta de fácil uso. Una definición comúnmente aceptada de validez de forma exige que el instrumento o prueba, al usarse "en una situación práctica, debe, además de tener validez pragmática y estadística parecer práctico, pertinente y relacionado con los

propósitos del instrumento (o prueba); es decir, no sólo debe ser válido, sino que debe parecer válido a los encuestados" (Nevo, 1985: 287). Por lo tanto, se puede señalar que la validez de forma tiene que ver con lo que los encuestados de poblaciones relevantes infieren con respecto a lo que está siendo medido.

La *validez convergente* se refiere al grado al cual están relacionadas dos medidas diseñadas para medir el mismo constructo. Existe convergencia si las dos diferentes medidas del mismo constructo están altamente correlacionadas. La *validez discriminante* evalúa el grado en que están relacionadas dos medidas diseñadas para medir constructos similares, pero conceptualmente diferentes. Se considera evidencia de validez discriminante una correlación baja a moderada. A menudo han sido usadas matrices multi-método y multi-tratamiento (MTMM, por sus siglas en inglés) para evaluar la validez convergente y discriminante en aquellos casos en que se requieren métodos muy diferenciados (v. gr., auto-reporte vs. observación) (Churchill, 1979; Peter, 1979, 1981).

La *validez nomológica* ha sido definida como el grado en que se confirman las predicciones de una red teórica formal que contiene el concepto bajo escrutinio (Campbell, 1960). Evalúa el grado en el cual están ligados empíricamente, los constructos que están ligados teóricamente (v. gr., medidas que correlacionan significativamente en la dirección pronosticada). Se han desarrollado lineamientos para establecer la validez nomológica de un constructo, pero no han escapado de la crítica (Peter, 1981). Al igual que para evaluar la validez convergente y discriminante y la

consistencia interna, se han desarrollado recientemente modelos de ecuaciones estructurales usados para evaluar la validez nomológica de las escalas de medición. Existen varios libros (Bollen, 1989; Schumacker y Lomas, 1996) y artículos (Baggozi, Yi y Phillips, 1991) que ilustran técnicas de modelado, criterios de evaluación y lineamientos para determinar la validez nomológica.

La *validez de grupo conocido*, tiene que ver con la habilidad de la medida para distinguir de forma confiable entre grupos de personas que deben contrastar con medidores altos y bajos a algún tratamiento. Como ejemplos, una persona que es verdaderamente conservadora debe tener una calificación alta al medirla en una escala de conservadurismo, mientras que una persona que es liberal debe tener una calificación baja en la misma escala; y la fuerza de ventas en el negocio de venta de autos al menúdeo y los de ventas de supercomputadoras para usos empresariales deben diferir en sus niveles de orientación al consumidor (Saxe y Weitz, 1982). Por lo tanto, las diferencias medias de las calificaciones entre grupos para un constructo dado pueden ser usadas como evidencia de validez de grupo conocido.

Procedimiento recomendado para el desarrollo de escalas de medición en ciencias administrativas y mercadotecnia

De lo comentado con anterioridad, resulta evidente que existen numerosos artículos y algunos libros dedicados a mostrar como desarrollar una escala de medición (Churchill, 1979; DeVellis, 1991; Nunnally y Bernstein, 1994). Los pasos y los procedimientos varían de autor en

autor, basados en los propósitos de la medición. Aun así, la mayoría de los autores comparten lineamientos comunes al desarrollar escalas de medición. El procedimiento propuesto a continuación es útil para el desarrollo de escalas de medición tipo auto-reporte (v. gr., encuestas) de constructos socio-psicológicos o mercadológicos latentes:

1. Definición del constructo y dominio del contenido. Asuntos que se deben considerar:

a) La importancia de una definición clara del constructo, su dominio y el papel de la teoría.
b) El enfoque sobre el "efecto de los ítems/indicadores vs. ítems/indicadores "formativos".

c) Dimensionalidad del constructo: ¿unidimensional, multidimensional o de orden superior?

2. Generación y análisis de los ítems de medición. Asuntos que se deben considerar:

a) Suposiciones teóricas de los ítems (v. gr., muestreo del dominio).

b) Generación de ítems potenciales y determinación del formato de respuesta.

c) ¿Cuántos ítems en la propuesta inicial?

d) ¿Formatos de respuesta dicotómicos o de respuesta múltiple?

e) Redacción de los ítems.

f) Análisis de la validez de contenido en relación a la dimensionalidad teórica.

g) Análisis de los ítems (juicio de expertos) —enfocándose en la validez de contenido y de forma.

3. Diseño y conducción de estudios para desarrollar y refinar la escala. Asuntos que se deben considerar:

a) Prueba piloto como procedimiento de revisión y ajuste de ítems.

b) Uso de varias muestras de poblaciones relevantes para el desarrollo de la escala.

c) Diseño de estudios para probar las propiedades psicométricas.

d) Análisis inicial de los ítems por medio de análisis exploratorio de factores (AEF, o EFA, por sus siglas en inglés).

e) Análisis inicial de los ítems en cuanto a estimación de la consistencia interna.

f) Estimación inicial de validez.

g) Retención de los ítems relevantes para el estudio posterior.

4. Finalización de la escala de medición. Asuntos que se deben considerar:

a) La importancia de tomar varias muestras de poblaciones relevantes.

b) Diseño de los estudios para determinar los diferentes tipos de validez.

c) Análisis de los ítems por medio de EFA.

d) Importancia de la consistencia entre los EFA del paso 3 y los del paso 4.

e) Derivación de una estructura inicial de factores —dimensionalidad teórica.

f) Análisis de los ítems por medio de análisis confirmatorio de factores (CFA, por sus siglas en inglés).

g) Contraste de la estructura teórica de los factores y las especificaciones del modelo.

h) Evaluación de los modelos de medición de los CFA.

i) Comprobación de invarianza de los factores del modelo a través de diferentes estudios (v. gr., análisis de grupos múltiples).

j) Análisis adicionales a los ítems a través de estimados de consistencia interna.

k) Estimados adicionales de validez.

l) Establecimiento de normas a través de los estudios.

m) Aplicación de teorías de generalización.

CONCLUSIONES

De lo analizado con anterioridad, se puede establecer que para que un constructo latente sea valioso, debe poseer relevancia teórica o práctica. Por lo tanto, debe considerarse cuidadosamente qué es lo que el constructo de interés predice. Mas aún, dada la importancia de la medición en las ciencias administrativas, cualquier medición debe ser validada con parámetros y métodos estrictos para que permita construir inferencias confiables a partir de estudios empíricos.

BIBLIOGRAFÍA

- Baggozzi, R. P., Y. Yi y L. W. Phillips (1991), "Assessing construct validity in organizational research", en *Administrative Science Quarterly*, 36 (3), pp. 421-458.
- Bearden, W. O. y R. G. Netemeyer (1998), *Handbook of marketing scales: Multi-item measures of marketing and consumer behavior research*, Thousand Oaks, CA, Sage.
- Bollen, K. A. (1989), *Structural equations with latent variables*, Nueva York, John Wiley & Sons.
- Bruner, G. y P. Hensel (1997), *Marketing scales handbook: A compilation of multi-item measures*, Chicago, American Marketing Association, 2a edición.
- Campbell, D. T. (1960), "Recommendations for APA test standards regarding construct, trait, or discriminant validity", en *American Psychologist*, núm. 15, pp. 546-553.
- Churchill, G. A. (1979), "A paradigm for developing better measures of marketing constructs", en *Journal of Marketing Research*, núm. 16, febrero, pp. 64-73.
- y J. P. Peter (1984), "Research design effects on the reliability of rating scales: A meta-analysis", en *Journal of Marketing Research*, núm. 2, noviembre, pp. 360-375.
- Cronbach, L. J. (1951), "Coefficient alpha and the internal structure of tests", en *Psychometrika*, núm. 31, pp. 93-96.
- Cronbach, L. J. y P. E. Meehl (1955), "Construct validity in psychological tests", en *Psychological Bulletin*, núm. 52, pp. 281-302.
- DeVellis, R. F. (1991), *Scale Development: Theory and applications*, Newbury Park, CA, Sage.
- Diamantopoulos, A., y H. M. Winklhofer (2001), "Index construction with formative indicators: an alternative to scale development", en *Journal of Marketing Research*, núm. 36, pp. 269-277.
- Fornell, C., y D. F. Larcker (1981), "Evaluating structural equation models with unobservable variables and measurement error", en *Journal of Marketing Research*, núm. 18, febrero, pp. 39-50.
- Gardner, H. (1993), *Frames of mind: The theory of multiple intelligences*, Nueva York, Basic Books.
- Gerbing, D. W. y J. C. Anderson (1984), "On the meaning of within-factor correlated measurement errors", en *Journal of Consumer Research*, núm. 11, junio, pp. 572-580.
- y J. C. Anderson (1984), "An updated paradigm for scale development incorporating unidimensionality and its assessment", en *Journal of Marketing Research*, 25, mayo, pp. 186-192.
- Kumar, A. y W. R. Dillon (1987), "Some further remarks on measurement-structure interaction and the unidimensionality of constructs", en *Journal of Marketing Research*, núm. 24, noviembre, pp. 438-444.
- Marcoulides, G. A. (1998), "Applied generalizability theory models", en Marcoulides, G. A. (ed.), *Modern methods for business research*, Mahwah, NJ, Lawrence Erlbaum, pp. 1-28.
- Messick, S. (1993), "Validity", en Linn, R. L. (ed.), *Educational Measurement*, Orix Press, 2a ed., pp. 105-146.
- Nevo, B. (1985), "Face Validity Revisited", en *Journal of Educational Measurement*, núm. 22, pp. 287-293.
- Nunnally, J. y I. H. Bernstein (1994), *Psychometric theory*, Nueva York, McGraw-Hill, 3a edición.
- Peter, J. P. (1979), "Reliability: A review of psychometric basics and recent marketing practices", en *Journal of Marketing Research*, núm. 16, febrero, pp. 6-17.
- (1981), "Construct validity: A review of basic issues and marketing practices", en *Journal of Marketing Research*, núm. 18, mayo, pp. 133-145.
- Robinson, J. P., P. R. Shaver y L. S. Wrightsman (1991), "Criteria for scale selection and evaluation", en Robinson, J. P., P. R. Shaver y L. S. Wrightsman (eds.), *Measures of Personality and Social Psychological Attitudes*, San Diego Academic Press, pp. 1-15.
- Saxe, R., y B. A. Weitz (1982), "The soco scale: A measure of the customer orientation of the salespeople", en *Journal of Marketing Research*, núm. 19, agosto, pp. 343-351.
- Schumacker, R. E. y R. G. Lomax (1996), *A beginner's guide to structural equation modeling*, Mahwah, NJ, Lawrence Erlbaum Associates.