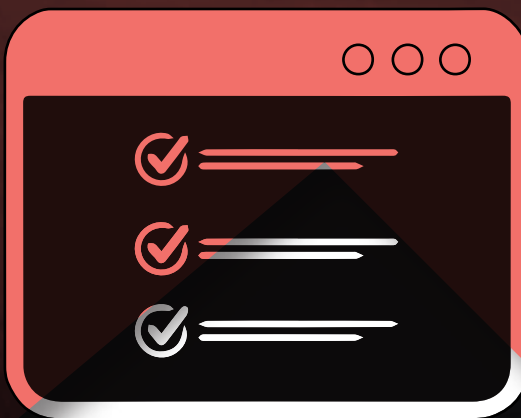
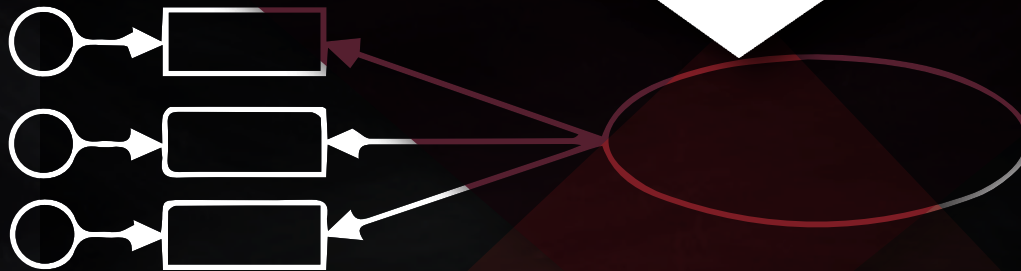




CREACIÓN DE ESCALAS EN LAS CIENCIAS DE LA ADMINISTRACIÓN

Dimensionalidad
Confiabilidad
Validez
Constructo
Diseño

Juan Mejía Trejo
Gonzalo Maldonado Guzmán



ISBN:



Creación de escalas en las ciencias de la administración

- Dimensionalidad**
- Confiabilidad**
- Validez**
- Constructo**
- Diseño**

Creación de escalas en las ciencias de la administración

- Dimensionalidad**
- Confiabilidad**
- Validez**
- Constructo**
- Diseño**

**JUAN MEJÍA TREJO
UNIVERSIDAD DE GUADALAJARA**

**GONZALO MALDONADO GUZMÁN
UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES**

Universidad de Guadalajara
2018

La presente obra fue sometida a un proceso de dictamen por pares de acuerdo a las normas establecidas por el Comité Editorial del Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara

D. R. © 2018, Universidad de Guadalajara
Centro Universitario de Ciencias
Económico Administrativas
Periferico Norte No. 779
Nucleo Universitario Los Belenes
Zapopan Jalisco, 45 100

ISBN: 978-607-547-184-6

<https://doi.org/10.55965/abib.9786075471846.2018a>

Impreso y hecho en México *Printed and made in Mexico*

Dedicatória

Índice

Introducción

Capítulo 1. La importancia de realizar mediciones en relación a las ciencias de la administración

¿Qué es medir?	17
¿Qué es una escala de medición?	18
La utilidad de realizar mediciones en las ciencias de la administración	20
Realizando la escala de constructos subyacentes con múltiples elementos ..	21
Constructo latente. Un ejemplo	23
Diagramas de trayectorias	23
La medición y su error	28
Tendencias en la dimensionamiento de los constructos para las ciencias de la administración	30
Marco teórico y validez de los constructos subyacentes	31
Importancia de la revisión de la literatura del estado del arte en el marco teórico	33
Propiedades de una medición: dimensionalidad, confiabilidad y validez	34
Dimensionalidad	34
Confiabilidad	35
Validez	37
El rol del investigador en las ciencias de la administración	39
El cientificismo	40
Determinación del problema a investigar	42
La medición en las ciencias de la administración	42
Concepto y variables: definición y tipos	46
Etapas para el desarrollo de una escala	46
Conclusión	47

Capítulo 2. Dimensionalidad

Unidimensionalidad de constructo, de indicadores y de conjunto de indicadores	49
Multidimensionalidad de constructo, de indicadores y de conjunto de indicadores	51
Implicaciones de la unidimensionalidad	54
Importancia de la unidimensionalidad	57
Evaluando la dimensionalidad de los constructos	58
Análisis Factorial Exploratorio (EFA. <i>Exploratory factor analysis</i>)	59
Análisis Factorial Confirmatorio (CFA. <i>Confirmatory factor analysis</i>)	67
Conclusión	69

Capítulo 3. Confiabilidad

La importancia de medir la confiabilidad	71
Confiabilidad test-retest	74
Confiabilidad alternativa	76
Confiabilidad por consistencia interna	76
Confiabilidad por consistencia interna: división en mitades.....	77
Confiabilidad por consistencia interna: coeficiente de alfa (?)	79
Coeficiente alfa. Explicación 1.	81
Ejemplo de caso basado en explicación 1	84
Coeficiente alfa. Explicación 2.	85
Ejemplo de caso basado en explicación 2	88
Coeficiente de alfa y dimensionalidad	94
Coeficiente de alfa, longitud de escala, correlación y redundancia entre indicadores	96
Índice de fiabilidad compuesta	98
Índice de la varianza extraída	101
Conceptos adicionales de la confiabilidad	102
Conclusión	103

Capítulo 4. Validez

Validez de versión (<i>translation validity</i>)	106
Validez aparente (<i>face validity</i>)	106

Validez de contenido (content validity)	107
Estudio de caso: administración del conocimiento OECD (2003)	109
Interpretación	111
Validez relacionada con el criterio (criterion-related validity)	111
Validez predictiva y post-dictiva (predictive and post-dictive validity) ..	111
Validez concurrente (concurrent validity)	112
Validez convergente (convergent validity)	112
Estudio de caso: administración del conocimiento OECD (2003)	113
Interpretación	114
Validez discriminante (discriminant validity)	117
Estudio de caso: administración del conocimiento OECD (2003)	120
Test de la diferencia entre las Chi cuadrado	121
Test del intervalo de confianza	122
Test de la varianza extraída	122
Validez del grupo conocido (known-group validity)	123
Validez nomológica (nomological validity)	124
Estudio de caso: administración del conocimiento OECD (2003)	126
Interpretación	128
Respuesta socialmente deseable (socially desirable responde bias)	129
Resumen	131

Capítulo 5. Definiendo el constructo y su contenido (etapas 1 y 2).

Etapla 1: El constructo

La importancia de la definición de un constructo claro y preciso	136
La importancia de una teoría sólida, revisión de la literatura y su análisis	136
El enfoque de indicadores reflectivos (de efecto) vs. indicadores formativos (de causa)	138
El papel de la teoría al especificar la dimensionalidad	140
Etapla 2: Generando y analizando los indicadores de contenido	141
La muestra y su contenido	141
Generando el pool de indicadores	142
Fuentes de indicadores	144
Escritura de los indicadores	145
Claridad de la redacción	145

La redacción redundante	145
La redacción en sentido positivo o negativo	147
a elección de formato de respuesta	147
Número de indicadores del pool inicial	150
Evaluando indicadores: validez de contenido (content validity) y validez aparente (face validity)	150
Guía para validez de ambas	151
Guía para el desarrollo de escalas	152
Determinar qué es lo que se pretende medir	152
Generar un listado de ítems	152
Determinar el formato de medida	153
Escala de Thurstone	153
Escala de Guttman	154
Escalas con ítems del mismo peso (Likert y diferencial semántico)	154
El listado de ítems debe ser revisado por expertos	157
Considerar la inclusión de ítems de validación	157
Administrar la escala a una muestra de prueba	158
Evaluando los ítems	158
Optimizar la longitud de la escala	158
Conclusión	159

Capítulo 6. diseñando la escala. análisis factorial exploratorio (etapa 3)

Prueba piloto	161
Análisis Factorial Exploratorio (EFA. <i>Exploratory factor analysis</i>)	163
Número de factores a extraer	165
Métodos de rotación	167
Retención de indicadores	167
Prueba inicial de indicadores y de confiabilidad	168
Consideraciones finales	169
Caso de estudio	
EFA: Condiciones iniciales	171
EFA: Resultados	174
EFA: Discusión y conclusión	174
Conclusión	176

Capítulo 7. Finalizando la escala. Análisis factorial confirmatorio y ecuación estructural (etapa 4)

CFA. Variables latentes vs. observadas	180
FA. Variables latentes exógenas vs endógena	181
CFA. El modelo de análisis factorial	182
CFA. Modelo completo de variable latente	183
CFA y SEM. Modelaje estadístico de propósito general y proceso	183
CFA. Notación simbólica	184
CFA. Diagrama de trayectoria	185
CFA y SEM. La ecuación estructural	187
CFA. Componentes no visibles de un modelo	192
Modelo SEM. Composición básica	192
EQS. Notación	214
EQS. Diagrama de trayectoria	214
EQS. Representación del sistema Bentler-Weeks	215
EQS. Uso del programa	217
Componentes del archivo de entrada	217
EQS. Reglas básicas en la creación de archivos de entrada	218
Keywords (palabras clave)	218
Descriptive Statements	218
File Editors	219
Basic Components of the EQS Input File	219
/TITLE (optional)	220
/SPECIFICATIONS (SPEC; Obligatorio)	220
Data (DA)	220
Variables (VAR)	221
Cases (CAS)	221
Method (ME)	221
Analysis (ANAL)	223
Matrix (MA)	223
/LABELS (opcional)	224
/EQUATIONS (EQU; obligatorio)	225
EQS. Cómo escribir ecuaciones	225
EQS. /VARIANCES (VAR; obligatorio)	227

EQS. /COVARIANCES (COV; Opcional)	228
EQS. Casos de ejemplos de archivos de entrada	228
CFA. Modelo de primer orden	229
El concepto de identificación de modelo.	232
El problema de la escala adecuada	233
CFA. Modelo de segundo orden	234
SEM. Modelo completo	235
CFA. Modelo de primer orden. Análisis de caso de estudio 1	237
CFA. Problema	239
CFA formulación de matriz	240
CFA. Condiciones de aplicabilidad	240
CFA. Identificación	244
CFA. Estimación	244
CFA y EQS. Entrada de datos al programa	247
CFA y EQS. Cómo ejecutar el programa	251
CFA y EQS. Análisis de resultados	252
CFA. Discusión	264
CFA y EQS. Reespecificación del modelo	265
CFA. Confiabilidad	266
CFA. Cálculo IFC	269
CFA y EQS. Redefiniendo el modelo	271
EQS. Recálculo de IFC y cálculo IVE	
CFA y EQS. Determinando la validez	276
FA y EQS. Cálculo de la validez discriminant.	
Test del intervalo de confianza 7.21.15.2.	278
CFA y EQS. Reportando resultados:	
consistencia interna y validez convergente	279
CFA Reporte de resultados: validez discriminante del modelo teórico	284
CFA. Conclusiones	285
SEM y EQS. Calculando la ecuación estructural (SEM)	286
SEM y EQS. Análisis de resultados	289
SEM y EQS. Discusión	293
SEM y EQS. Escribiendo la comprobación de hipótesis	294
Conclusión	295

CFA. Modelo de segundo orden. Análisi de caso de estudio 2	295
CFA, SEM y EQS. Entrada de datos al programa	296
CFA, SEM y EQS. Análisis de resultados	303
CFA, SEM y EQS. Eliminando V1	312
CFA, SEM y EQS. Discusión rápida de resultados, sin V1	313
CFA , SEM y EQS. Eliminando V3	318
CFA, SEM y EQS. Discusión rápida de resultados sin V1 y V3	318
CFA, SEM y EQS. Cálculo del alfa de Cronbach sin V1 y V3	321
CFA, SEM y EQS. Cálculo de IFC sin V1 y V3	324
CFA, SEM y EQS. Cálculo de IVE sin V1 y V3	325
CFA, SEM y EQS. Análisis de la validez de la escala, sin V1 y V3	326
Validez de contenido	327
Validez de constructo o de concepto	328
Validez discriminante	329
Validez nomológica	329
CFA, SEM y EQS. Reportando resultados: consistencia interna y validez convergente	331
Conclusión	323
CFA. Modelo de segundo orden. Análisis de caso de estudio 3	335
CFA, SEM y EQS. Entrada de datos al programa	337
CFA, SEM y EQS. Discusión rápida de resultados	340
CFA, SEM y EQS. Discusión rápida de resultados sin V14, V15 y V16	344
CFA, SEM y EQS. Cálculo del alfa de Cronbach sin V14,V15 y V16	344
CFA, SEM y EQS. Cálculo de IFC, sin V14,V15 y V16	347
CFA, SEM y EQS. Cálculo de IVE, sin V14,V15 y V16	352
CFA, SEM y EQS. Análisis de la validez de la escala, sin V14,V15 y V16	352
Validez de contenido	352
alidez de constructo o de concepto	355
Cálculo de validez discriminante	355
Cálculo de validez nomológica	359
CFA, SEM y EQS. Reportando resultados: consistencia interna y validez convergente	362
Conclusión	362
Bibliografía	366

Introducción

Obtener reconocimiento y credibilidad en resultados de investigación científica depende de qué estrategia esclarece y justifica mejor, la técnica de medición empleada. Así, es posible plantear diversas estrategias que permiten desarrollar y afinar las mediciones, sin embargo, su verdadero impacto dependerá del tipo de fenómeno científico que sea medido, tanto con variables observables directamente como de sus relaciones subyacentes. Es así que las ciencias de la administración deberán centrar su interés en el planteamiento, diseño, desarrollo y validación de constructos tanto de relaciones directas como subyacentes. Particularmente, este último tipo de relaciones implica constructos que al representar abstracciones sólo pueden evaluarse de manera indirecta. La evaluación indirecta implica el diseño y uso de múltiples elementos (en nuestro caso, indicadores) que miden al constructo, es decir, "plantean la escala para medir". Así, partiendo de un problema de investigación determinado y caracterizados los factores, variables e indicadores que mejor la describan en forma de constructos, el objetivo principal de este documento es el de plantear y diseñar las relaciones de dichos factores, variables e indicadores para discutir cómo medirlos y validarlos a través de la creación de escalas.

Para lograr este objetivo, en el capítulo 1, se justifica y enfatiza la importancia de la teoría del desarrollo de escalas a partir de explicar conceptos clave y básicos, como son: dimensionalidad, confiabilidad y validez. Además se destaca la importancia del enfoque de cuatro pasos en el desarrollo de una escala que involucra a estos tres conceptos.

En los capítulos 2. Dimensionalidad, 3. Confiabilidad y 4. Validez se plantea y se describe con mayor detalle cada uno los conceptos clave. Los capítulos 5, 6 y 7 hacen énfasis y detallan el enfoque en cuatro etapas y ofrecen ejemplos empíricos relevantes para cada paso.

Particularmente se remarca la importancia del análisis factorial tanto exploratorio como factorial. El enfoque en cuatro etapas está basado en el precedente de

múltiples trabajos de creación de escalas de diversos autores (Churchill, 1979; Clark y Watson, 1995; DeVellis, 1991; Haynes, Nelson y Blaine, 1999; Nunnally y Bernstein, 1994; Spector, 1992).

Cabe destacar que el diseño de una escala con base en constructos que definen finalmente un modelo, son el cimiento de la investigación científica, los cuales, son de carácter más que nada iterativo, en vez de lineal. Esto quiere decir que, en lugar de seguir pasos y actividades lineales o consecutivas, como todo proceso lógico y secuencial, el proceso sugerido de desarrollo de la escala tiende fuertemente a ser un procedimiento iterativo y continuo en el que "se reinician todos los procedimientos de creación de escala". Esto es debido a que los investigadores, conscientes, aprenden de sus esfuerzos y errores, por lo que se hacen necesarias las revisiones, incluyendo las que sean de etapas tempranas como la definición de factores, variables e indicadores a nivel conceptual, así como la definición de la dimensionalidad de constructo.

CAPÍTULO 1

La importancia de realizar mediciones en relación con las ciencias de la administración

En este capítulo se aclaran el qué es medir, qué es una escala para qué realizarlo, su utilidad en las ciencias de la administración, el planteamiento de escalas con constructos subyacentes con múltiples indicadores, así como su importancia, las implicaciones de los conceptos básicos de dimensionalidad, confiabilidad y validez de escala para finalmente, revisar el proceso de las cuatro etapas que las involucran.

¿Qué es medir?

Medir es una de las principales tareas de la ciencia y de la que es posible afirmar que "si no se mide, no se controla". De acuerdo con el Diccionario de la lengua española, medir es: "comparar una cantidad con su respectiva unidad, con el fin de averiguar cuántas veces la segunda está contenida en la primera" (RAE, año), es decir, determinar una medida. Así, tenemos que la definición de medida, es "un conjunto de reglas que permiten asignar números a los objetos observados de tal forma que representen de manera adecuada la cantidad de un determinado atributo que poseen" (Nunnally y Bernstein, 1994). En algunos casos, esas reglas son muy obvias, como cuando se trata de medir la altura de un individuo con una cinta métrica. Pero, desgraciadamente, obviada no es común en las ciencias sociales. Así, las reglas para determinar el grado de timidez, de etnocentrismo o de inteligencia de un consumidor, o el grado de orientación al mercado de una empresa, no son tan intuitivas.

Para nuestro caso, medir consiste en aplicar reglas que asignan símbolos a objetos para representar numéricamente cantidades de atributos. Medir, incluye la evaluación de números tales que reflejen los diferentes niveles del atributo que se está evaluando (DeVellis, 1991; Haynes et al., 1999; Nunnally y Bernstein, 1994).

Se destaca que, en las ciencias de la administración, la mayoría de las veces los "objetos de estudio" son individuos, empresas, etc. y las reglas implican una

asignación explícita de números, y los atributos son rasgos que caracterizan de forma muy particular a cada uno de los objetos de estudio que se miden. Lo anterior quiere decir que los individuos o empresas no se miden, pero sí sus atributos, por ejemplo: el liderazgo o la innovación.

En cuanto a las reglas de medida requieren un poco más de explicación ya que algunas reglas son obvias y universales, como medir la distancia en kilómetros o millas. Sin embargo, las reglas que se emplean para medir los constructos socio-psicológicos no son tan evidentes. Por ejemplo, ¿cuáles son las reglas apropiadas para medir constructos como el liderazgo de los directores, la satisfacción de los clientes la autoestima de los empleados, la satisfacción en el trabajo y/o la autoconfianza del consumidor?

Se destaca que no existen reglas universales para realizar la medición de tales constructos, pero sí el desarrollo de reglas que son finalmente aceptadas como factor importante para la estandarización y el establecimiento de normas. Así, una medición se dice está estandarizada cuando (a) las reglas de medición son claras, (b) es práctico de aplicar, (c) no es exigente del administrador y/o supervisor, y (d) los resultados no dependen del administrador (Nunnally y Bernstein, 1994). Dicha medición tiende a producir resultados similares en todas las aplicaciones (es decir, la medición ya es confiable) y ofrece puntuaciones que pueden ser fácilmente interpretadas como bajas, medias y altas. Por otro lado, cabe aclarar que el enfoque de la medición de los atributos también requiere precisarse.

Como se ha explicado, realizar mediciones, medir a un individuo o empresa no es hacerlo per se, sino que se están midiendo sus atributos.

Esta distinción es importante porque enfatiza la naturaleza abstracta de la medición en las ciencias de la administración, es decir, se debe realizar un alto nivel de abstracción del atributo a medir. En las ciencias de la administración existen una gran variedad de estudios que intentan determinar relaciones entre dos atributos (por ejemplo, competitividad e innovación, o liderazgo con personalidad, etc.). Para evitar la confusión entre los atributos relacionados, la naturaleza exacta del atributo debe ser cuidadosamente determinada y especificada (marco teórico de la investigación, por ejemplo). Además, debe realizarse una evaluación de la factibilidad sobre si el atributo puede medirse en absoluto ya que algunos atributos son tan abstractos que pueden no ser susceptibles de medición, por ejemplo, la clarividencia (Nunnally y Bernstein, 1994).

¿Qué es una escala de medición?

Para medir, es necesario crear (o utilizar si ya existe) una escala de medida. Una escala de medida es un conjunto de *ítems*, frases o preguntas que permiten medir el

nivel que alcanza un atributo determinado (etnocentrismo, orientación al mercado) no directamente observable en un objeto (un consumidor, una empresa). De lo anterior, podemos afirmar que la escala es "un herramental diseñado por el investigador de las ciencias de la administración que representa directamente el modelo conformado por uno o varios constructos basados en factores, variables, dimensiones e indicadores tanto directos como subyacentes, que implementado, le permiten medir el fenómeno objeto de estudio, determinando su magnitud y sentido". ¿Qué ventajas aporta a la investigación la utilización de escalas estandarizadas de medida? Podemos resumirlas en las siguientes: (Aldás-Manzano y Maldonado-Guzmán, 2016)

Objetividad. Un principio científico básico es que cualquier afirmación efectuada por un investigador, ha de poder ser verificada independientemente por cualquier otro repitiendo el experimento en las mismas circunstancias. Aunque este principio es de limitada aplicación en ciencias sociales, no deja de ser cierto que si, ni siquiera se dispone de un instrumento para medir el atributo, difícilmente podrán compararse los resultados.

Cuantificación. Los resultados numéricos de medidas estandarizadas tienen la ventaja de permitir la utilización de técnicas estadísticas avanzadas.

Comunicación. Las medidas objetivas permiten la fácil comunicación de los resultados de las investigaciones entre los distintos científicos.

Economía. Desarrollar una escala adecuadamente es una tarea costosa, pero una vez creadas suponen un gran ahorro de tiempo. Así, por ejemplo, un investigador preparado puede juzgar bastante acertadamente el grado de innovación de un sector manteniendo con su CEO una entrevista en profundidad. Valorar esta misma variable en 100 empresas con 100 entrevistas CEO le resultará, sin embargo, muchísimo más costoso que administrarles un cuestionario con las preguntas que conformen la escala, por ejemplo, a diseño sugerido por el Manual de Oslo de la Organización para la Cooperación y el Desarrollo Económicos (OECD, 2005).

Desafortunadamente, aún y sus evidentes ventajas, no todas las escalas son desarrolladas con el suficiente cuidado en su diseño y aplicación. Se debe tomar en cuenta que una escala no es un conjunto de preguntas unidas sin más, sino que su diseño implica necesariamente el cubrir una serie de propiedades adecuadas conocidas como propiedades psicométricas, las cuales son básicamente tres: dimensionalidad, confiabilidad y validez. Por lo anterior, se hace imperioso al aspirante a investigación que conozca y aplique de manera sistemática un esquema de desarrollo de escalas que le permita generar sus instrumentos de forma adecuada para las medidas que pretenda analizar.

La utilidad de realizar mediciones en las ciencias de la administración

Dado que las ciencias de la administración tienen un fuerte componente psicológico individual, así como social, se hace notar que existen múltiples criterios basados en propiedades psicométricas y/o de conducta, que se utilizan para evaluar las mediciones. Los criterios más relevantes, son dependientes de los objetivos de la evaluación y del esfuerzo científico emprendido, de tal manera que las ciencias de la administración se centran en realizar propuestas que incluyen constructos que se describan con escalas de factores, variables y/o indicadores socio-psicológicas latentes.

Esto ha generado opiniones y debates, incluso muy divergentes, pero una postura que es común en la mayoría de científicos de las disciplinas de económico-administrativas es que los resultados de las mediciones deben ser repetibles y estandarizados ya que son conceptos relacionados.

Por ejemplo, un hallazgo de investigación basado en circunstancias similares y en las mismas mediciones debería replicarse. Este es el principio básico de la repetibilidad, que la medición se realice de forma fiable bajo condiciones de prueba similares. Los procedimientos, por ejemplo, de estímulos visuales y/o auditivos para la evaluación de atributos de un producto, a fin desarrollar de la escala incluyen el establecimiento de normas. Cuando estas normas, al ser interpretadas describiendo a una persona como baja, media o alta en un atributo, la medida se considera estandarizada. La estandarización tiene varias ventajas:

A. Aunque medimos percepciones que por su propia naturaleza son subjetivas, una medida estandarizada aumenta la objetividad de las ciencias de la administración. Cuando entre investigadores verifican de manera independiente una o varias relaciones entre sus constructos, se aumenta la objetividad, dado que las medidas utilizadas son las mismas y están estandarizadas. Si existe desacuerdo en cuanto a la adecuación de las medidas utilizadas para obtener el hallazgo, la objetividad se verá comprometida o en duda. En las ciencias de la administración, a menudo probamos teorías, que sólo son posibles hacerlas en la medida en que los atributos de la teoría (constructos) sean adecuadamente medidas. Cuando se acuerdan procedimientos para medir los atributos de interés, se aumenta la objetividad de las pruebas teóricas.

B. La estandarización produce resultados numéricos cuantificables que permite la creación de categorías (por ejemplo, baja, media, alta) para análisis matemáticos y estadísticos (ANOVA) o para uso como niveles de factor en diseños experimentales. La cuantificación también mejora la comunicación y la generalización de los resultados. El conocimiento se acumula en las ciencias de la administración

cuando los investigadores comparan sus resultados con los resultados de estudios previos. Cuando se utilizan las mismas medidas estandarizadas en aplicaciones científicas, los resultados son denominados como bajos en el liderazgo o altos en la innovación tienen un significado común entre los investigadores. Esto mejora la comunicación de los resultados y la generalización de los hallazgos.

C. Finalmente, el desarrollo de mediciones/escalas es una tarea que requiere mucho tiempo. Sin embargo, si una medición ha sido bien desarrollada, ésta recompensa en el futuro. Una vez que se produce la estandarización, la medida y su medición estarán disponibles para su uso con poco o ningún tiempo invertido debido a las normas acordadas. En el centro mismo de la repetibilidad y la estandarización están las propiedades de medición de fiabilidad y validez.

Realizando la escala de constructos subyacentes con múltiples elementos

En la literatura de hoy, se observa un número importante de conceptos básicos de la medición, así, es posible afirmar que los investigadores de las ciencias de la administración, enfocan sus esfuerzos en proponer constructos y modelos que expliquen, a través de las mediciones de los atributos de sus objetos de estudio que tienden a ser altamente abstractos y, por lo tanto, latentes o subyacentes de origen. *Esto significa que no son directamente determinados y, por lo tanto, observables y/o cuantificables de forma directa.* Los *constructos subyacentes* son variables en sentido y magnitud de las calificaciones y cambian con el tiempo. Por ejemplo, la necesidad de pertenencia representa un constructo latente o subyacente (es decir, un *atributo personal*). Esto significa que no puede ser observado directamente por un investigador y por lo tanto requiere del diseño de una escala que estime su magnitud y sentido real en un momento dado. Además, el nivel de necesidad de pertenencia de un individuo hacia un grupo puede cambiar con el tiempo. Pitonyak (2010) describe como el sentido de pertenencia está en función del grado de soledad, de la sensación de cercanía con otra persona, de la habilidad de relacionarse para hacer amistades, de los niveles de exclusión social que el individuo presenta, entre otros factores subyacentes, los cuales deben ser justificados dentro del marco teórico que se plantee. Por otro lado, algunos constructos son de naturaleza empírica o sin base teórica (*ateórica*). Las encuestas de opinión muchas veces evalúan constructos altamente empíricos. No se está sugiriendo que tales constructos y/o enfoques de medición tengan poco valor en las ciencias de la administración. De hecho, tienen un potencial relevante de utilidad, en el desarrollo del marco teórico. Por ejemplo, una encuesta de opinión (empírica y deseablemente teórica) que evalúa el nivel de

asimilación tecnológica en una firma y que favorece o no favorece el diseño y desarrollo de productos innovadores bélicos podría revelar que aquellas firmas que son innovadores tempranos, benefician el control de armas más que aquellos que son innovadores tardíos. Lo anterior es sólo para recalcar la importancia vital del marco teórico como sustento de los constructos de un modelo a fin de poderlo explicar. Se acepta generalmente que las medidas de los constructos teóricos latentes o subyacentes requieren múltiples elementos o afirmaciones para revelar más exactamente los niveles de las variables de dichos constructos; es decir, son escalados (Clark y Watson, 1995; DeVellis, 1991; Haynes, Richard y Kubany, 1995; Nunnally y Bernstein, 1994).

Es posible afirmar que, dado el nivel de un objeto de estudio sobre un atributo que es latente y con tendencia socio-psicológica a ser altamente abstracto no puede medirse directamente en el contexto económico-administrativo, debe construirse una escala. Algunas ocasiones, le será posible inferir un nivel de un constructo socio-psicológico latente a través del comportamiento; por ejemplo, en la compra repetida de la marca, uno infiere que un individuo cree que la marca representa un buen valor para el dinero (constructo latente). Muchas veces un comportamiento no puede ser indicativo de una construcción latente. Finalmente, se debe reiterar que la escala, y no la indexación, es nuestro enfoque. En la escala, las puntuaciones de los indicadores son teóricamente impulsadas por el constructo latente o subyacente; es decir, son un *efecto reflectivo (reflected)*. Con un índice, los puntajes en los indicadores impulsan la puntuación total del índice; es decir, los indicadores causan el índice construido.

Aunque todavía latentes en muchos aspectos, los indicadores de causa o formativos (*formativos*) no se consideran escalas porque sus puntuaciones no se reflejan necesariamente en la construcción latente. Un ejemplo muy frecuente usado en indicadores de causa o formativo (*formative*), que resultan en un índice, es el nivel de competitividad que se incluye en el premio nacional de calidad de México (PNC, 2017). Los indicadores incluyen nivel de liderazgo, generación de valor al consumidor, planeación estratégica, administración del cambio, compromiso social, salud, inclusión, administración del conocimiento y agilidad. Ciertamente, varios de estos indicadores tienen la propiedad latente de no ser directamente observables, sus puntuaciones se consideran el índice de competitividad y no viceversa. Para más información véase a Bollen y Lennox (1991), Diamantopoulus y Winklhofer (2001), MacCallum y Browne (1993) y Smith y McCarthy (1995), que se hacen referencia a las medidas formativa (*formative*) vs. reflectiva (*reflective*).

Constructo latente

El fenómeno subyacente que una escala pretende medir se denomina con frecuencia *variable latente*. Por ejemplo, el concepto de la relación innovación-administración del conocimiento del consumidor, escala desarrollada por Mejía-Trejo et al. (2016) que muestra tanto a la innovación y su relación con la administración del conocimiento del consumidor como *constructos latentes* dado que no son directamente observables. Además, de que también la administración del conocimiento del consumidor, se manifiesta a través de otros constructos o variables, como las citadas: CKM as a driver of innovation (CKMADI); (H) CKM support (CKMS); (I) CKM other sources of knowledge (CKMOSK); (J) CKM, satisfaction, experience and performance (CKMSEP), (tabla 1).

La administración del conocimiento del consumidor es también *variable* en cuanto que no es constante, es decir, no está fija en el tiempo. Una escala de medición como la administración del conocimiento del consumidor, buscará estimar la magnitud de esta variable latente en el momento y entorno geográfico actual, lo que denominamos, puntuación verdadera. La noción de constructo latente, implica que debe existir cierta relación entre ella y los ítems que componen la escala.

El constructo latente significa si la escala se ha desarrollado correctamente, la causa de la puntuación obtenida en los distintos ítems de la escala. La forma más común de representar las relaciones entre los distintos ítems de una escala y el concepto que miden, son los diagramas de trayectorias, facilitando la comprensión de técnicas estadísticas de validación de escalas, como el análisis factorial confirmatorio (CFA. *Confirmatory Factor Analysis*).

Diagramas de trayectorias

Los diagramas de trayectorias son un método para representar las relaciones causales entre las variables, y están sujetos a una serie de convenciones, tales como:

1. Una flecha (\rightarrow) entre dos variables, indica que una es la causa de otra, así $A \rightarrow B$ señalaría que A es la causa de B. Si la variable a la que se hace referencia es una variable latente, ésta se representa en un óvalo o círculos. Si es una variable observable o manifiesta, esto es directamente medible como son los indicadores de la escala, y se representa en un cuadrado (figura 1).
2. Otra convención de los diagramas de trayectorias, es la representación del término de error, que se realiza mediante una nueva variable causal (E ó D en la figura 1). E es el error de medida asociado a un constructo observable (causado)

**TABLA 1. Modelo innovación-administración del conocimiento del consumidor.
Constructo: administración del conocimiento del consumidor**

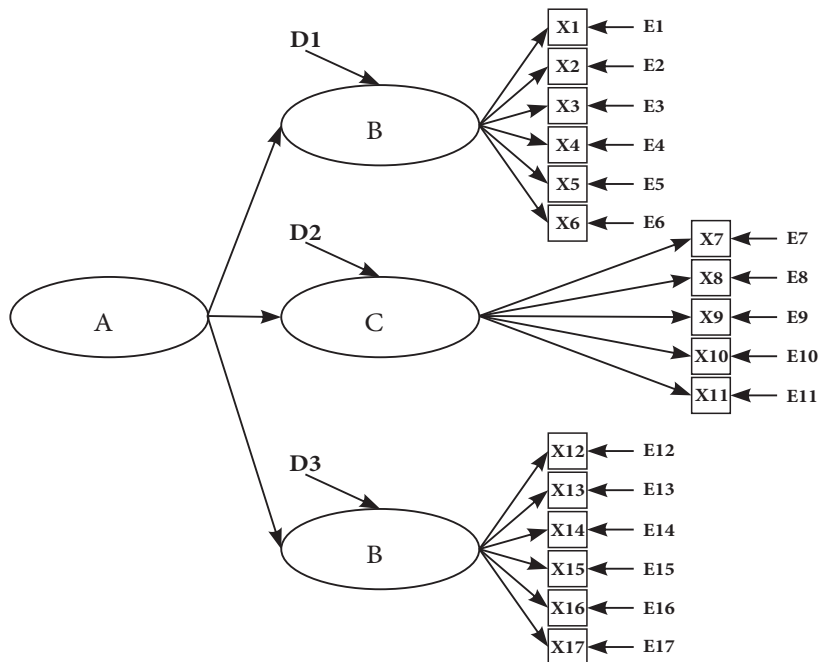
Customer Knowledge				
F	Variable	Indicator	Q	Author
(G)	34).- Information from Costumer (IFMC)	Customer is Resource of NPDIdeation; Customer Driven-Innovation (Innovation from Customers). Mutual Innovation.	78	Nambisan (2002), Desouza et al. (2007), Gibbert and Probst (2002), Chiu & Fogel (2014)
	35).- Information about the Customer (IABC)	Strategy of close collaboration with customers. Communities of creation.	79	Nambisan (2012), Gibbert and Probst (2002)
	36).- Information for Customer (IFRC)	Customer as a User collaborates intensively in the product testing and suport. Customer Focused Innovation (Innovation for Customers)	80	Nambisan (2002) Desauza et al. (2007)
	37).- Information as a Customer Co-creator (with)(IWIC)	Customer as a Co-creator helps over NPDI desing and development, Customer Centered Innovation (Innovation with Customers); Prosumerism; Team-Based-Colearning. Joint Intellectual Property	81	Nicalai et al. (2007), Desauza et al. (2007) Gibbert and Probst (2002)
	38).- Negative side effects of customer integration(NSEC)	The firm is warned about the dependence on customer´s personality (NSEC1)	82	Gassmann et al. (2012)
		The firm is warned about the dependence on customer´s experience (NSEC2)	83	
		The firm is warned about the dependence on customer´s point of view (NSEC3)	84	
		The firm is warned about to choose the wrong customer (NSEC4)	85	
		The firm is warned about the risk to integrate the customer to the company´s side (NSEC5)	86	
(H)	39).- Know ledge incentives (IWIC)	Salary associated with the ability and willingness to share knowledge (Ki1)	87	Nicolai et al. (2001). OECD (2003)
		salary dtetermined by willingness to impove skills and upgrade knowledge (K12)	88	

La importancia de realizar mediciones en relación con las ciencias de la administración

Customer Knowledge				
F	Variable	Indicator	Q	Author
		Tolerance of Failure (K13)	89	Gloet and samson
		Rewards and recognition(K14)	90	(2013)
	40).- Knowledge Fluence (KF)	Exchange the knowledge between employees across departments (KF1)	91	Nicolai et al. (2011), OECD (2003), Chiu and Fogel (2014)
		Communication among employees and management (Kf2)	92	
	41).- Knowledge and ICT (KICT)	ICT to support and control the Customer Knowledge Management	93	Laudon and Laudon (2012), Mejía-trejo and Sánchez-Gutiérrez (2013)
		Technical Services (ISOK1)	94	
		Engicering department (ISOK2)	95	
(I)	42).- Internal Sources of Knowledge (ISOK)	Research and design development (ISOK3)	96	García-Murillo and Annabi (2002)
		Production (ISOK4)	97	
		Marketing and sales (ISOK5)	98	
		Purchasing and Supply (ISOK6)	99	
		Other Employees (ISOK7)	100	
		Suplier (ESOKL1)	1	
	43).- External Sources of Knowledge (ESOK)	Scientist, Universities, Patents, Exhibitions Technological Consultant (ESOK2)	2	García-Murillo and Annabi (2002)
		Distributor Agents (ESOK3)	3	
		Competitor (ESOK4)	4	
		Of Only We Knew What We Know (KM) as a Customer retention (PAR1)	5	Gibbert and Probst (2002),
(J)	44).- Paradigm(PAR)	Retenntion is Cheaper than Acquisition (CRM) as a Custormer Satisfaction (PAR2)	6	García-Murillo and Annabi (2002)
		If We Only Knew What Our Customer (CKM) Knows as a Custtomer Experience and Creativity (PAR3)	7	
		Performance against budget; Customer retention rate. (KM) (PER1)	8	
	45).- Performance (PER)	Performace in terms of customer satisfaction and loyalty (PER2)	9	
		Performance againt competitirs in innovation and growth; Contribution to customer success (CKM) (PER3)	10	

Note: Factor (F); (A) Innovation Value Added (IVADD); (B) Innovation Income Items (IIIT); © Innovation Process (INPROC); (D) Innovation Outcome Items(IOITT); (E) Innovation Performance (IPERF); (F) Innovation Feedback Items IFEEED); (G) CKM as drive of Innovation (CKMADI); (H) CKM Support (CKMS); (I) CKM other Souces of Knowledge (CKMOSK); (J) CKM, Sastisfaction, Experience And Performance (CKMSEP).

FIGURA 1. Diagrama de trayectoria conceptual



Fuente: De Villis (1991).

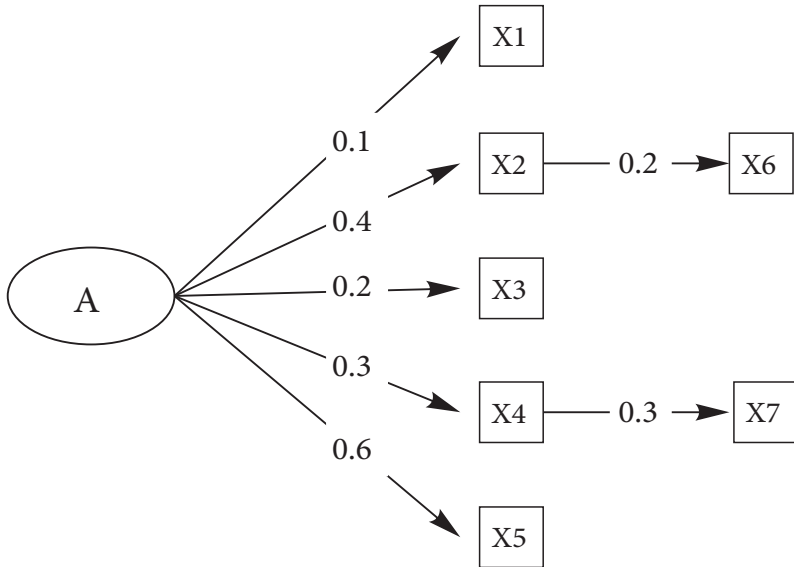
(causado)Z (causado, por ejemplo, por la incorrecta interpretación de las preguntas del cuestionario realizado por una empresa, o en su formulación por parte del entrevistador). El error D (*disturbance*) hace referencia al error en la predicción de una variable o constructo latente.

- Otro concepto fundamental en los diagramas de trayectorias, es el del coeficiente estandarizado de la relación causal (cargas factoriales en ecuaciones estructurales y análisis factoriales), que son los números que, una vez estimado el modelo, aparecen sobre las flechas (figura 2).

Estos coeficientes expresan la fuerza de la relación causal entre las variables unidas por la flecha. El hecho de que estos coeficientes se hayan estandarizado, indica que están medidos en la misma escala y son directamente comparables entre ellos. Así, en la figura 2, A es la causa mucho más intensa de X₅ que de X₁.

Existe una útil relación entre los coeficientes estandarizados y las correlaciones entre las variables manifiestas (indicadores de la escala): la correlación entre dos variables manifiestas causadas por una misma variable latente, es igual al producto

FIGURA 2. Diagrama hipotético de trayectorias por coeficiente estandarizado de la relación causal.



Fuente: Elaboración propia

(causado)Z (causado, por ejemplo, por la incorrecta interpretación de las preguntas del cuestionario realizado por una empresa, o en su formulación por parte del entrevistador). El error D (*disturbance*) hace referencia al error en la predicción de una variable o constructo latente.

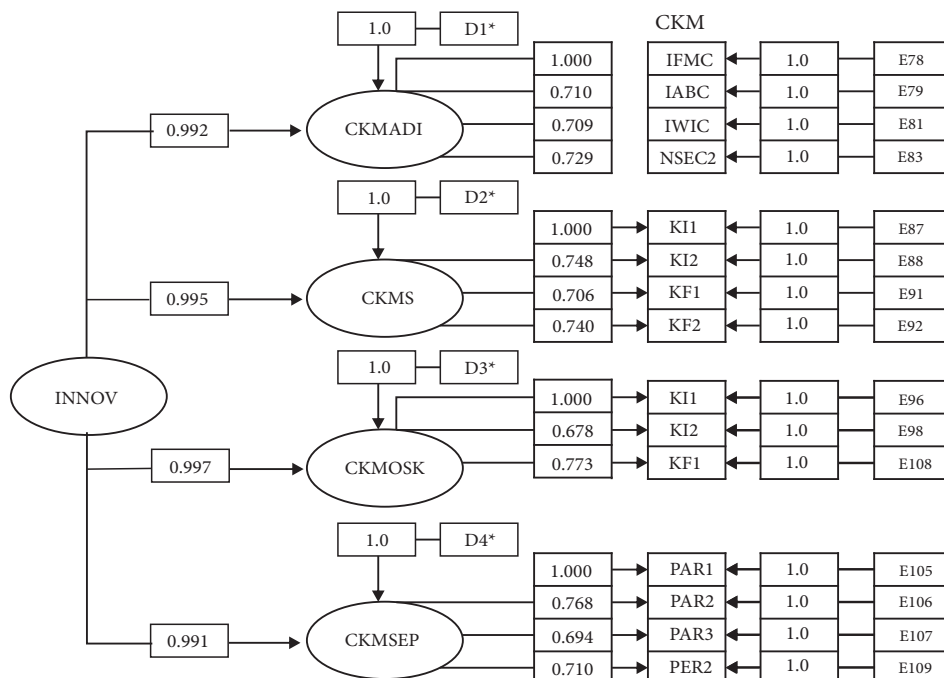
3. Otro concepto fundamental en los diagramas de trayectorias, es el del coeficiente estandarizado de la relación causal (cargas factoriales en ecuaciones estructurales y análisis factoriales), que son los números que, una vez estimado el modelo, aparecen sobre las flechas (figura 2).

Estos coeficientes expresan la fuerza de la relación causal entre las variables unidas por la flecha. El hecho de que estos coeficientes se hayan estandarizado, indica que están medidos en la misma escala y son directamente comparables entre ellos. Así, en la figura 2, A es la causa mucho más intensa de X₅ que de X₁.

Existe una útil relación entre los coeficientes estandarizados y las correlaciones entre las variables manifiestas (indicadores de la escala): la correlación entre dos variables manifiestas causadas por una misma variable latente, es igual al producto de sus respectivos coeficientes estandarizados. Por ejemplo, la correlación entre X₁ y

(causado)Z (causado, por ejemplo, por la incorrecta interpretación de las

FIGURA 3. Diagrama de trayectoria de la relación de innovación-administración del conocimiento del consumidor.



Hypothesized second-order factorial model of Customer Knowledge Management influenced by Innovation for SSG. Notes: E(n) -error disturbance; D(n)- variances of the disturbances. Because the estimation of all higher order factor loadings are typically of interest in second-order models, the variance of the single higher order factor (CKMS) has been constrained to 1.0; note also that the variances of the disturbances (the D's) are designated as freely estimated. Relatedly, their paths are automatically constrained to 1.0 by the program (Byrne, 2006).

Fuente: Mejía-Trejo *et al.* (2016).

La medición y su error

Es de gran importancia que se forme una idea más clara de cuáles pueden ser las fuentes de error en las mediciones de una variable o constructo latente por medio de una escala y, posteriormente, especificaremos los supuestos estadísticos que sean necesarios realizar. Así, se consideran ocho fuentes de variación o fuentes de

La medición y su error

Es de gran importancia que se forme una idea más clara de cuáles pueden ser las fuentes de error en las mediciones de una variable o constructo latente por medio de una escala y, posteriormente, especificaremos los supuestos estadísticos que sean necesarios realizar. Así, se consideran ocho fuentes de variación o fuentes de error que darán lugar a diferencias en los valores obtenidos en la medición de un concepto Churchill y Iacobucci (2002) en (Aldás-Manzano y Maldonado-Guzmán, 2015):

1. *Diferencias verdaderas en la característica que se pretende medir.* Que diversos individuos tengan distintas opiniones sobre la variable estudiada, provoca la deseada diferencia entre las puntuaciones.
2. *Diferencias verdaderas en otras características estables del individuo que afectan al valor resultante.* Hace referencia a la influencia que variables como, el tamaño de la empresa, pueden ejercer sobre el valor de la variable latente que se mide, digamos, la orientación al mercado.
3. *Diferencias debidas a características transitorias del individuo,* se refieren al estado de ánimo, el cansancio o las prisas.
4. *Diferencias causadas por factores circunstanciales del entorno,* como que el individuo entrevistado esté solo o en compañía, en su hogar o en la calle.
5. *Diferencias ocasionadas por variaciones en la recolección de datos,* como sesgos introducidos por el encuestador.
6. *Diferencias debidas a la muestra de preguntas,* como cambios introducidos en el enunciado de las mismas por el entrevistador.
7. *Diferencias debidas a la falta de claridad en el instrumento de medición,* que hacen que el entrevistado no comprenda adecuadamente la pregunta.
8. *Diferencias causadas por factores mecánicos,* como falta de espacio para registrar la respuesta, o un error en el registro de la misma por el entrevistador.

La existencia de estas posibles fuentes de variación, obliga a realizar una serie de supuestos para que sea posible asumir que las medidas obtenidas de una variable latente son razonablemente próximas a la realidad. Estos supuestos son los siguientes:

- a. Sea X_{ij} la respuesta dada por el individuo i al ítem j de la escala de medida de la constructo latente Y (por ejemplo, la orientación a la innovación en Mejía-Trejo *et al.* (2016).
- b. En el contexto de la teoría clásica de la medición, X_{ij} puede considerarse como

función de la puntuación verdadera que tiene para el individuo i la variable medida (la orientación a la innovación para la empresa i), que denotaremos m_i y un error e_{ij} provocado por cualquiera de las fuentes de variación antes señaladas. Entonces:

$$X_{ij} = m_i + e_{ij}$$

- c. Pues bien, para poder desarrollar una escala como instrumento de medida del constructo o variable latente Y , y poder comprobar sus propiedades psicométricas (confiabilidad y validez) mediante los procedimientos que posteriormente expondremos, es necesario asumir los siguientes supuestos:
1. El error asociado a un ítem determinado es aleatorio, es decir, tiene una media de 0 cuando se considera una muestra amplia de individuos que responden al mismo.
 2. El error de un ítem de la escala no está correlacionado con el de otro ítem. La única conexión entre los ítems de la escala pasa por la variable latente, nunca a través de ningún término de error. Nótese en la figura 1 que no hay flechas entre los términos de error.
 3. Los términos de error no están correlacionados en la puntuación verdadera del constructo o variable latente. Nótese en la figura 1 que las flechas que salen de la variable latente van hasta los ítems de la escala, nunca llegan al término de error.

Tendencias en el dimensionamiento de los constructos para las ciencias de la administración

Es un hecho que las ciencias de la administración han tenido un repunte de crecimiento a partir de los años noventa, debido a factores como: el desarrollo de software estadístico (R, SPSS, SAS, BMDP, LISREL, EQS, CALIS, AMOS, etc.) cada vez más versátil y de fácil uso, disponible en versiones PC y laptop, para el diseño, creación y acceso a bases de datos con contenidos afines, así como la aportación cada vez más notable y creciente de investigadores incipientes y con experiencia que aportan al conocimiento de las ciencias de la administración con nuevos modelos, es que "...el desarrollo de la escala sigue siendo una industria en crecimiento....>" (Clark y Watson, 1995: 309). Esta declaración se aplica a todos los campos relacionados con las ciencias de la administración como la psicología, los negocios, la mercadotecnia, la competitividad, la administración del conocimiento, la innovación, la contabilidad, sistemas de información de gestión, gestión estratégica, comportamiento organizacional, etc. De hecho, tenemos varios textos que han realizado el esfuerzo por hacer un compilado de las escalas usadas con éxito en la mercadotec-

nia, así como el comportamiento organizacional, tales como: Price y Mueller (1986), Robinson, Shaver y Wrightsman (1991), Bruner y Hensel (1997), Bearden y Netemeyer (1998).

Para las ciencias de la administración, el hecho de crear modelos basados en constructos cada vez más sólidos, hace más evidente, el interés por diseñar, desarrollar y aplicar escalas que faciliten la explicación de problemas donde intervienen factores y variables cada vez más complejos. Lo anterior obliga a comprobarlos de manera objetiva de manera que los modelos aspiren a ser teorías a partir de la operacionalización de los constructos de interés. Así ocurre que cuando los constructos son medidos de forma correcta, basados en su confiabilidad y, se logran mejores bases para la prueba de la teoría. Cabe destacar, que el factor humano, implicado en todos los modelos de las ciencias de la administración obliga a que una escala utilizada una sola vez, requiera ser actualizada o refinada para reflejar y explicar mejor el constructo de interés. Así que no se debe descartar que nuevos artículos y/o tesis que presenten la medición de nuevas escalas, sean más bien actualizaciones derivadas de otras mediciones existentes, a fin de que se sienten con mayor precisión y expliquen más eficientemente los constructos de interés.

Es importante considerar que aún y cuando las tecnologías de información presenten avances notables en el software estadístico que permiten notablemente desarrollos más rápidos las escalas, no necesariamente hacen un mejor desarrollo de la escala. Se sugiere ser muy observante de los procedimientos recomendados que se propongan en los textos afines, así como de libros y artículos que han demostrado solidez en el desarrollo de escalas en entre ellos (Clark y Watson, 1995; DeVellis, 1991; Haynes *et al.*, 1999; Haynes *et al.*, 1995; Nunnally y Bernstein, 1994; Spector, 1992).

De esta manera se aprecia que no sólo ha habido una tendencia hacia más escalas, sino que también de los procedimientos utilizados para desarrollar y validar las mismas. Así, se han hecho evidentes pruebas de dimensionalidad de la escala, efectos de los métodos y distribución de la varianza (Bearden y Netemeyer, 1998). También es evidente una preocupación más pronunciada por el contenido y la validez de los elementos en la etapa de desarrollo, así como las consideraciones de longitud de escala (Clark y Watson, 1995; Haynes *et al.*, 1995).

Marco teórico y validez de los constructos subyacentes

Como se ha explicado hasta el momento, los constructos subyacentes no son directamente observables, por lo que no es posible cuantificarlos directamente, así

también las evaluaciones en sus mediciones tienen la posibilidad de ser altamente variables. Se debe hacer la anotación de que existen ya una serie de constructos latentes en las ciencias de la administración, que van desde aquellas que son muy amplias, como los de innovación abierta (Mejía-Trejo, 2017a) hasta constructos más específicos que pueden considerarse subcomponentes de constructos más amplios, como el de modelo de negocios abierto (Mejía-Trejo, 2017b). Lo anterior, hace notar, que los constructos subyacentes requieren un tratamiento más bien efecto en cuanto al nivel de abstracción y especificidad. Como es de esperar, es aquí donde se hace relevante la importancia de plantear y desarrollar un marco teórico sólido a fin de que las mediciones de los constructos de observación directa y/o subyacente no sean exagerados y por lo tanto, sobredimensionado. Esto no es nada nuevo, se tienen los trabajos de los clásicos sobre medición y validez, tales como Cronbach y Meehl (1955) y Loevinger (1957) quienes son los pioneros en manifestar de manera muy concreta, la importancia de la teoría en la medición.

Cabe destacar que para que las mediciones de constructos subyacentes tengan la pertinencia, significancia y, por lo tanto, relevancia en las ciencias de la administración, es fundamental y obligatoriamente necesaria, la sustentación del marco teórico. Es más aún, todo constructo subyacente extremadamente abstracto, pero discutido lo suficiente a nivel de marco teórico se dice que es más útil como antecedente o de consecuencia, incorporado en otros constructos subyacentes cuando están incrustados debidamente, en el marco teórico. Así, es posible afirmar que, la importancia de un o varios constructos subyacentes de un modelo que pretende explicar un fenómeno de las ciencias de la administración depende en gran medida de las teorías en las que se basa y formula. En otras palabras, está la interrogante: ¿qué predice la construcción latente de interés y/o qué predice la construcción latente? Las relaciones entre factores, variables y dimensiones que la explican han sido referidas como una red nomológica de un constructo subyacente (Cronbach y Meehl, 1955).

Hay enfatizar que el marco teórico debe involucrar modelos con base en constructos de factores y variables visibles, así como no visibles (subyacentes) de interés, sino también a la validez de la medición del constructo, el cual indica el verdadero impacto del modelo al demostrar su comprobación.

Pero, ¿qué significa la validez de constructo? En términos simples, es una evaluación del grado en que una medición mide realmente el constructo subyacente a la que se destina para medir. Cronbach y Meehl (1955) afirmaron que la validez del constructo implica al menos tres pasos:

- ▶ Especificar el conjunto de constructos teóricos y sus relaciones basados (una teoría);

- ▶ Desarrollar métodos para medir los constructos de la teoría; y
- ▶ Probar empíricamente cuán bien los indicadores manifiestos (observables) miden los constructos en la teoría, probando las relaciones hipotéticas entre los constructos de la teoría (es decir, la red nomológica).

Además, la evaluación de la validez de constructo es un proceso continuo. Un estudio que apoya la validez de una estructura no es suficiente para concluir que la medida ha sido validada. Se requieren múltiples pruebas y aplicaciones a lo largo del tiempo, y algunas de ellas pueden requerir un refinamiento del constructo mismo, así como su medición. Según Clark y Watson, "...las mediciones más precisas y eficientes son aquellas que se basan en una validez de constructo establecida... una teoría articulada que está bien soportada por datos empíricos..." (1995: 310).

Importancia de la revisión de la literatura del estado del arte en el marco teórico

El marco teórico es la base para formular y desarrollar una teoría ya que comienza con conceptualizaciones basadas en una revisión exhaustiva de la literatura del estado del arte. Estas revisiones de la literatura sirven a varios propósitos para los investigadores, entre ellos (más en el apartado La importancia de una teoría sólida: revisión de la literatura y su análisis):

1. *Debe alertar los intentos anteriores delimitando de intereses y teorías en las que el constructo puede resultar útil como variable independiente o dependiente.* Esto es, debe establecer los límites de los conceptos de factores, variables e indicadores de los modelos, sus fronteras de contenido, antecedentes, así como previsión de alcances de los mismos.
2. *Recopilar y revisar de forma rigurosa, escalas previas e intentos de implementación para medir el constructo.* Así, con base en la experiencia e intereses del investigador éste deberá determinar fortalezas, debilidades e incluso oportunidades y amenazas implícitas en la propuesta del modelo, tomando en cuenta incluso, factores políticos, económicos, tecnológicos, sociales, organizacionales, ambientales, etcétera.
3. *La implementación de cualquier escala, debe tomar en cuenta la relación objeto y sujeto de estudio.* En este sentido, es de relevante importancia la metodología empleada para la recopilación de los datos del sujeto de estudio, a fin de evitar en lo posible datos omisos, sesgos, errores, etc. que tiendan a desviar el resultado de la escala.

4. Antes de iniciar el trabajo de crear una escala, se surge la pregunta, de manera recursiva e insistente: ¿es necesario crear una nueva escala? Se debe reconocer que el proceso de generación, implementación y prueba de una nueva escala es un proceso largo y conflictivo. El desarrollo y la validación de escala es un proceso largo que implica esfuerzos y costos en el levantamiento de datos. De esta manera, la revisión exhaustiva y profunda de la literatura del estado del arte ayudará a responderla. Si ya existen previas escalas y mediciones del constructo, el valor de una nueva medida puede ser pequeño en relación con los costos involucrados en el desarrollo. Una nueva escala de medición debe mostrar pertinencia teórica y empírica sobre las escalas y mediciones de modelos existentes del mismo constructo para ser útiles.
5. Para que una nueva escala tenga al menos validez incremental sobre las medidas existentes, debe hacer una precisión mayor del constructo problema o hacerlo más eficiente (por ejemplo, que logre niveles notables de optimización para afirmar que es más barato, más corto, más fáciles de usar, de responder, que logre mejores niveles, etc.) que las mediciones existentes.
6. El investigador debe tener en mente que una revisión exhaustiva de la literatura del estado del arte, es la búsqueda de una oportunidad que le ayude a evitar la redundancia de desarrollar otra escala para evaluar un constructo ya bien medido.

Propiedades de una medición: dimensionalidad, confiabilidad y validez

Como se ha visto, el proceso de desarrollo de una escala comienza con una revisión exhaustiva de la literatura en la que se determina una definición teórica sólida del constructo y su dominio, de manera recursiva. Esta definición y descripción del constructo debe incluir lo que está dentro del dominio del constructo y lo que está excluido del dominio del constructo, así como su dimensionalidad derivados de la revisión exhaustiva de la literatura del estado del arte existente y, más aún, siendo altamente deseable el recopilar la opinión de expertos. En esencia, la definición del constructo y el dominio del contenido determinan la dimensionalidad teórica.

Dimensionalidad

La dimensionalidad de una medición se refiere a la homogeneidad de los ítems o indicadores. Básicamente, hay que considerar que una medición de tipo unidimensional, tiene propiedades estadísticas que demuestran que sus elementos

subyacen en un solo constructo o factor. Cuando la medición es multidimensional, los elementos pulsán más de una dimensión o factor.

El dominio de un constructo puede ser hipotético unidimensional, multidimensional, y/o como un factor de orden superior. Por lo tanto, la escala (o subescalas/factores) utilizada para operacionalizar el constructo debe reflejar la dimensión hipotética. Dado que la unidimensionalidad de escala (factor) es considerada un requisito previo para la confiabilidad y validez, la evaluación de la unidimensionalidad debe ser primordial (Cortina, 1993; Hattie, 1985; Schmitt, 1996).

Al momento, han sido empleados varios procedimientos para verificar la dimensionalidad de una escala (por ejemplo, el análisis de ítems o indicadores y el análisis factorial exploratorio, entre los más comunes). Una técnica muy conocida es el análisis factorial confirmatorio, en el que pueden especificarse y evaluarse varios factores con diversos indicadores (y las relaciones entre ellos) sobre los criterios utilizados para evaluar la dimensionalidad, por ejemplo: índices de ajuste, presencia de errores de medición correlacionados y grado de carga cruzada, entre otros. (Anderson y Gerbing, 1988; Clark y Watson, 1995; Floyd y Widaman, 1995; Gerbing y Anderson, 1988; Hattie, 1985; Kumar y Dillon, 1987).

Confiabilidad

La confiabilidad se refiere a la parte de la medición dada por efectos permanentes que persisten de una muestra a otra. Hay dos tipos de confiabilidad basados en la literatura psicométrica, referente en las ciencias de la administración:

Test-retest (o de estabilidad temporal): se trata de la correlación entre la puntuación de la misma persona en el mismo conjunto de ítems o indicadores, pero en dos puntos diferentes en el tiempo. Refiere a la estabilidad de las respuestas de un participante a lo largo del tiempo. Un coeficiente test-retest o "de estabilidad" generalmente se estima por la magnitud de la correlación entre las mismas medidas (y muestra) en diferentes ocasiones de evaluación. Si el coeficiente de estabilidad es bajo en magnitud, sin cambios en el constructo a lo largo del tiempo, la fiabilidad de la medida está en duda. Por lo tanto, la confiabilidad test-retest es útil porque ofrece información sobre el grado de confianza que se tiene de que la medición refleja del constructo y que es generalizable para otras ocasiones de evaluación (Haynes *et al.*, 1999). Curiosamente, la fiabilidad test-retest no se ha evaluado en el uso o desarrollo de la escala tan frecuentemente como la consistencia interna (Robinson *et al.*, 1991). Al momento, las estimaciones de test-retest están disponibles en muy pocas escalas de las ciencias de la administración, por lo que se sugiere utilizarla además de usar otros procedimientos de evaluación de confiabilidad y validez.

Consistencia interna (de evaluación de interrelación entre los indicadores): los

indicadores o ítems que componen la escala (o subescala) deben mostrar altos niveles de consistencia interna. Algunos criterios comúnmente utilizados para evaluar la consistencia interna son correlaciones individuales corregidas de ítem a total (*individual corrected ítem-to-total correlations*), el promedio de la correlación interitem entre los ítems de la escala (*the average interitem correlation among escale ítems*) y una serie de fiabilidad de coeficientes (*reliability coefficients*) (Churchill, 1979; Cortina, 1993; DeVellis, 1991; Nunnally y Bernstein, 1994; Robinson *et al.*, 1991).

El coeficiente de confiabilidad de consistencia interna (internal consistency reliability coefficient) más ampliamente utilizado, es el coeficiente alfa de Cronbach (Cronbach's coefficient alpha, 1951). Si bien, existen algunas reglas generales que se refieren a lo que constituye un nivel aceptable de coeficiente alfa, la longitud de la escala debe ser considerada. Esto es, a medida que aumenta el número de ítems o indicadores, el alfa tenderá a aumentar. Debido a que la parsimonia también es una preocupación en la medición (Clark y Watson, 1995; Cortina, 1993). Importante es preguntarnos ¿cuántos elementos se necesitan para medir un constructo? La respuesta depende parcialmente del dominio y las dimensiones del constructo. Naturalmente, un constructo con un dominio amplio y múltiples dimensiones requerirá más indicadores para aprovechar adecuadamente la relación dominio/dimensiones que un constructo con un dominio estrecho y de pocas dimensiones. Dado que la mayoría de las escalas son autoadministradas, la fatiga y/o la falta de cooperación de los encuestados deben ser consideradas, por lo que la brevedad de la escala es a menudo ventajosa (Churchill y Peter, 1984; Cortina, 1993; DeVellis, 1991; Nunnally y Bernstein, 1994). Con el advenimiento del modelado de ecuaciones estructurales, otras pruebas de consistencia interna o de estructura interna (o de estabilidad interna) se hicieron disponibles.

La confiabilidad compuesta (composite reliability o confiabilidad del constructo), que es similar al coeficiente alfa, se puede calcular directamente a partir de la producción de LISREL, CALIS, EQS O AMOS (Fornell y Larcker, 1981). Una prueba más estricta de la consistencia interna (estructura interna o estabilidad interna) implica evaluar la cantidad de varianza capturada por la medición de un constructo en relación con la cantidad de varianza debida al error de medida, esto es la varianza media extraída (*AVE. Average variance extracted*). Por medio del uso de una combinación de los criterios anteriores, es decir, correlaciones individuales corregidas de ítem a total (*individual corrected ítem-to-total correlations*), el promedio de la correlación interitem entre los ítems de la escala (*the average interitem correlation among escale ítems*), el alfa de Cronbach (*Cronbach's coefficient alpha*), la confiabilidad compuesta (*composite reliability*) y el AVE, las escalas

pueden desarrollarse de manera eficiente sin sacrificar la consistencia interna.

Validez

La *validez de constructo* se refiere a qué tan bien una medición mide realmente el constructo. La validez del constructo es el objetivo final en el desarrollo de un instrumento de evaluación y abarca todas las pruebas relativas a una medición (Haynes *et al.*, 1999). Aún existen desacuerdos en cuanto a la clasificación y los tipos de validez de constructo a aplicar, por lo que se tienen por el momento: convergente, discriminante, relacionado con el criterio (*criterion-related o predictiva*), nomológica y la validez de grupo conocido (*known-group validity*). Hasta aquí, podemos observar que todas las pruebas relativas a una medición contribuyen a establecer la validez del constructo, por lo que una medición también debe *a priori* mostrar su dimensionalidad teórica, así como dar evidencia de confiabilidad para ser considerada válida. La dimensionalidad y la fiabilidad son necesarias, pero las condiciones pueden ser insuficientes para la validez de constructo. Nuevamente, existen varios procedimientos para establecer la validez del constructo, siendo algunas:

Validez de contenido.

Se refiere al contenido de los ítems o indicadores. Se clasifican en dos subtipos: el contenido (*content validity*) y la validez aparente (*face validity*):

El término *validez de contenido* se ha definido de muchas maneras; en la mayoría se enfatiza que los ítems o indicadores de una medida son una muestra apropiada del dominio teórico de la construcción (Messick, 1993; Nunnally y Bernstein, 1994). La mayoría de las definiciones son consistentes con lo siguiente: la validez del contenido refleja el grado en que los elementos de un instrumento de evaluación son relevantes y representativos del constructo objetivo para un propósito de evaluación particular (Haynes *et al.*, 1995: 238). Los *elementos* se refieren al contenido de los ítems o indicadores individuales, formatos de respuesta e instrucciones a los encuestados, y representatividad se entiende como el grado en que los elementos son proporcionales a las facetas (*dominios*) del constructo objetivo y en la medida en que se ha muestreado todo el dominio. Es decir, los ítems deben aparecer consistentes con el dominio teórico del constructo en todos los aspectos, incluyendo formatos de respuesta e instrucciones.

En las escalas de desarrollo que son válidas para el contenido se recomienda, por lo regular, que se generen una serie de ítems o indicadores que aprovechen el dominio del constructo, que los indicadores sean examinados por expertos en la literatura, y que las pruebas piloto sean muestras de las poblaciones pertinentes

para refinar el conjunto de ítems o indicadores (DeVellis, 1991; Robinson *et al.*, 1991). Haynes *et al.* (1995) ofrecen una excelente descripción de los procedimientos para establecer la validez del contenido.

La validez aparente (face validity), trata de la simple apariencia que una medida tiene validez (Kaplan y Saccuzzo, 1997: 132). Aunque los términos validez aparente y validez de contenido (*content validity*) se han utilizado indistintamente, algunos argumentan que la validez aparente debe estar separada de la validez de contenido (Anastasi y Urbina, 1998; Nevo, 1985). Otros van un poco más lejos e indican que la validez aparente vs. la validez del contenido se define en términos de investigador y encuestado. Un instrumento con una gran carga de validez aparente mejora su uso en situaciones prácticas, entre otras cosas, induciendo la cooperación de los encuestados a través de la facilidad de uso, el nivel de lectura adecuado, la claridad, la facilidad de leer las instrucciones y formatos de respuestas fáciles de usar. Una definición un tanto aceptada de la validez aparente implica que un instrumento o prueba, cuando se utiliza "...debe ser práctico, pertinente y relacionado con los propósitos del instrumento [prueba], además de tener validez pragmática o estadística", es decir, no sólo debe ser válida, sino que también debe parecer válida para los encuestados (Nevo, 1985: 228). Por lo tanto, la validez aparente puede estar más preocupada con lo que los encuestados de las poblaciones relevantes infieran con respecto a lo que se está midiendo.

Validez convergente es el grado en que se relacionan dos medidas diseñadas para medir el mismo constructo. La convergencia se encuentra si las dos medidas diferentes del mismo constructo están altamente correlacionadas.

Validez discriminante valora el grado en que se relacionan dos medidas diseñadas para medir constructos similares pero conceptualmente diferentes. Una correlación de baja a moderada a menudo se considera una evidencia de validez discriminante. Las matrices multi-trama-multimétodo (MTMM *Multitrait-multimethod matrices*) se utilizan con frecuencia para evaluar la validez convergente y discriminante cuando se requieren métodos de medición diferentes, es decir, autoinforme vs. observacional (Campbell y Fiske, 1959).

La validez nomológica se ha definido como el grado en que las predicciones de una red teórica formal que contiene el concepto bajo escrutinio son confirmadas (Campbell, 1960). Evalúa el grado en el que los constructos que están teóricamente relacionados lo están también empíricamente (es decir, sus medidas se correlacionan significativamente en la dirección prevista). También existen pautas para establecer la validez nomológica, pero también han sido criticadas (Peter, 1981).

Al igual que con la consistencia interna y la validación convergente y discriminante, recientemente se han utilizado paquetes de ecuaciones estructurales para evaluar la validez nomológica de las medidas de escala. Varios libros (por ejemplo, Bollen, 1989; Byrne, 2001; Hayduk, 1996; Hoyle, 1995; Schumacker y Lomax, 1996) y artículos (por ejemplo, Anderson y Gerbing, 1988; Bagozzi, Yi y Phillips, 1991; Bentler y Chou, 1987) ilustran técnicas de modelado, criterios evaluativos y guías para lo que constituye validez nomológica.

Validez relacionada con el criterio (criterion-related validity) varían, y algunas definiciones son similares a otros tipos de validez. Por ejemplo, se denomina validez relacionada con el criterio al grado en que una medida covaría con medidas previamente validadas (*gold-standard*) de los mismos constructos (Haynes *et al.*, 1999). Esta definición es similar a la de la validez convergente. La validez relacionada con el criterio significa también la medida en que una medida corresponde a otra medida de interés (Kaplan y Saccuzzo, 1997). Algunos afirman que la validez basada en criterio es la misma que la validez predictiva, o relación funcional entre un predictor y un criterio antes, durante o después de aplicar un predictor (Nunnally y Bernstein, 1994). Este enfoque se basa en la relación temporal entre el predictor y su criterio, es decir, la validez pos-dictiva, concurrente y predictiva. Lo que la mayoría de las definiciones tienen en común es que la validez relacionada con los criterios se evalúa mediante un patrón de relaciones teóricamente especificadas entre una medida y un criterio a menudo denominado coeficiente de validez (*validity coefficient*).

Validez del grupo conocido (known-group validity) implica la capacidad de la medida para distinguir de manera fiable entre grupos de personas que deben anotar alto en el rasgo y bajo en el rasgo. Como ejemplo, una persona que es verdaderamente conservadora debe obtener una puntuación significativamente mayor en una escala de conservadurismo que una persona liberal y las personas de ventas en el negocio de inmobiliarias y grandes empresas farmacéuticas deben diferir en sus niveles de orientación al cliente (Saxe y Weitz, 1982). Por lo tanto, las diferencias de puntuación media entre los grupos de un constructo dado pueden utilizarse como evidencia de la validez del grupo conocido. Ver más en Jarvis y Petty (1996).

El rol del investigador en las ciencias de la administración

Los investigadores deben mantener un perfil adecuado con sus funciones en la creación de escalas, tal como se sugiere en la tabla 2.

TABLA 2. Rol del investigador en la creación de escalas

Rol	Descripción
Creativo	Es fundamental no relegar a un segundo término el ingenio y la creatividad; la creatividad es absolutamente necesaria, pese a que tiende a utilizar métodos poco aceptados.
Competencia	Consiste en aplicar cada método en su justa medida y en su ámbito de acción; la metodología científica no puede utilizarse para explicar qué está bien y qué está mal.
Ético	El investigador no escapa a la influencia que pueda ejercer sobre ciertos colectivos al utilizar determinados métodos, pero debe respetar las normas de comportamiento al desarrollar su estudio.
Técnico	Para realizar una investigación con garantías es necesario que el investigador estar adiestrado en las técnicas de investigación, poseer suficiente bagaje metodológico, ser conocedor avanzado de su campo de estudio y dominar las técnicas cuantitativas.
Comunicador	Se puede decir que no existe la investigación que no se comunica, por ello las dos tareas fundamentales son: a) Informar a la comunidad sus líneas de trabajo, con ello se suelen evitar duplicaciones y se incrementa la cooperación entre colegas. b) Informar a los colegas, las empresas y la sociedad de los hallazgos realizados, de su aplicabilidad y de su posible idoneidad para mejorar aspectos concretos.
Imparcialidad	El investigador necesita cierto grado de aislamiento respecto de su entorno y de inmunidad respecto del sistema burocrático de su organización, tratando de ser lo más neutral y objetivo posible.
Escepticismo	El científico no es un creyente que trata de validar sus opiniones o presunciones, sino que debe, por principio, dudar de las explicaciones oportunistas, vulgares, demagógicas o que no se han obtenido mediante una metodología científica.

Fuente: (Aldás-Manzano y Maldonado-Guzmán, 2016).

El cientificismo

Uno de los peligros que acecha al investigador es el entendido como tal, a toda aquella persona, argumento o propuesta que tiene la pretensión no legítima de tener o aparentar naturaleza científica. En esta definición hay tres elementos importantes a considerar (Aldás-Manzano y Maldonado-Guzmán, 2016):

- ▶ La pretensión, es la actitud o comportamiento de un individuo.
- ▶ La ilegitimidad, refiere a no guardar las normas de la comunidad científica o a

utilizar argumentos, elementos y teorías no aceptadas en ella.

- La naturaleza científica del fenómeno o la calificación profesional de los sujetos que sustentan un argumento o propuesta.

No obstante, tratar de sustentar la validez universal de esta definición y de sus connotaciones sería ya de por sí una actitud inadecuada, pues hay importantes “divergencias sobre lo que está bien o lo que está mal, sobre lo que es legítimo e ilegítimo, lo que es científico y lo que es científicista: lo que para unos es ciencia, para otros es científicismo”.

TABLA 3. Características para diseñar una escala de medición

Característica	Descripción
Viable	Significa que el estudio debe ser realizable, y que el problema objeto de análisis es alcanzable; muchas veces puede haber problemas derivados del propio investigador o del tema elegido.
Científico	El concepto a transmitir es claro: el tem a debe tener valor científico; para que suceda la metodología debe ser rigurosa, adaptarse al objetivo de estudio y ser factible de analizar y aplicar.
Propio	El problema debe enmarcarse dentro del área de conocimiento esto no quiere decir que haya que huir de temáticas multidisciplinares, sino que hay que rodearse de un equipo también multidisciplinar.
Preciso	El problema a investigar debe quedar claramente formulado y delimitado; que omitir temas ambiguos y genéricos.
Significante	Aunque es perfectamente posible realizar un estudio sobre un caso aislado, lo ideal es que el tema elegido permita obtener hallazgos susceptibles de generalización o, en su defecto, que representen el comportamiento de un colectivo o un fenómeno para un periodo de tiempo lo más amplio posible.
Importante	La investigación está para buscar conocimiento, por lo que el “ <i>problema elegido debe representar alguna novedad</i> ”, en determinadas ocasiones, comprobar ciertos hechos o comportamientos es una tarea razonable y justificada.
Motivación	Llama la atención la gran cantidad de investigadores escasamente motivados por su temática; si no hay predisposición y entusiasmo hacia el tema que se estudia, se termina aplicando la ley del <i>mínimo esfuerzo</i> , lo que produce detrimento en la calidad del trabajo.
Enfoque	Hay que destacar la fuerte tendencia de muchos investigadores a no plantear cuestiones sobre las teorías existentes y, así, analizar y describir fenómenos o modelos procedentes de dichas teorías.

Fuente: (Aldás-Manzano y Maldonado-Guzmán, 2016).

Los argumentos científicas, muchas veces, suelen basarse en analogías y pretenden:

- ▶ Ampliar una disciplina a ámbitos no aceptados actualmente.
- ▶ Aplicar directamente conocimientos de otras disciplinas sin considerar su naturaleza, objeto y condiciones.
- ▶ Dar por válidos argumentos o teorías no adecuados o insuficientemente probados por la comunidad científica.

Determinación del problema a investigar

La investigación inicia cuando se determina qué problema o fenómeno hay que estudiar y por qué hay que hacerlo, y no es suficiente saber la temática sobre la que se ha de trabajar. También hay que tener en claro el por qué de la elección del tema (importancia, idoneidad, oportunidad), para que al momento de elegir un tema, éste debe ser de acuerdo con las características indicadas en la tabla 3.

La medición en las ciencias de la administración

El investigador debe siempre tomar en cuenta que su aportación en las ciencias de la administración se basará en el diseño pertinente y adecuado de una escala de medición, por lo que se sugiere considerar los siguientes aspectos (tabla 4).

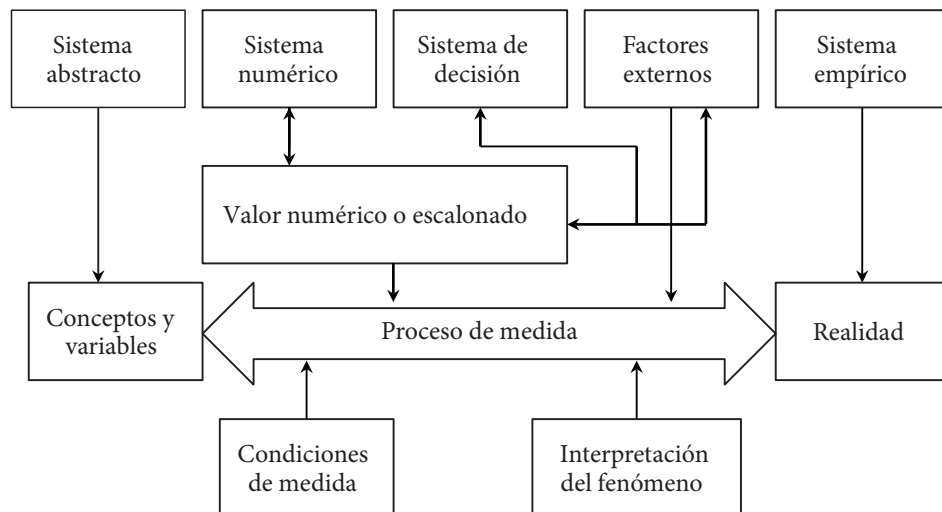
Churchill (1979) afirma que cuando un investigador pone en conexión el sistema abstracto con el sistema empírico utilizando números para valorar e interpretar los fenómenos objeto de su estudio, es cuando se produce una medición (tabla 5).

TABLA 4. Puntos básicos en el diseño de una escala de medición

Las escalas conforman la parte fundamental del cuestionario
La escala está compuesta por diferentes ítems o preguntas
Los enunciados de los ítems deben ser: <ul style="list-style-type: none">◆ Cortos (20 palabras aproximadamente).◆ Lenguaje simple.◆ Expresen una sola idea.◆ Significado unívoco, sin ambigüedades.◆ Evitar enunciados con negaciones.◆ Misma proporción de elementos en sentido favorable y desfavorable.◆ Evitar términos absolutos (siempre, todas, nadie...) salvo que el objetivo sea la polarización.

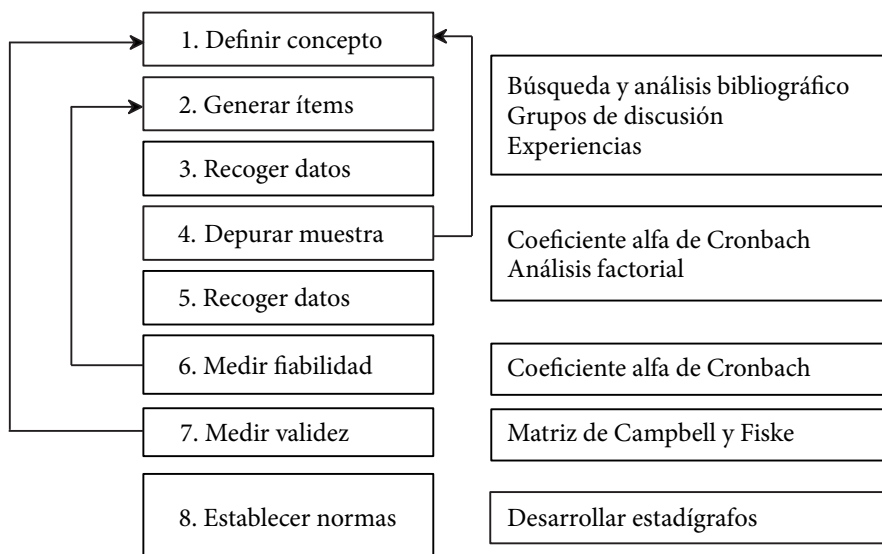
Fuente: Elaboración propia.

TABLA 5. Proceso para generar escala de medidas



Fuente: Fuente: Churchill (1979) en (Aldás-Manzano y Maldonado-Guzmán, 2016).

TABLA 6. Proceso de creación de escalas



Fuente: Churchil (1979).

TABLA 7. Concepto y variables: definición y tipos

Conceptos y variables		
<ul style="list-style-type: none"> - El primer paso para poder medir es desarrollar los conceptos entendiendo que concepto es “una representación literaria y abstracta de una realidad o idea” - Para que sea completo el concepto es necesario que todas las dimensiones de la mencionada realidad o idea estén perfectamente delimitadas, clarificadas, expresadas en lenguaje consensuado y basadas, siempre que sea posible, en desarrollos anteriores que le sean afines. - Por otra parte, que el concepto sea una representación literaria no impone que su estructura no sea científica, o bien, al contrario, al enunciar un concepto el propósito ha de ser científico. - En muchos casos, los conceptos se construyen a partir de generalizaciones de casos concretos o se limitan a plantear representaciones ideales de algún aspecto de la realidad. En cualquier caso, es imprescindible que la construcción de conceptos cumpla al menos con los siguientes criterios: <ul style="list-style-type: none"> - El contenido no ha de ser circular - Los elementos que lo definen han de ser precisos y unitarios, por ello debe rechazarse todo concepto que en su formulación utilice expresiones dialógicas. - La expresión (definición) del concepto ha de señalar los elementos esenciales de lo que se trata de definir. - Debe tener <i>potencial operacionalizador</i> es decir, contar con la capacidad para ser puestos a prueba. No se deben considerar aquellos conceptos que por su propia naturaleza no puedan ser comprobados directa o indirectamente. - Los conceptos pueden ser clasificados en <i>objetivos, subjetivos, científicos y literarios</i>. Los científicos se pueden dividir, a su vez, en generales (sirven para muchas disciplinas, desarrollando fenómenos o ideas de rápida, fácil y amplia aplicación) y específicos (sólo tienen sentido operativo en un campo concreto). Un ejemplo de ello se muestra a continuación: 		
	Objetivos	Subjetivos
Generales	Dimensión Saturación	Calidad/e stilo de vida Liderazgo
Específicos	Apalancamiento Participación de mercado	Flexibilidad organizativa Lealtad a la marca
Requisitos de variables e índices		
<p>Así también, se deberá guardar una clara diferencia entre variables e índices, como se describe en el recuadro</p>		
Variable	Índice	
<ul style="list-style-type: none"> - Responden a una definición formal. - Relacionarse con el contenido significativo del concepto. - Discriminación del contenido. - Capacidad para ser medida. 	<ul style="list-style-type: none"> - Capacidad definitoria. - Debe tener relación con el contenido significativo de la variable. - Internamente consistente. - Ofrecer una valoración específica. 	

Continúa...

Ejemplos de lo anterior		
Conceptos	VARIABLES	Índice (para cada variable)
Relaciones en el canal de distribución.	Conflictos en el canal Control del canal. Poder en el canal.	Escala de Schul (1987). Escala de Dwey y Welsh (1985). Escala de El-Ansary y Stern (1972).
Segmento de mercado.	Demográfica económica. Psicográfica /comportamiento.	Densidad de población. Nivel de ingresos VALS Frecuencia de uso/común.
Comportamiento de compra.	Motivación social. Grado de innovación. Fidelidad a la marca.	Índice de Moschis (1978). Escala de Hawes y Lumpkin (1984). Tasa de recompensa.

Valor numérico y escalas naturales

Una vez definidas las variables, se debe reconocer cuál es la naturaleza de su valor numérico o como se sabe, las valoraciones que se otorgan a las variables provienen del sistema numérico, con las siguientes características:

Universalidad: el proceso de numeración es estándar y consensuado

Operacionalidad: los números admiten operaciones y manipulaciones

No Limitación: el conjunto de los números es ilimitado, tanto categóricamente como dentro de un intervalo o rango dado. No obstante, la serie de los números se puede truncar a voluntad por el investigador.

Identificación: cada número representa una situación concreta de la cualidad que representa e implica una valoración exclusiva.

Ordenación: la asignación de números a diversos sujetos considerando una cualidad determinada implica una orden natural si dichos sujetos se comparan. Los números poseen una ordenación (0, 1, 2, ..., 9) y la transmiten cuando son asignados a cualidades de los objetos o de los fenómenos.

Distancia: se pueden cuantificar las diferencias entre los números, y tales diferencias tienen sentido matemático.

Origen: la medición tiene un origen, cuyo significado es la ausencia de valoración (0). Los números negativos indican la dirección del sistema y de su interpretación

Problemas en la definición de escalas

No todas las mediciones son fáciles de realizar; la mayoría de las variables utilizadas no son directamente observables (calidad de servicio) o no se les puede otorgar una valoración directa (imagen de marca), en estos casos se debe crear un instrumento de medida que permita medir la variable y estimar la naturaleza del concepto de forma fiable y válida. En las ciencias de la administración, donde el estudio por encuesta está bastante generalizado, se utilizan escalas compuestas como las psicológicas (actitudes) y las sociológicas. En el campo del *marketing*, Bruner y Hensel (1997) han recopilado más de 800 instrumentos (escalas), y Bearden y Netemeyer (1998) han analizado otros 124. De cualquier forma, un investigador que necesite realizar un estudio puede acudir a las publicaciones o desarrollar su propio instrumento de medida.

Fuente: Netenmeyer *et al.* (2002)..

Los argumentos científicistas, muchas veces, suelen basarse en analogías y pretenden:

- ▶ Ampliar una disciplina a ámbitos no aceptados actualmente.
- ▶ Aplicar directamente conocimientos de otras disciplinas sin considerar su naturaleza, objeto y condiciones.
- ▶ Dar por válidos argumentos o teorías no adecuados o insuficientemente probados por la comunidad científica.

Determinación del problema a investigar

La investigación inicia cuando se determina qué problema o fenómeno hay que estudiar y por qué hay que hacerlo, y no es suficiente saber la temática sobre la que se ha de trabajar. También hay que tener en claro el por qué de la elección del tema (importancia, idoneidad, oportunidad), para que al momento de elegir un tema, éste debe ser de acuerdo con las características indicadas en la tabla 3.

La medición en las ciencias de la administración

El investigador debe siempre tomar en cuenta que su aportación en las ciencias de la administración se basará en el diseño pertinente y adecuado de una escala de medición, por lo que se sugiere considerar los siguientes aspectos (tabla 4).

Churchill (1979) afirma que cuando un investigador pone en conexión el sistema abstracto con el sistema empírico utilizando números para valorar e interpretar los fenómenos objeto de su estudio, es cuando se produce una medición (tabla 5).

Diversas operativas se han sugerido para crear medidas, siendo la de Churchill (1979) la más aceptada por la comunidad investigadora (tabla 6).

Concepto y variables: definición y tipos

A manera de repaso, mostramos la tabla 7 (Mejía-Trejo, 2017c, capítulos 2 y 3)

Etapas para el desarrollo de una escala

Existen numerosos artículos y libros acerca de cómo desarrollar una escala, entre ellos Churchill (1979); DeVellis (1991); Spector (1992); Nunnally y Bernstein (1994); Clark y Watson (1995); Haynes et al. (1999) de los que sugieren pasos y procedimientos que varían con base en los objetivos y propósitos de la medición,

pero que comparten un conjunto común de pautas para el desarrollo de la escala (tabla 8).

Conclusión

Para que un constructo sea valioso, debe tener relevancia teórica y/o práctica para el investigador en ciencias de la administración. Por lo tanto, debe considerarse cuidadosamente lo que predice el constructo de interés y/o lo que predice el constructo de interés. Aquí, la noción de teoría y conocer la literatura es muy importante. Además, dada la importancia de la medición en las ciencias sociales, cualquier medida debe ser válida para permitir construir inferencias confiables a partir de estudios empíricos. Dicha validez se basa en cuán bien el constructo subyacente que se mide se basa en teoría. También en este capítulo introductorio, hemos resaltado los conceptos de dimensionalidad, fiabilidad y validez, así como resumimos una serie de pasos para derivar medidas con propiedades adecuadas. El resto de nuestro texto elabora sobre dimensionalidad, fiabilidad y validez, y los cuatro pasos en la construcción de escala.

TABLA 8. Etapas para la creación de una escala

Etapa	Descripción
<p>1 Definición del constructo y dominio de contenido</p>	<p>Como hemos dicho a lo largo de este capítulo introductorio, la importancia de la teoría en el desarrollo de la escala no puede ser exagerada, y desarrollar y refinar una teoría requiere una revisión bibliográfica exhaustiva. Durante la revisión de la literatura del estado del arte y los procesos de desarrollo de la teoría, se deben enfatizar varias cuestiones:</p> <ol style="list-style-type: none"> a. La importancia de una definición clara del constructo, el dominio del contenido y el rol de la teoría y el marco teórico de soporte; b. El enfoque de los ítems o indicadores como <i>efecto reflectivo (reflective)</i> vs. indicadores formativos (<i>formative</i>). c. Dimensionalidad del constructo: ¿unidimensional, multidimensional o construcción de orden superior?
<p>2 Generación y clasificación de indicadores de medición</p>	<p>Se deben considerar varias cuestiones, como:</p> <ol style="list-style-type: none"> a. Suposiciones teóricas sobre los ítems o indicadores. b. Generar indicadores potenciales y determinar el formato de respuesta, es decir: <ul style="list-style-type: none"> - Cuántos ítems como grupo inicial. - Formato dicotómico vs. multicotómico. - Problemas con las palabras en los indicadores. c. El enfoque en la validez de contenido en relación con la dimensionalidad teórica, y d. Clasificación de indicadores por académicos, expertos e interesados validez por contenido (<i>content validity</i>) y validez aparente (<i>face validity</i>).
<p>3 Diseño de la escala</p>	<p>Una vez que se ha generado y juzgado un grupo adecuado de ítems, las pruebas empíricas de los ítems en muestras relevantes son el siguiente paso. Entre las cuestiones y procedimientos que deben considerarse se incluyen:</p> <ol style="list-style-type: none"> a. La prueba piloto como procedimiento de reducción-optimización de ítems o indicadores; b. El uso de varias muestras de poblaciones relevantes para el desarrollo de la escala; c. El diseño de estudios para probar las propiedades solicitadas como ciencias de la administración; d. Análisis inicial de ítems o indicadores vía análisis factorial exploratorio (EFA. <i>exploratory factor analysis</i>); e. Análisis inicial de ítems o indicadores y estimados de consistencia interna; f. Estimados iniciales de validez; y g. Retención de ítems para el siguiente conjunto de estudios.
<p>4 Finalización de la escala</p>	<p>Se deben utilizar varios estudios para ayudar a finalizar la escala. Muchos de los procedimientos utilizados y cuestiones relacionadas con el perfeccionamiento de la escala también serán aplicables para derivar la forma final de la escala. Esto incluye:</p> <ol style="list-style-type: none"> a. La importancia de varias muestras de las poblaciones pertinentes; b. El diseño de los estudios para probar los diferentes tipos de validez; c. El análisis de los ítems o indicadores a través del EFA: <ul style="list-style-type: none"> -Consistencia del EFA en los resultados entre los pasos 3 al 4; -Derivación de una estructura factorial inicial con teoría y dimensionalidad; d. Análisis de ítems e indicadores con análisis factorial confirmatorio (CFA. <i>confirmatory factor analysis</i>); <ul style="list-style-type: none"> - Prueba de la estructura factorial teórica y el modelo especificado; - Estudios de cruce de modelo factorial de invarianza (por ejemplo, análisis multigrupo) e. Análisis de ítem o indicadores adicionales vía estimados de consistencia interna. f. Estimaciones adicionales de validez; g. Establecimiento de normas entre los estudios, y

Fuente: Netenmeyer *et al.* (2003).

CAPÍTULO 2

Dimensionalidad

Esta primera etapa de la dimensionalidad de los constructos, es una parte importante del proceso de desarrollo de la escala. Es prácticamente imposible desarrollar buenas medidas de un constructo sin conocimiento de su dimensionalidad. Por ejemplo, considere la tarea de desarrollar una escala para medir el liderazgo. ¿Tiene el liderazgo una sola faceta o dimensión (es decir, es unidimensional), o tiene múltiples facetas o dimensiones (es decir, es multidimensional)? Si el liderazgo es un constructo unidimensional, entonces se requiere un sólo número para medirlo. Sin embargo, si se suscribe a la teoría MLQ5X (*multifactor leadership questionnaire level 5*) de Avolio y Bass (2004; Mejía-Trejo *et al.*, 2014), el liderazgo comprende las siguientes cuatro variables: transformacional, transaccional, pasivo y resultados de liderazgo con doce dimensiones y 45 indicadores. Aún así, se tiene que la dimensionalidad ha llegado a significar aspectos muy diferentes a los investigadores sustantivos. El propósito de este capítulo es discutir el significado de dimensionalidad de constructos y su relación con la confiabilidad.

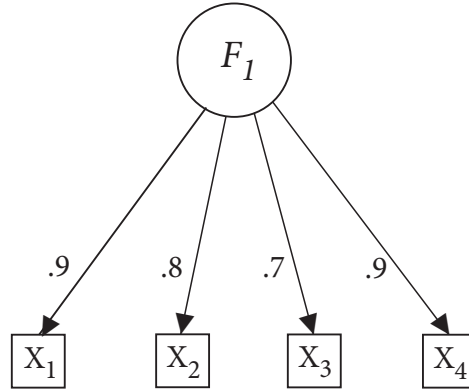
Específicamente, abordaremos las siguientes dos preguntas:

- ▶ Qué es la unidimensionalidad? y
- ▶ ¿Existe una distinción entre la dimensionalidad de un constructo, la dimensionalidad de un conjunto de elementos manifiestos utilizados para medir el constructo y la dimensionalidad de sólo los elementos manifiestos utilizados para medir el constructo?

Unidimensionalidad de constructo, de indicadores y de conjunto de indicadores

En la figura 1, se aprecia como la variable F1 es la causa del efecto de las variables X_1 , X_2 , X_3 y X_4 .

FIGURA 1. Modelo de 1 factor con relación total de las variables.



Fuente: Elaboración propia

A su vez, es presentada la matriz de correlación de la tabla 1, en la que de acuerdo con nuestro enfoque de mediciones efecto reflectivo (*reflective*) o causal formativo (*formative*) se puede apreciar en la figura 1 que F_1 afecta X_1 , X_2 , X_3 y X_4 . También se supone que las variables X se pueden medir u observar, mientras que F_1 no se puede medir ni observar.

TABLA 1. Matriz de correlación del modelo de la figura 1.

	F_1	x_1	x_2	x_3	x_4
F_1	1.00				
x_1	0.90	1.00			
x_2	0.80	0.72	1.00		
x_3	0.70	0.63	0.56	1.00	
x_4	0.90	0.81	0.72	0.63	1.00

Fuente: SPSS IBM 21, con datos propios.

Es decir, F_1 típicamente se denomina como constructo no observable, latente o subyacente, y las variables X se designan indicadores, elementos o variables manifiestas del constructo latente o subyacente. La correlación parcial, $r_{12 \cdot 3}$, entre cualquier par de variables (es decir, las variables 1 y 2) después de *remover* o *parcializar* el efecto de una tercera variable (es decir, la variable 3) está dada por la ecuación 1.

ECUACIÓN 1. Correlación parcial $r_{12\cdot3}$

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}} .$$

Así, utilizando los valores de la tabla 1 y la ecuación 1, la correlación parcial entre X_1 , X_2 después de la aparición del efecto de F_1 es igual a

$$\frac{0.72 - 0.9 \times 0.8}{\sqrt{1 - .9^2} \sqrt{1 - .8^2}} = 0.$$

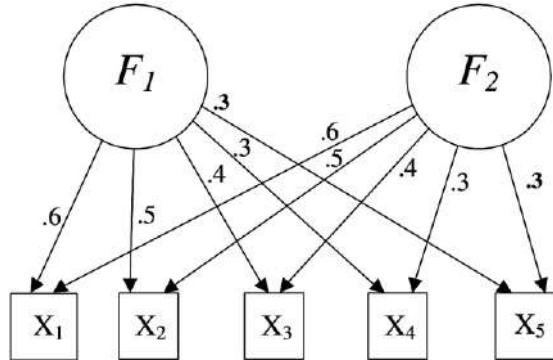
Así, al usar la ecuación 1, se puede demostrar fácilmente que las correlaciones parciales entre todos los pares de variables X dados en la tabla 1 son iguales a cero. Es decir, una vez que el efecto de F_1 ha sido removido o parcializado, las correlaciones parciales o las relaciones entre las variables X desaparecen. En otras palabras, F_1 es responsable de todas las relaciones entre las variables X ; por lo tanto, F_1 se conoce como *factor común*.

El conjunto de indicadores X_1 , X_2 , X_3 y X_4 se considera que es unidimensional porque las correlaciones entre ellos, después de haber sido parcializadas para el efecto de un único factor común (es decir, F_1), son iguales a cero. Por lo tanto, un conjunto de elementos se considera unidimensional si las correlaciones entre ellos pueden ser contabilizados por un solo factor común (McDonald, 1981; Hattie, 1985). Observe que cada uno de los cuatro índicadores de la figura 1 es una medida de uno y sólo un constructo y, por lo tanto, cada indicador es unidimensional. Es decir, se considera que un elemento es unidimensional si es una medida de un solo constructo o factor latente o subyacente. Ahora bien, si se toman múltiples conjuntos de n elementos del dominio del constructo y las correlaciones parciales entre cada conjunto son iguales a cero, entonces se dice que el constructo es unidimensional. Sin embargo, es posible que un conjunto de indicadores pueda ser unidimensional, pero el constructo o los indicadores individuales pueden no ser unidimensionales, circunstancia que se explicará más adelante.

Multidimensionalidad de constructo, de indicadores y de conjunto de indicadores

Observe la figura 2, la cual muestra como la variable F_1 es la causa del efecto de las variables X_1 , X_2 , X_3 , X_4 y X_5 , así como la variable F_2 es la causa del efecto de las variables X_1 , X_2 , X_3 , X_4 y X_5 .

FIGURA 2. Modelo de dos factores con relación total de las variables.



Fuente: Elaboración propia

La tabla 2 presenta la matriz de correlación para el modelo de dos factores dado en la figura 2.

TABLA 2. Matriz de correlación del modelo de la figura 2.

	F_1	F_2	x_1	x_2	x_3	x_4	x_5
F_1	1.00						
F_2	0.00	1.00					
x_1	0.60	0.60	1.00				
x_2	0.50	0.50	0.60	1.00			
x_3	0.40	0.40	0.48	0.40	1.00		
x_4	0.30	0.30	0.36	0.30	0.24	1.00	
x_5	0.30	0.30	0.36	0.30	0.24	0.18	1.00

Fuente: SPSS IBM 21, con datos propios.

En la que las variables X se ven afectadas por los factores comunes F_1 y F_2 . La correlación parcial entre X_1 y X_2 después de que el efecto de F_1 es parcializado, es igual a:

$$\frac{0.60 - 0.6 \times 0.5}{\sqrt{1 - 0.6^2} \sqrt{1 - 0.5^2}} = 0.43.$$

Similarmente, la correlación parcial entre X_1 y X_2 después de que el efecto de F_2 es parcializado, es igual a:

$$\frac{0.60 - 0.6 \times 0.5}{\sqrt{1 - 0.6^2} \sqrt{1 - 0.5^2}} = 0.43.$$

Por lo tanto, ambas correlaciones parciales no son iguales a cero. Esto sugiere que las correlaciones entre las variables no pueden ser contabilizadas por un solo factor, de tal manera que el conjunto de elementos no es unidimensional. La correlación parcial entre cualquiera de las dos variables 1 y 2 después de parcializar el efecto de las variables 3 y 4 está dada por la ecuación 2.

ECUACIÓN 2. Correlación parcial $r_{12 \cdot 34}$

$$r_{12 \cdot 34} = \frac{r_{12 \cdot 4} - r_{13 \cdot 4} \times r_{23 \cdot 4}}{\sqrt{1 - r_{13 \cdot 4}^2} \sqrt{1 - r_{23 \cdot 4}^2}}$$

De esta forma, al usar la ecuación 2, la correlación parcial entre X_1 y X_2 después de la parcelación de los efectos de F_1 y F_2 es igual:

$$r_{12 \cdot F_1} = \frac{.60 - .6 \times .50}{\sqrt{1 - .6^2} \sqrt{1 - .5^2}} = .43$$

$$r_{12 \cdot F_2} = \frac{.60 - .6 \times .50}{\sqrt{1 - .6^2} \sqrt{1 - .5^2}} = .43$$

$$r_{12 \cdot F_1 F_2} = \frac{r_{12 \cdot F_2} - r_{1F_1 \cdot F_2} \times r_{2F_1 \cdot F_2}}{\sqrt{1 - r_{1F_1 \cdot F_2}^2} \sqrt{1 - r_{2F_1 \cdot F_2}^2}}$$

$$r_{1F_1 \cdot F_2} = \frac{r_{1F_1} - r_{1F_2} \times r_{F_1 F_2}}{\sqrt{1 - r_{1F_2}^2} \sqrt{1 - r_{F_1 F_2}^2}} = \frac{.6 - .6 \times 0}{\sqrt{1 - .6^2} \sqrt{1 - 0}} = .75$$

$$r_{2F_1 \cdot F_2} = \frac{r_{2F_1} - r_{2F_2} \times r_{F_1 F_2}}{\sqrt{1 - r_{2F_2}^2} \sqrt{1 - r_{F_1 F_2}^2}} = \frac{.5 - .5 \times 0}{\sqrt{1 - .5^2} \sqrt{1 - 0}} = .58$$

$$r_{12 \cdot F_1 F_2} = \frac{.43 - .75 \times .58}{\sqrt{1 - .75^2} \sqrt{1 - .58^2}} \approx 0$$

Es posible demostrar fácilmente que las correlaciones parciales entre el conjunto de variables de X_1 a X_5 , después de controlar para F_1 y F_2 son todas iguales a cero. Es decir, son necesarios dos factores o constructos latentes o subyacentes para explicar las correlaciones entre las variables X_s y, por lo tanto, dos dimensiones representan el conjunto de variables X_1 a X_5 . La dimensionalidad de un conjunto dado de variables es igual al número de constructos latentes o subyacentes necesarios para dar cuenta de las correlaciones entre las variables.

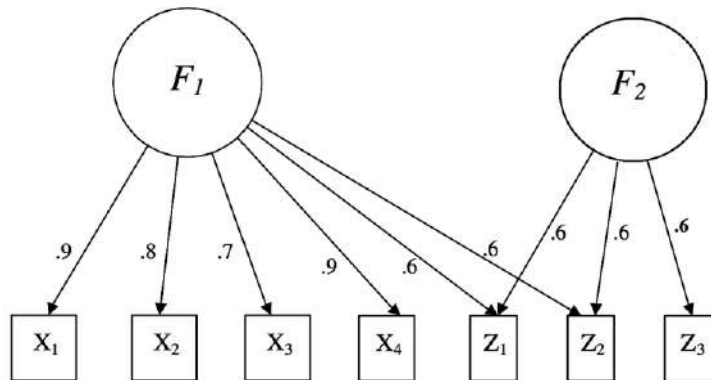
Observe además que cada uno de los cinco indicadores representa a dos constructos. Es decir, la dimensionalidad de cada indicador o constructo es igual a dos, porque cada uno de estos elementos es una medida de dos constructos. Una vez más, si se toman múltiples conjuntos de elementos del dominio de los dos constructos y las correlaciones parciales entre las variables de cada conjunto después de eliminar el efecto de F_1 y F_2 son iguales a cero, entonces se dice que el constructo es multidimensional (de dos dimensiones en este caso).

El razonamiento anterior se puede extender a más de dos constructos. Si dos factores no reducen las correlaciones parciales a cero, entonces el constructo tiene más de dos dimensiones. En general, la dimensionalidad de un conjunto de elementos es igual al número de constructos o factores necesarios para reducir las correlaciones parciales a cero. Además, si se extraen múltiples conjuntos de elementos del dominio del constructo y se necesitan factores n para tener en cuenta las correlaciones entre los ítems, entonces se dice que la dimensionalidad del constructo es n .

Implicaciones de la unidimensionalidad

En la figura 3 la variable F_1 es la causa del efecto de las variables X_1 , X_2 , X_3 y X_4 , así como del efecto de las variables Z_1 y Z_2 . Aunado a lo anterior, la variable F_2 es la causa del efecto de las variables Z_1 , Z_2 y Z_3 .

FIGURA 3. Modelo de dos factores con relación parcial de las variables.



Fuente: Elaboración propia

TABLA 3. Matriz de correlación del modelo de la figura 3 .

	F_1	F_2	x_1	x_2	x_3	x_4	z_1	z_2	z_3
F_1	1.00								
F_2	0.00	1.00							
x_1	0.90	0.00	1.00						
x_2	0.80	0.00	0.72	1.00					
x_3	0.70	0.00	0.63	0.56	1.00				
x_4	0.90	0.00	0.81	0.72	0.63	1.00			
z_1	0.60	0.60	0.54	0.48	0.42	0.54	1.00		
z_2	0.60	0.60	0.54	0.48	0.42	0.54	0.72	1.00	
z_3	0.00	0.60	0.00	0.00	0.00	0.00	0.36	0.36	1.00

Fuente: SPSS IBM 21, con datos propios.

Observe que X_1 a X_4 son indicadores que sólo provienen a causa de F_1 ; Z_1 y Z_2 son indicadores provenientes a causa de F_1 y F_2 y Z_3 es un indicador proveniente sólo de F_2 . La tabla 4 muestra las correlaciones parciales entre varios conjuntos de variables pertenecientes a la figura 3. A continuación se presentan algunas de las observaciones que pueden extraerse de los resultados expuestos de las tablas 4a a 4e.

TABLA 4a. Matriz de correlación del modelo de la figura 3. Caso 1. Parcializando el efecto de F_1 desde X_1 a X_4

	x_1	x_2	x_3	x_4
x_1	1.00			
x_2	0.00	1.00		
x_3	0.00	0.00	1.00	
x_4	0.00	0.00	0.00	1.00

Fuente: SPSS IBM 21, con datos propios.

Caso 1. En el caso mostrado en la tabla 4a, el conjunto X_1 a X_4 de indicadores es unidimensional, ya que un factor común explica las correlaciones entre los ítems, y cada ítem es unidimensional, ya que mide uno y sólo un constructo.

Caso 2. En el ejemplo mostrado en las tablas 4b y 4c, el conjunto de elementos X_1 a X_4 y Z_1 es unidimensional, ya que las correlaciones entre los ítems pueden

TABLA 4b. Matriz de correlación del modelo de la figura 3. Caso 2: Parcializando el efecto de F_1 desde X_1 a X_4 y Z_1 .

	x_1	x_2	x_3	x_4	z_1
x_1	1.00				
x_2	0.00	1.00			
x_3	0.00	0.00	1.00		
x_4	0.00	0.00	0.00	1.00	
z_1	0.00	0.00	0.00	0.00	1.00

Fuente: SPSS IBM 21, con datos propios.

TABLA 4c. Matriz de correlación del modelo de la figura 3. Caso 2: Parcializando el efecto de F_1 desde X_1 a X_4 y Z_2 .

	x_1	x_2	x_3	x_4	z_2
x_1	1.00				
x_2	0.00	1.00			
x_3	0.00	0.00	1.00		
x_4	0.00	0.00	0.00	1.00	
z_2	0.00	0.00	0.00	0.00	1.00

Fuente: SPSS IBM 21, con datos propios.

ser contabilizadas por un solo factor. El indicador Z_1 no es unidimensional, sin embargo, es una medida de dos factores. Es decir, es posible que un conjunto de indicadores sea unidimensional, sin embargo, cada elemento del conjunto puede o no ser unidimensional. Del mismo modo, el conjunto de elementos X_1 a X_4 y Z_2 es unidimensional; sin embargo, el indicador Z_2 no es unidimensional.

Caso 3. En las tablas 4d y 4e, el conjunto de elementos X_1 a X_4 , Z_1 y Z_2 no es unidimensional, ya que la correlación entre Z_1 y Z_2 no puede ser explicada por un factor. Las correlaciones entre este conjunto de indicadores se pueden explicar por dos factores. Por lo tanto, este conjunto de elementos es multidimensional “en este caso, de dos dimensiones). Además, los indicadores Z_1 y Z_2 son multidimensionales y no unidimensionales. El lector puede ver fácilmente que los ítems X_1 a X_4 y Z_1 - Z_3 no son unidimensionales, ya que las correlaciones parciales entre ellos no pueden ser contabilizadas por un solo factor.

TABLA 4d. Matriz de correlación del modelo de la figura 3. Caso 3: Parcializando el efecto de F_1 desde X_1 a X_4 y Z_1 y Z_2 .

	x_1	x_2	x_3	x_4	z_1	z_2
x_1	1.00					
x_2	0.00	1.00				
x_3	0.00	0.00	1.00			
x_4	0.00	0.00	0.00	1.00		
z_1	0.00	0.00	0.00	0.00	1.00	
z_2	0.00	0.00	0.00	0.00	0.225	1.00

Fuente: SPSS IBM 21, con datos propios.

TABLA 4e. Matriz de correlación del modelo de la figura 2.3. Caso 3: Parcializando el efecto de F_1 y F_2 desde X_1 a X_4 y Z_1 y Z_2 .

	x_1	x_2	x_3	x_4	z_1	z_2
x_1	1.00					
x_2	0.00	1.00				
x_3	0.00	0.00	1.00			
x_4	0.00	0.00	0.00	1.00		
z_1	0.00	0.00	0.00	0.00	1.00	
z_2	0.00	0.00	0.00	0.00	0.00	1.00

Fuente: SPSS IBM 21, con datos propios.

Con base en las observaciones anteriores, se puede concluir que, si la definición conceptual del constructo permanece igual, entonces el constructo F_1 no es unidimensional, ya que todos los conjuntos de elementos del dominio del constructo no son unidimensionales. Normalmente, cuando esto sucede, se suele decir que el constructo depende del contexto, ya que su dimensionalidad y quizás el significado conceptual podría cambiar dependiendo de qué conjunto de indicadores se utilizan para medir el constructo.

Importancia de la unidimensionalidad

Una revisión de la literatura existente utilizando el análisis factorial confirmatorio (CFA. *Confirmatory Factor Analysis*), el modelo de estructura de covarianza o

modelo de ecuaciones estructurales (SEM. *Structural Equations Modeling*) sugiere fuertemente la preferencia por el uso de indicadores que sean unidimensionales y, sólo en el contexto de un estudio de investigación específico, y sólo en un factor (Gerbing y Anderson, 1988; Neuberg, West, Thompson y Judice, 1997).

A pesar de que no ha habido una investigación sistemática de por qué estos indicadores deben ser preferidos, a primera vista, la fuerte preferencia por estos elementos parece históricamente fundamentada en la práctica de utilizar la suma de una puntuación de los indicadores en escalas específicas en ANOVA (análisis de la varianza) y análisis de regresión en lugar de indicadores individuales. De hecho, incluso cuando se usan indicadores individuales en modelos de estructura de covarianza (por ejemplo, el SEM), la tendencia es reportar estadísticas tales como correlaciones corregidas de indicadores totales (*corrected ítem-to-total correlations*) y coeficientes alfa, que son características de, o implican, puntuaciones sumadas. Si el investigador utiliza puntajes compuestos en un modelo de estructura de covarianza, entonces, desde un punto de vista conceptual, es deseable que los indicadores que forman la puntuación sumada sean unidimensionales (Floyd y Widman, 1995; Neuberg et al., 1997).

Este es el argumento más convincente a favor de usar elementos que sean unidimensionales. Si, por otra parte, el conjunto de elementos es unidimensional, pero los elementos individuales no son unidimensionales, entonces el uso de las puntuaciones sumadas no es apropiado, ya que las puntuaciones sumadas también contendrían el efecto de otros factores sobre los cuales se cargan los ítems. En estas situaciones, los investigadores podrían utilizar las puntuaciones de los factores en lugar de las puntuaciones sumadas como insumo para un análisis posterior, a pesar de los problemas conocidos con la indeterminación y la estimación de los puntajes de los factores (Sharma, 1996). Las puntuaciones de los factores se forman como combinaciones lineales de los indicadores de cada factor usando los coeficientes de regresión de la puntuación del factor como pesos. Este enfoque tiene la ventaja de que los elementos no necesitan cargas distintas de cero sólo en un factor latente.

Cabe señalar, sin embargo, que las relaciones estructurales con base en las puntuaciones de los factores aún pueden atenuarse, porque las puntuaciones de los factores contienen error. Sin embargo, este enfoque puede ser menos problemático que el uso de compuestos basados en las puntuaciones sumadas.

Evaluando la dimensionalidad de los constructos

El coeficiente alfa (Cronbach, 1951) es la medida más popular que se informa para evaluar la consistencia interna de las escalas y, en muchos casos, un alto coeficiente

alfa se considera evidencia de la unidimensionalidad del constructo. Desafortunadamente, esto no es verdad. Por ejemplo, la estimación del coeficiente alfa para el conjunto de elementos X_1 a X_4 y Z_1 a Z_3 de la figura 3 es igual a .852, y se podría concluir que el conjunto de elementos es unidimensional. Ahora bien, se acepta que el coeficiente alfa es significativo sólo para un conjunto unidimensional de elementos (Clark y Watson, 1995; Cortina, 1993). Para un conjunto de elementos que no son unidimensionales, existe confusión en cuanto a lo que es exactamente evaluado por el coeficiente alfa. Es muy importante garantizar la relación entre consistencia interna y dimensionalidad, así como conocer y aplicar procedimientos para evaluar la dimensionalidad de los constructos. Como se mencionó anteriormente, la dimensionalidad se define como el número de factores comunes o construcciones latentes o subyacentes necesarias para explicar la correlación entre las variables. Por lo tanto, el análisis de factores es un método apropiado y popular para evaluar la dimensionalidad de las construcciones. Podría utilizarse tanto el análisis factorial exploratorio (EFA. *Exploratory Factor Analysis*) como el análisis factorial confirmatorio o ambos para evaluar la dimensionalidad de los constructos.

Análisis factorial exploratorio

El supuesto implícito que subyace en el uso del análisis factorial exploratorio (EFA. *Exploratory Factor Analysis*) es que el investigador generalmente tiene una idea limitada con respecto a la dimensionalidad de los constructos y qué elementos pertenecen o cargan sobre qué factor. Además, la EFA se lleva a cabo típicamente durante la etapa inicial del desarrollo de la escala. Aún así, el EFA puede usarse para obtener información sobre la potencial dimensionalidad de los elementos y escalas (Mejía-Trejo, 2017c, capítulo 12: Análisis factorial). Por ejemplo, suponga la aplicación del EFA a los datos de la tabla 3, se obtiene la tabla 5

El número de factores que explican las correlaciones entre las variables representa la dimensionalidad de un conjunto de variables. Para determinar el número de factores se utilizan una serie de reglas empíricas o heurísticas, tales como los criterios:

- a. El eigenvalor con valor >1 .
- b. El gráfico de autovalor para el criterio de contraste de caída (Mejía-Trejo, 2017c: 22).
- c. El gráfico de autovalor para el criterio de contraste de caída con análisis paralelo.

TABLA 5. Salida parcial de SPSS para los datos de la tabla 3.

<i>Initial Eigenvalues</i>			
<i>Factor</i>	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>
1	3.961	56.580	56.580
2	1.337	19.094	75.674
3	.500	7.141	82.815
4	.437	6.249	89.064
5	.296	4.222	93.286
6	.280	4.000	97.286
7	.190	2.714	100.000

Extraction Method: Principal Axis Factoring.

Fuente: SPSS IBM 21, con datos propios.

Así, es posible afirmar:

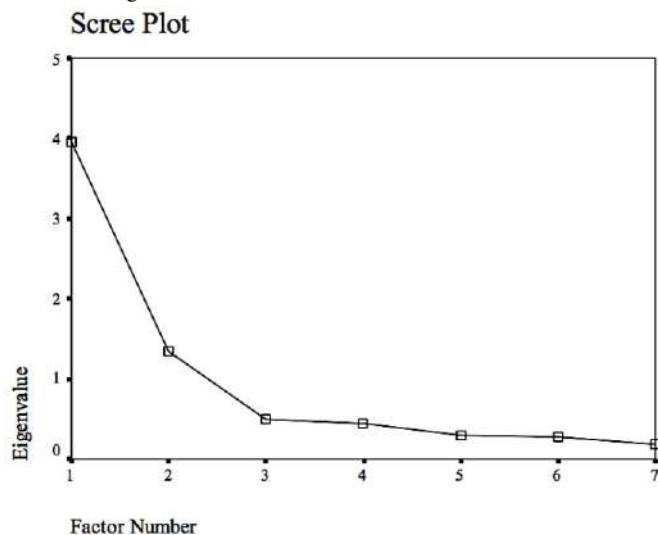
- ▶ De acuerdo con el criterio a, el número de factores es igual al número de autovalores mayores a 1. Un factor dado debe dar cuenta de por lo menos tanta variación como puede ser explicado por un solo elemento o variable. En este caso, el inciso a sugiere la presencia de dos factores, por lo que concluye que el conjunto de elementos no es unidimensional. Cabe destacar que este criterio ha tenido críticas de uso muy interesantes, sobre la utilidad de identificar el número de factores (Cliff, 1988).
- ▶ En cuanto al criterio b propuesto por Cattell (1966), es otra técnica popular. Es un diagrama de valores propios en función del número de factores, y uno busca un codo que significa una fuerte caída en la varianza explicada por un factor dado. Se supone que los factores en o más allá del codo son factores que ya no aportan más valor y tienden a representar errores o componentes únicos (gráfico 1).

Por ejemplo, en la gráfica 1 el codo sugiere la presencia de dos factores.

El conjunto de elementos es multidimensional, sin embargo, en muchos casos, no es posible identificar completamente el codo (gráfico 2).

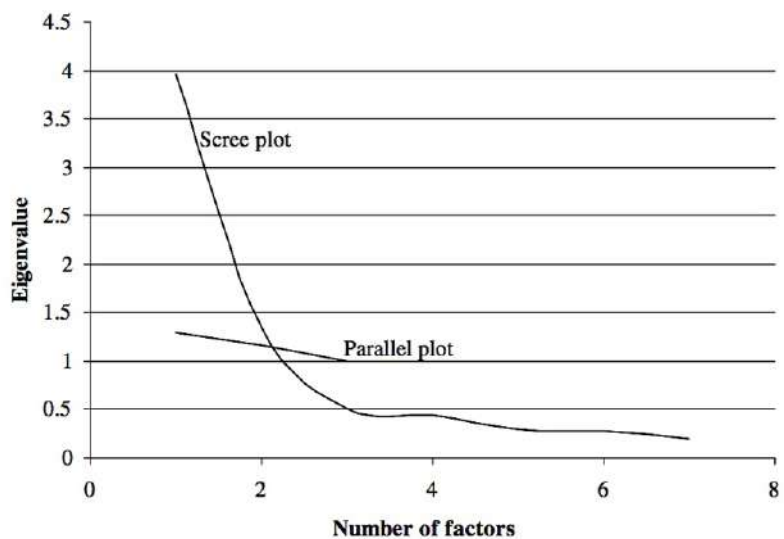
En tales casos, se puede utilizar el procedimiento de trazado paralelo sugerido por Horn (1965). Con este procedimiento, el gráfico en paralelo representa los valores propios que resultarían si el conjunto de datos no tuviera factores comunes; es decir, las correlaciones entre las variables se deben completamente al error de muestreo. Se requieren simulaciones extensas para estimar los autovalores para el gráfico paralelo; sin embargo, basadas en ecuaciones empíricamente derivadas,

GRAFICO 1. El gráfico de autovalor para el criterio de contraste de caída, con codo declarado (scree test) o gráfico de sedimentación.



Fuente: SPSS IBM 21, con datos propios.

GRÁFICO 2. Gráfico de sedimentación (de autovalor para el criterio de contraste de caída, sin codo declarado).



Fuente: SPSS IBM 21, con datos propios.

Allen y Hubbard (1986) han desarrollado la ecuación 3 para estimar los valores propios y obtener el gráfico paralelo:

ECUACIÓN 3. Estimación de valores para obtener el gráfico paralelo.

$$\ln \lambda_k = a_k + b_k \ln (n - 1) + c_k \ln \{(p - k - 1) / (p - k + 2)/2\} + d_k \ln \lambda_{k-1}$$

donde:

λ_k : es la estimación del késimo valor propio,

p : es el número de variables o ítems,

n : es el número de observaciones,

a_k, b_k, c_k y d_k : son coeficientes de regresión, y

$\ln \lambda_0$: es considerado como 1.

Fuente: Allen y Hubbard (1986).

La tabla 6 muestra los coeficientes empíricamente derivados para la ecuación anterior.

TABLA 6. Coeficientes de regresión por el método de componentes principales.

<i>Root (k)</i>	<i>Number of Points^a</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>R²</i>
1	62	.9794	-.2059	.1226	0.0000	.931
2	62	-.3781	.0461	.0040	1.0578	.998
3	62	-.3306	.0424	.0003	1.0805	.998
4	55	-.2795	.0364	-.0003	1.0714	.998
5	55	-.2670	.0360	-.0024	1.0899	.998
6	55	-.2632	.0368	-.0040	1.1039	.998
7	55	-.2580	.0360	-.0039	1.1173	.998
8	55	-.2544	.0373	-.0064	1.1421	.998
9	48	-.2111	.0329	-.0079	1.1229	.998
10	48	-.1964	.0310	-.0083	1.1320	.998
11	48	-.1858	.0288	-.0073	1.1284	.999
12	48	-.1701	.0276	-.0090	1.1534	.998
13	48	-.1697	.0266	-.0075	1.1632	.998
14	41	-.1226	.0229	-.0113	1.1462	.999
15	41	-.1005	.0212	-.0133	1.1668	.999
16	41	-.1079	.0193	-.0088	1.1374	.999

<i>Root (k)</i>	<i>Number of Points^a</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>R²</i>
17	41	-.0866	.0177	-.0110	1.1718	.999
18	41	-.0743	.0139	-.0081	1.1571	.999
19	34	-.0910	.0152	-.0056	1.0934	.999
20	34	-.0879	.0145	-.0051	1.1005	.999
21	34	-.0666	.0118	-.0056	1.1111	.999+
22	34	-.0865	.0124	-.0022	1.0990	.999+
23	34	-.0919	.0123	-.0009	1.0831	.999+
24	29	-.0838	.0116	-.0016	1.0835	.999+
25	28	-.0392	.0083	-.0053	1.1109	.999+
26	28	-.0338	.0065	-.0039	1.1091	.999+
27	28	.0057	.0015	-.0049	1.1276	.999+
28	28	.0017	.0011	-.0034	1.1185	.999+
29	22	-.0214	.0048	-.0041	1.0915	.999+
30	22	-.0364	.0063	-.0030	1.0875	.999+
31	22	-.0041	.0022	-.0033	1.0991	.999+
32	22	.0598	-.0067	-.0032	1.1307	.999+
33	21	.0534	-.0062	-.0023	1.1238	.999+
34	16	.0301	-.0032	-.0027	1.0978	.999+
35	16	.0071	.0009	-.0038	1.0895	.999+
36	16	.0521	-.0052	-.0030	1.1095	.999+
37	16	.0824	-.0105	-.0014	1.1209	.999+
38	16	.1865	-.0235	-.0033	1.1567	.999+
39	10	.0075	.0009	-.0039	1.0773	.999+
40	10	.0050	-.0021	.0025	1.0802	.999+
41	10	.0695	-.0087	-.0016	1.0978	.999+

Fuente: Datos tomados de la tabla 1 de Regression Equations for the Latent Roots of Random Data Correlation Matrices With Unities on the Diagonal, Multivariate Behavioral Research, 21, pp. 393-398 citado en Netemeyer et al. (2003).

a. The number of points used in the regression.

Usando la ecuación 3, el primer valor propio es igual a:

$$\ln \lambda_1 = .974 - .2059 \ln(200-1) + .1226 \ln[(7-1-1)(7-1+2)/2] + 0.00x1 = .257$$

$$\lambda_1 = e^{.257} = 1.293$$

y el segundo valor propio es igual a

$$\lambda_2 = -.3781 + .0461 \ln(200-1) + .004 \ln[(7-2-1)(7-2+2)/2] + .0578x.257 = .148$$

$$\lambda_2 = e^{.148} = 1.160$$

El procedimiento anterior se repite para calcular los autovalores restantes (de ser negativos, se consideran cero). El gráfico 2 muestra el gráfico paralelo. La trama sugiere dos factores; es decir, la dimensionalidad del conjunto de elementos es dos. El número de factores identificados por estos procedimientos debe ser confirmado adicionalmente, asegurando que el número de factores es suficiente para explicar todas las correlaciones entre las variables, ya que este es un objetivo clave del análisis de factores. Es decir, el investigador siempre está atento a determinar el ajuste del modelo de factor a los datos. Dos medidas de uso común son la raíz cuadrada media residual (RMSR. *Root-Mean-Square Residual*) y la raíz cuadrada media de correlaciones parciales entre las variables (RMSP. *root-mean-square of partial correlations among the variables*). Éstos se calculan a partir de la matriz de correlación residual (diferencia entre la matriz de correlación de muestras y la matriz de correlación calculada o estimada usando cargas factoriales estimadas) y la matriz de correlación parcial, antes definida. La RMSR y la RMSP pueden calcularse como lo indicado en las ecuaciones 4.

ECUACIÓN 4. Cálculo de la RMSR y la RMSP.

$$\text{RMSR} = \sqrt{\frac{2 \sum_{i=1}^k \sum_{j=i}^k \text{res}_{ij}^2}{k(k-1)}}$$

$$\text{RMSP} = \sqrt{\frac{2 \sum_{i=1}^k \sum_{j=i}^k \text{pc}_{ij}^2}{k(k-1)}}$$

donde res_{ij} y pc_{ij} son, respectivamente, la correlación residual y la correlación parcial entre las variables i y j .

Las tablas 7 y 8 muestran la matriz de correlación residual proporcionada por SPSS para los modelos de uno y dos factores.

En la ecuación 4, la RMSR para un modelo de dos factores es .00021, y de un factor es .215. Aunque no hay pautas sobre cuán bajo es bajo, hay una diferencia sustancial entre la RMSR para los modelos de uno y dos factores; por lo tanto, el modelo de dos factores proporciona la mejor explicación de las correlaciones entre las variables.

(Nota: SPSS no proporciona la RMSP. SAS, por otra parte, proporciona ambas matrices y también calcula los dos índices).

TABLA 7. Salida SPSS parcial para solución de un factor .

Reproduced Correlations										
	X1	X2	X3	X4	Z1	Z2	Z3			
Reproduced Correlation	X1	.754 ^b	.674	.592	.754	.606	.606	.123		
	X2	.674	.602 ^b	.529	.674	.541	.541	.110		
	X3	.592	.529	.464 ^b	.592	.475	.475	9.646E-02		
	X4	.754	.674	.592	.754 ^b	.606	.606	.123		
	Z1	.606	.541	.475	.606	.487 ^b	.487	9.880E-02		
	Z2	.606	.541	.475	.606	.487	.487 ^b	9.880E-02		
	Z3	.123	.110	9.646E-02	.123	9.880E-02	9.880E-02	2.005E-02 ^b		
Residual ^a	X1	4.613E-02	4.613E-02	3.836E-02	5.574E-02	-6.605E-02	-6.605E-02	-.123		
	X2	4.613E-02	3.142E-02	3.142E-02	4.613E-02	-6.145E-02	-6.145E-02	-.110		
	X3	3.836E-02	3.142E-02	3.836E-02	3.836E-02	-5.538E-02	-5.538E-02	-9.646E-02		
	X4	5.574E-02	4.613E-02	3.836E-02	5.574E-02	-6.605E-02	-6.605E-02	-.123		
	Z1	-6.605E-02	-6.145E-02	-5.538E-02	-6.605E-02	.233	.233	.261		
	Z2	-6.605E-02	-6.145E-02	-5.538E-02	-6.605E-02	.233	.233	.261		
	Z3	-.123	-.110	-9.646E-02	-.123	.261	.261	.261		

Extraction Method: Principal Axis Factoring.

a. Residuals are computed between observed and reproduced correlations. There are 16 (76.0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

Fuente: SPSS IBM 21, con datos Netemeyer *et al.* (2003).

TABLA 8. Salida SPSS parcial para soluciones de dos factores.

		Reproduced Correlation						
		X1	X2	X3	X4	Z1	Z2	Z3
Reproduced Correlation	X1	.810 ^b	.720	.630	.810	.540	.540	1.893E-04
	X2	.720	.640 ^b	.560	.720	.480	.480	3.751E-05
	X3	.630	.560	.490 ^b	.630	.420	.420	5.103E-05
	X4	.810	.720	.630	.810 ^b	.540	.540	1.893E-04
	Z1	.540	.480	.420	.540	.720 ^b	.720	.360
	Z2	.540	.480	.420	.540	.720	.720 ^b	.360
	Z3	1.893E-04	3.751E-05	5.103E-05	1.893E-04	.360	.360	.359 ^b
Residual ^a	X1		-5.284E-05	-5.296E-06	2.717E-04	9.458E-05	9.458E-05	-1.893E-04
	X2	-5.284E-05		-2.337E-04	-5.284E-05	1.876E-05	1.876E-05	-3.751E-05
	X3	-5.296E-06	-2.337E-04		-5.296E-06	2.548E-05	2.548E-05	-5.103E-05
	X4	2.717E-04	-5.284E-05	-5.296E-06		9.458E-05	9.458E-05	-1.893E-04
	Z1	9.458E-05	1.876E-05	2.548E-05	9.458E-05		2.049E-04	3.388E-04
	Z2	9.458E-05	1.876E-05	2.548E-05	9.458E-05	2.049E-04		3.388E-04
	Z3	-1.893E-04	-3.751E-05	-5.103E-05	-1.893E-04	3.388E-04	3.388E-04	

Extraction Method: Principal Axis Factoring.

a. Residuals are computed between observed and reproduced correlations. There are 0 (.0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

Fuente: SPSS IBM 21, con datos Netemeyer *et al.* (2003).

Análisis factorial confirmatorio

Aunque el análisis factorial exploratorio (EFA) da una idea de la dimensionalidad, el CFA (Análisis Factorial Confirmatorio / *Confirmatory Factor Analysis*) como su nombre implica, se centra esencialmente en si un hipotético modelo por factores se ajusta o no a los datos. El CFA es al momento, un método comúnmente aceptado para probar/ confirmar la dimensionalidad. En este método, se especifican a priori el número de factores, la estructura del factor (es decir, qué elementos cargan sobre qué factores) y la relación entre factores (o sea, si los factores están correlacionados).

Por ejemplo, al considerar el modelo de factores en la figura 3, se plantea la hipótesis de que el conjunto de ítems mide dos factores y que estos factores no están correlacionados. Además, se supone que los ítems X_1 a X_4 son medidas del primer factor F_1 , los ítems Z_1 y Z_2 miden los factores F_1 y F_2 , y Z_3 miden F_2 . El objetivo entonces es determinar si los datos apoyan el modelo hipotético o no. Esencialmente, las hipótesis nula y alternativa son:

H_0 : “El modelo se ajusta a los datos”.

H_1 : “El modelo no se ajusta a los datos”.

Obviamente, se desea no rechazar la hipótesis nula. Estas hipótesis anteriores pueden probarse mediante la estimación de los parámetros del modelo usando la estimación de máxima verosimilitud (MLE. *Maximum Likelihood Estimation*) o una técnica relacionada a través del CFA, y empleando la estadística de bondad de ajuste del chi-cuadrado resultante. Una serie de paquetes de software, como el SAS, LISREL, EQS, Y AMOS, se puede utilizar para realizar el CFA. La tabla 9 muestra la salida de PROC CALIS en el SAS.

El estadístico de la prueba Chi cuadrado es cero con 12 grados de libertad y no es significativo en $p < 0.05$, lo que implica que la hipótesis nula no puede ser rechazada, esto es, el modelo se encuentra suficientemente soportado por los datos. (Cabe señalar que el ajuste del modelo es perfecto, ya que se utilizaron datos hipotéticos para el modelo de la figura 3. Normalmente, para datos reales y tamaños de muestra grandes, el ajuste no será perfecto y la prueba estadística de chi-cuadrado será grande y significativa).

Por el contrario, la estadística de la prueba Chi-cuadrado para el modelo de un factor en el que todos los indicadores cargan en el factor único es 126.25 con 14 grados de libertad y es estadísticamente significativa en $p < 0.05$. Es decir, el modelo de un factor no está soportado por los datos. Las cargas factoriales estimadas del

TABLA 9. Salida parcial de resultados del SAS sobre los datos de la tabla 3.

Goodness-of-Fit Results

Fit Function	0.0000
Goodness of Fit Index (GFI)	1.0000
GFI Adjusted for Degrees of Freedom (AGFI)	1.0000
Root Mean Square Residual (RMR)	0.0000
Parsimonious GFI (Mulaik, 1989)	0.5714
Chi-Square	0.0000
Chi-Square DF	12
Pr > Chi-Square	1.0000
Independence Model Chi-Square	797.28
Independence Model Chi-Square DF	21
RMSEA Estimate	0.0000
RMSEA 90% Lower Confidence Limit	.
RMSEA 90% Upper Confidence Limit	.
ECVI Estimate	0.1675
ECVI 90% Lower Confidence Limit	.
ECVI 90% Upper Confidence Limit	.
Probability of Close Fit	1.0000
Bentler's Comparative Fit Index	1.0000
Normal Theory Reweighted LS Chi-Square	0.0000
Akaike's Information Criterion	-24.0000
Bozdogan's (1987) CAIC	-75.5798
Schwarz's Bayesian Criterion	-63.5798
McDonald's (1989) Centrality	1.0305
Bentler & Bonett's (1980) Non-normed Index	1.0271
Bentler & Bonett's (1980) NFI	1.0000
James, Mulaik, & Brett (1982) Parsimonious NFI	0.5714
Z-Test of Wilson & Hilferty (1931)	-7.2124
Bollen (1986) Normed Index Rho1	1.0000
Bollen (1988) Non-normed Index Delta2	1.0153
Hoelter's (1983) Critical N	.

Fuente: EQS 6.2, con datos propios..

modelo de dos factores son, dentro del error de redondeo, los mismos que los reportados en la figura 3.

Hay una serie de cuestiones relacionadas para el uso de la prueba chi-cuadrado estadística para la prueba de hipótesis, el más importante es su sensibilidad al tamaño de la muestra (Bearden, Sharma y Teel, 1982; Hoyle, 1995;

McDonald y Marsh, 1990). Por esta razón, los investigadores han propuesto una serie de pruebas estadísticas de bondad de ajuste alternativo que no son sensibles al tamaño de la muestra ya otros parámetros del modelo.

Conclusión

El concepto de unidimensionalidad fue discutido en este capítulo. Los siguientes puntos resumen los conceptos presentados:

1. Un conjunto de elementos es unidimensional si las correlaciones entre los ítems después del efecto de un solo factor son parcializados en cero. Es decir, un conjunto de elementos es unidimensional si un modelo de factor único se ajusta a los datos.
2. La dimensionalidad de un conjunto de elementos es el número de factores comunes necesarios para reducir las correlaciones parciales entre los elementos a cero.
3. Un ítem dado es unidimensional si mide una y sólo un constructo. La dimensionalidad de un elemento es igual al número de factores que mide.
4. Es posible que el conjunto de elementos de un conjunto no sea unidimensional.
5. Un constructo es unidimensional si cada conjunto de n ítems tomados aleatoriamente del dominio del constructo es unidimensional.
6. Los análisis factoriales exploratorios y confirmatorios son dos métodos para evaluar la dimensionalidad de los constructos.



CAPÍTULO 3

Confiabilidad

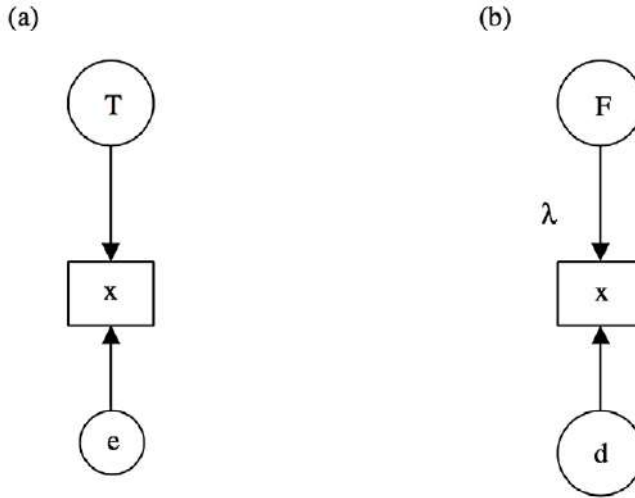
Hemos discutido problemas relacionados con la dimensionalidad de los constructos latentes y sus elementos de medición. Llegado a la conclusión de que el conocimiento de la dimensionalidad de un constructo es fundamental para el desarrollo de los indicadores que miden dicho constructo. Después de que los indicadores se han desarrollado y acotado, el siguiente paso importante a seguir, es examinar la confiabilidad y la validez de la escala. El propósito de este capítulo es discutir el concepto de confiabilidad y proporcionar procedimientos sugeridos para evaluarlo. La validez de una escala y los procedimientos para evaluarla se abordarán en el próximo capítulo.

Si un conjunto de indicadores de una escala está midiendo una misma variable o constructo latente, cabrá esperar que sus puntuaciones estén fuertemente correlacionadas entre sí, es decir, que sean internamente consistentes. Es muy importante señalar que la confiabilidad de una escala indica solamente que los distintos ítems que la componen, al estar muy correlacionados entre sí, están midiendo la misma variable o constructo latente. Pero que una escala sea confiable no quiere decir que dicha variable o constructo latente que está midiendo, sea la que tiene que medir, es decir, que sea válida. Por ejemplo, supongamos que en un examen de innovación tecnológica hemos diseñado un test de 50 preguntas y que por error el día del examen repartimos otro test diseñado para el examen de innovación de procesos. Este test será probablemente confiable, en cuanto que todas las preguntas obtendrán puntuaciones muy correlacionadas entre sí, dado que están midiendo lo mismo: el conocimiento de la industria por las prácticas de innovación de procesos. Lo que ocurre es que no son estos conocimientos que queríamos medir, sino los de innovación tecnológica. La escala diseñada es fiable, pero no es válida. Sin embargo, una escala no podrá ser válida, si no es confiable, de tal forma que la confiabilidad se convierte en una condición necesaria, aunque no suficiente, de la validez.

La importancia de medir la confiabilidad

La confiabilidad de las mediciones está estrechamente asociada con el concepto del modelo de puntuación verdadera (true-score model, Coker y Algina, 1986) de la teoría de la prueba clásica (figura 1a).

FIGURA 1. Modelo de 1 factor con relación total de las variables.



Fuente: Elaboración propia

La cual describe a X como la medida observada de un constructo latente. Según el modelo de puntuación verdadera, la puntuación observada, X , puede dividirse en verdadera puntuación y error (es decir, error de medición), (ecuación 1).

ECUACIÓN 1. Puntuación verdadera

$$X = T + e$$

donde:

X : es la puntuación observada,

T : es la puntuación verdadera, y

e : es el error de medición aleatorio.

Cuando la puntuación verdadera, T , varía, también lo hace la puntuación observada, X , porque el constructo latente influye en la puntuación observada (es

decir, X es un indicador de efecto). Suponiendo que la puntuación verdadera y el error no están correlacionados, es decir:

$$\text{Cov}(T, e) = 0$$

Puede demostrarse que la varianza de X es igual a la varianza de la puntuación verdadera más la varianza del error. Es decir:

$$\text{Var}(X) = \text{Var}(T) + \text{Var}(e)$$

Así, la confiabilidad de una medición se define como la relación de la varianza de puntuación verdadera relacionada con la varianza de puntuación observada (ecuación 2).

ECUACIÓN 2. Confiabilidad basada en la varianza.

$$\text{Reliability} = \rho_{xx} = \frac{\text{Var}(T)}{\text{Var}(x)} = \frac{\sigma_T^2}{\sigma_x^2}$$

Como se muestra en la figura 1b, en la medición de un modelo de un factor, la relación entre un elemento y el constructo latente está dada por la ecuación 3.

ECUACIÓN 3. Relación entre un elemento y el constructo latente.

$$X = \lambda F + \delta$$

donde

X: es la puntuación observada,

F: es el factor latente,

δ : es el error y

λ : carga factorial, es la medida en que el factor latente F afecta a la puntuación observada X.

Una vez más, como F varía, también lo hace la puntuación observada; sin embargo, el grado en que varía está determinado por el valor de λ . Está claro que el término λF es equivalente a la puntuación verdadera en el modelo (*true-score model*) y δ representa el error (en sentido estricto, se podría argumentar que conceptualmente los modelos de factor y el de verdadera puntuación no son equivalentes. En el modelo de factor, δ además del error de medición también incluye el error único o específico. Sin embargo, empíricamente, es imposible separar los dos. En consecuencia, a nivel empírico no es posible diferenciar entre la puntuación verdadera y el factor modelo). (ecuación 4).

ECUACIÓN 4. Confiabilidad de la medición observada.

$$\text{Reliability} = \frac{\text{Var}(T)}{\text{Var}(x)} = \frac{\lambda^2 \text{Var}(F)}{\text{Var}(x)} = \frac{\lambda^2 \sigma_F^2}{\sigma_x^2}.$$

Veamos que los numeradores en las cuaciones 2 y 4 son equivalentes en el sentido de que miden la varianza debido al efecto del constructo latente en la puntuación observada.

Las técnicas de confiabilidad más utilizadas son: test-retest, alternativa y de consistencia interna.

Confiabilidad test-retest

Imaginemos que una agencia de mercadotecnia digital, está interesado en medir la propensión de sus clientes potenciales a suscribirse al diseño de una página web, denominada (CP) con la siguiente afirmación X_1 : “Suscribirme a la página web me hace sentir bien”. También, considere que pide a diez sujetos que indiquen su grado de aceptación utilizando una escala de tipo Likert de 7 puntos de acuerdo-desacuerdo y que lo hagan en dos momentos: hoy y dos semanas después. Los datos hipotéticos resultantes se presentan en la tabla 1.

TABLA 1. Datos hipotéticos para la prueba de confiabilidad test-retest.

Sujeto	Ocasión 1	Ocasión 2
1	4	5
2	4	5
3	6	5
4	5	3
5	5	4
6	3	3
7	2	3
8	5	4
9	3	3
10	7	7

Fuente: Elaboración propia

Asumiendo que la propensión a la suscripción de la página web, por parte de los encuestados no cambia durante el periodo de dos semanas y no hay error de medición, entonces la correlación entre las medidas tomadas con dos semanas de diferencia hipotéticamente debería ser igual a uno. Un valor de menos de uno para

el coeficiente de correlación puede atribuirse al error de medición. La confiabilidad de la medida, pues, viene dada por la correlación de las medidas tomadas en la ocasión 1 y la ocasión 2, que para los presentes datos hipotéticos es igual a 0.74. Esta medida se conoce como confiabilidad test-retest.

Tome en cuenta que la fiabilidad test-retest se refiere a la estabilidad de las respuestas a los indicadores a través del tiempo. Un test-retest o coeficiente de estabilidad generalmente se estima por la magnitud de la correlación entre las mismas medidas (de la muestra) en diferentes ocasiones de evaluación. Si el coeficiente de estabilidad es bajo en magnitud, sin cambios en el constructo a lo largo del tiempo, la confiabilidad de la medición está en duda.

Si el coeficiente de estabilidad es alto en magnitud, sin cambio en el constructo a lo largo del tiempo, se mejora la confiabilidad de la medida. La razón para la confiabilidad test-retest es que si una medición refleja verdaderamente el constructo pretendido, debe ser capaz de evaluar el constructo en diferentes ocasiones. La puntuación verdadera en el constructo latente debe ser reflejada por el constructo en dos ocasiones de manera comparable. Por lo tanto, la confiabilidad test-retest es útil porque ofrece información sobre el grado de confianza que se tiene de que la medida refleje el constructo y es generalizable a otras ocasiones de evaluación (Haynes et al., 1999). En resumen, la correlación test-retest teóricamente representa el grado en que la constructo latente determina las puntuaciones observadas en el tiempo (DeVellis, 1991; Nunnally y Bernstein, 1994).

Existen algunas desventajas clave asociadas con la confiabilidad test-retest que limitan su utilidad como un coeficiente de confiabilidad teórica:

1. ¿Cuánto tiempo debe transcurrir entre las administraciones? No hay respuestas claras, pero para los constructos basados en la opinión (actitudinal), se ha propugnado un periodo de un mínimo de dos semanas (Robinson *et al.*, 1991).
2. En relación con la cantidad de tiempo entre administraciones, si se obtiene un bajo coeficiente test-retest, ¿significa esto que la escala no es confiable, o implica que lo que la escala ha cambiado con el tiempo? Si este último es el caso, entonces se ha violado un principio básico de la teoría de la prueba clásica y el coeficiente test-retest no sería una medición adecuada de confiabilidad (Crocker y Algina, 1986). Es decir, debe considerarse la naturaleza del constructo y la posibilidad de que cambie con el tiempo. Es posible que la posición de un encuestado sobre el constructo de interés (su puntuación verdadera) haya cambiado con el tiempo.

Por otra parte, si el intervalo de tiempo entre las administraciones es

demasiado corto, la puntuación del entrevistado del tiempo 1 al tiempo 2 puede reflejar simplemente la memoria con una apariencia en consistencia. Por lo tanto, el periodo debe ser lo suficientemente largo como para permitir que los efectos de un sesgo de respuesta de memoria se desvanezcan (un efecto de prueba de ser permanentemente consistente desde el tiempo 1 al tiempo 2), pero no tanto como para reflejar cambios históricos en la puntuación verdadera. Para otras deficiencias asociadas con el coeficiente test-retest, consulte a Kelley y McGrath (1988) y Nunnally y Bernstein (1994).

Confiabilidad alternativa

Considere que la siguiente afirmación, preferiblemente desarrollada independientemente de la primera, también está disponible para medir el constructo de la propensión a la suscripción de una página web (CP) con la afirmación X2: “Me gusta navegar en distintas opciones de páginas web para informarme”. Asuma que en la ocasión 1, la declaración X1 se utiliza, y en la ocasión 2, la declaración X2 es la que se emplea. La tabla 2 presenta los datos para las dos formas alternativas (es decir, las declaraciones) para medir el constructo CP.

TABLA 2. Datos hipotéticos para la prueba de confiabilidad alternativa.

Sujeto	Ocasión 1 (Declaración X1)	Ocasión 2 (Declaración X2)
1	4	4
2	4	4
3	6	5
4	5	3
5	5	4
6	3	5
7	2	4
8	5	4
9	3	5
10	7	7

Fuente: Elaboración propia

La correlación entre las dos respuestas (ocasión 1 y ocasión 2), que en el presente caso es igual a .41, se denomina confiabilidad de forma alternativa. Se deben anotar las similitudes y diferencias entre el test-retest y la confiabilidad de la

forma alternativa. En ambos casos, la confiabilidad se estima correlacionando dos conjuntos de respuestas tomadas en dos ocasiones.

En el caso test-retest, los dos conjuntos de respuestas son con respecto a la misma declaración o medida. Por otro lado, en la confiabilidad de la forma alternativa, los dos conjuntos de respuestas se refieren a dos declaraciones diferentes desarrolladas para medir el constructo. Los problemas relevantes para la confiabilidad de la forma alternativa son los mismos que para la confiabilidad prueba-retest.

Confiabilidad por consistencia interna

Debido a limitaciones como el tiempo, el costo y la disponibilidad de los mismos sujetos de estudio en múltiples ocasiones, no siempre es posible tomar medidas repetidas o usar formas alternativas. En tales casos, el concepto de consistencia interna se puede utilizar para estimar la confiabilidad. Este concepto requiere sólo una sola administración de los indicadores a los encuestados; sin embargo, asume la disponibilidad de múltiples medidas o elementos para medir una constructo dado. Antes de discutir la medida más utilizada de consistencia interna (coeficiente alfa), se presenta una breve discusión de la confiabilidad de la división en mitades, que es una forma de confiabilidad de consistencia interna.

Confiabilidad por consistencia interna: división en mitades

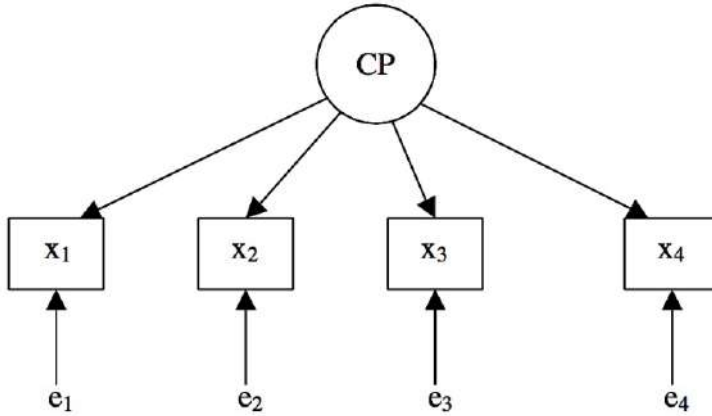
Suponga que existen cuatro indicadores que se utilizan para medir el constructo CP. Vea figura 2, el cual se basa en cuatro afirmaciones:

- X_1 : “Suscribirme a la página web me hace sentir bien”.
- X_2 : “Me gusta navegar en distintas opciones de páginas web para informarme”.
- X_3 : “Cuando logro la suscripción a la página web, siento que he realizado un gran trato”.
- X_4 : “Disfruto cuando me suscribo a una página web, sin importar el tiempo utilizado”.

Se pide a diez sujetos que indiquen su grado de acuerdo o desacuerdo sobre las declaraciones anteriores usando una escala de tipo Likert de 7 puntos (tabla 3).

Ahora suponga, que tomamos las dos primeras y dos últimas declaraciones y calculamos la puntuación total para cada tema, que se muestra en las dos últimas columnas de la tabla 3. La correlación entre las dos puntuaciones totales es igual a .77 y se denomina confiabilidad por división de mitades. Es obvio que pueden

FIGURA 2. Modelo de 1 factor .



Fuente: Elaboración propia

TABLA 2. Datos hipotéticos para la prueba de confiabilidad alternativa.

Sujeto	Declaraciones				Puntuación total	
	X1	X2	X3	X4	X1- X2	X3- X4
1	4	4	5	5	8	10
2	4	4	5	5	8	10
3	6	5	6	5	11	11
4	5	3	5	3	8	8
5	5	4	3	4	9	7
6	3	5	5	3	8	8
7	2	4	5	3	6	8
8	5	4	3	4	9	7
9	3	5	5	3	8	8
10	7	7	7	7	14	14

Fuente: Elaboración propia

formarse numerosas mitades divididas y la correlación para cada división será diferente (el número de divisiones que puede formarse es igual a $(2n')! / 2(n')^2$ donde $n' = n/2$ y n es el número de enunciados).

Confiabilidad por consistencia interna: coeficiente de alfa (α)

Cronbach (1951) ha demostrado que el promedio de la correlación de todas las divisiones posibles es igual al coeficiente de alfa.

El coeficiente alfa o alfa de Cronbach (α) se puede calcular utilizando la ecuación 4.

ECUACIÓN 5. Confiabilidad de la medición observada.

$$\alpha = \frac{k}{k-1} \left(\frac{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \text{Cov}(x_i x_j)}{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \text{Cov}(x_i x_j) + \sum_{i=1}^k \text{Var}(x_i)} \right)$$

$$= \frac{k}{k-1} \left(\frac{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij}}{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} + \sum_{i=1}^k \sigma_i^2} \right).$$

donde:

α : es el coeficiente alfa,

X_i : es medida para el indicador i , y

k : es el número de ítems o indicadores.

Debido a que α es la medida de confiabilidad más utilizada, se analiza su relación con la dimensionalidad, la longitud de la escala, las correlaciones entre índices y la redundancia de indicadores.

El coeficiente alfa se refiere al grado de interrelación entre un conjunto de indicadores diseñados para medir un solo constructo, con una varianza común entre los ítems. Para un conjunto de indicadores que componen una escala, la varianza en ese conjunto está compuesta por la varianza verdadera (la varianza entre individuos en el constructo que mide la escala) y error de varianza (toda otra varianza no contabilizada por la variable varianza verdadera, es decir, varianza no compartida o varianza única). El cálculo de α es un medio para dividir la varianza de puntuación total en sus componentes: verdadero y de error. En términos muy simples, *1 - error de varianza = α* , y *1 - α = error de varianza*. Por lo tanto, α representa la proporción de la varianza total de una escala que es atribuible a una fuente común, siendo esa fuente común la puntuación verdadera del constructo latente que se está midiendo.

En la figura 2, la varianza en cada ítem o indicador (es decir, X_i) que se debe a la variable latente CP se considera varianza común o compartida. Cuando CP varía,

también lo hacen las puntuaciones de los indicadores individuales, porque el constructo latente influye en las puntuaciones de los indicadores. Por lo tanto, las calificaciones de todos los indicadores varían conjuntamente con el constructo latente CP, lo que implica teóricamente que todos los indicadores están correlacionados. Los términos de error (e_i) representados en la figura 2 se consideran únicos para cada indicador. Es decir, representan una varianza que no es atribuible al constructo latente CP, y de acuerdo con la teoría de la prueba clásica, los términos de error no están correlacionados. Tanto la puntuación individual como la puntuación global de toda la escala, varían como funciones de dos fuentes:

- a. La fuente de variación común a sí misma (la puntuación global) y otros indicadores y
- b. La variación no compartida o única asociada con ese indicador en particular. Por lo tanto, la varianza total de la escala y la varianza para cada indicador son las que son atribuibles a fuentes: comunes y únicas (error). Así, alfa es conceptualmente equivalente a la relación entre la variación de la fuente común y la variación total (Cortina, 1993; DeVellis, 1991; Nunnally y Bernstein, 1994).

Por otro lado, y a considerar, tenemos la matriz de covarianza de la escala de CP para 4 ítems, la cual está dada por C, de la matriz 1.

MATRIZ 1. Varianzas-covarianzas.

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Donde el arreglo de C representa una matriz de varianzas-covarianzas, la cual es una generalización de una matriz de correlaciones, sólo que sus elementos no están estandarizados. Los elementos σ_i^2 de la diagonal expresan la varianza de cada ítem i de la escala. Los elementos σ_{ij} de fuera de la diagonal indican la covarianza entre dos ítems i y j de la escala. Así, se ha construido la escala para medir el constructo X (por ejemplo, administración del conocimiento del consumidor) y por ello asumiremos que X está bien representado por la suma de los k ítems de la escala que llamaremos Y. Las matrices de varianzas-covarianzas tienen un conjunto de propiedades muy útiles. Si sumamos todos los elementos de la matriz C tenemos la varianza total de la escala Y.

Coefficiente alfa. Explicación 1

Supongamos que formamos la puntuación total suma, TS, como:

$$TS = X_1 + X_2 + X_3 + X_4$$

La varianza de TS, es:

$$\begin{aligned} \text{Var (TS)} &= s_{TS}^2 \\ &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_{12} + \sigma_{13} + \sigma_{14} + \sigma_{21} + \sigma_{23} + \sigma_{24} + \sigma_{31} + \sigma_{32} \\ &\quad + \sigma_{34} + \sigma_{41} + \sigma_{42} + \sigma_{43} \\ &= \sum_{i=1}^k \sigma_i^2 + \sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} \end{aligned}$$

Observe que la varianza de puntuación total suma, TS, es igual a la suma de todas las varianzas y covarianzas en la matriz de covarianza. Para particionar el total de la varianza en, varianza común (es decir, varianza de puntuación verdadera) y varianza única (es decir, variación debida al error de medición), debe considerarse lo siguiente:

- ▶ Los *indicadores en diagonal* representan esencialmente la covarianza de un indicador consigo mismo, es decir, la variabilidad en el puntaje de un indicador proveniente de una muestra dada de individuos o casos. Como tales, los indicadores diagonales son fuentes únicas de varianza y no de la varianza que es común o compartida entre los indicadores.
- ▶ Los *indicadores fuera de la diagonal* son covarianzas que representan la varianza que es común o compartida por cualquier par de elementos en la escala. Así, las entradas en la matriz de covarianza consisten en: varianza única (error) y varianza común (*compartida/conjunta*).
- ▶ Así, lo que es único, es representado a lo largo de la diagonal principal ($\sum \sigma_i^2$), lo que es común es representado por los indicadores fuera de la diagonal, y la varianza total (σ^2_{TS}) es igual a la suma TS de todas las entradas de la matriz. Como tal, la proporción de varianza única (*no común*) a la varianza total viene dada por:

$$\frac{\sum_{i=1}^k \sigma_i^2}{\sigma_{TS}^2}$$

Se deduce de la expresión anterior, la ecuación 5.

ECUACIÓN 5.

$$\begin{aligned}
 \text{Confiabilidad} &= 1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_{TS}^2} \\
 &= \frac{\sigma_{TS}^2 - \sum_{i=1}^k \sigma_i^2}{\sigma_{TS}^2} \\
 &= \frac{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij}}{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} + \sum_{i=1}^k \sigma_i^2}
 \end{aligned}$$

Proporción de la variación conjunta/común de alfa captura el concepto de variación común inherente a la definición del coeficiente de alfa, debe realizarse un ajuste para el número de ítems en una escala. El número total de indicadores en la matriz de covarianza es igual a k^2 , donde k es el número de indicadores en la escala. Por lo tanto, el denominador de la ecuación 5 es la suma de los indicadores k^2 . Del mismo modo, el numerador está en la suma de $k^2 - k$ indicadores. Para ajustar la confiabilidad dada por la ecuación 5, el numerador se divide por $k^2 - k$ y el denominador por k^2 , como se aprecia en la ecuación 6.

ECUACIÓN 6.

$$\begin{aligned}
 &\frac{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} / (k^2 - k)}{\left(\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} + \sum_{i=1}^k \sigma_i^2 \right) / k^2} \\
 &= \frac{k}{k-1} \left(\frac{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij}}{\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} + \sum_{i=1}^k \sigma_i^2} \right)
 \end{aligned}$$

Ajuste de la proporción de la variación conjunta/común de alfa el cual, que es el mismo que el coeficiente α dado por la ecuación 4.

Teóricamente, α varía de 0 a 1, por lo que son altamente deseables valores altos de α . También existe una fórmula de indicadores estandarizada análoga a la

ecuación 6 que utiliza una matriz de correlación con 1s en la diagonal principal (es decir, la correlación de un indicador consigo misma o desviaciones estandarizadas) y correlaciones (r) entre pares de indicadores como elementos fuera de la diagonal (es decir, *covarianzas estandarizadas*). Así:

- ▶ El indicador estandarizado α se considera como una evaluación apropiada de consistencia interna si se suman las puntuaciones estandarizadas de indicadores para formar una escala.
- ▶ Sin embargo, el indicador estandarizado α no se considera apropiado, cuando el puntaje bruto de los indicadores son utilizados para sumar una escala, porque las diferencias en las variaciones de los indicadores pueden afectar la puntuación total (Cortina, 1993), (ecuación 7).

ECUACIÓN 7. Elemento estandarizado alfa.

$$\frac{kr}{1 + (k - 1)r}$$

donde:

k : es el número de elementos en la escala, y

r : es la correlación promedio entre los indicadores de la escala.

Finalmente, para los indicadores que se califican dicotómicamente, se usa la fórmula de Kuder-Richardson 20 (KR-20) para calcular el coeficiente alfa. Esta fórmula, es idéntica de la ecuación 6 con la excepción de que la expresión $\sum pq$ sustituye a la expresión $\sum \sigma_i^2$. La expresión $\sum pq$ especifica que la varianza de cada indicador se calcula, y luego estas varianzas se suman para todos los indicadores, donde:

p : representa cada media de indicadores, y

q : es 1-media del indicador.

Así que la varianza de un indicador se convierte en la expresión pq

Para saber más, vea: Crocker and Algina, 1986, p. 139-140, en la aplicación de la fórmula KR-20. Vea la Ecuación 3.8

ECUACIÓN 8. Cálculo de coeficiente de alfa con la fórmula KR20 (indicadores dicotómicos)

$$\alpha = \frac{k}{k - 1} \left(1 - \frac{\sum pq}{\sigma_x^2} \right)$$

Ejemplo de caso basado en explicación 1

Para demostrar el cálculo del coeficiente alfa (ecuación 6) y su forma estandarizada de ítem, partiremos del ejemplo sobre la agencia de mercadotecnia digital, utilizando la calificación hipotética de cinco indicadores medidos en escala de Likert de 7 puntos siguiente:

- X1: "Suscribirme a la página web me hace sentir bien".
- X2: "Me gusta navegar en distintas opciones de páginas web para informarme".
- X3: "Cuando logro la suscripción a la página web, siento que he realizado un gran trato".
- X4: "Disfruto cuando me suscribo a una página web, sin importar el tiempo utilizado".
- X5: "Cuando me suscribo a una página web, considero que estoy ganando dinero".

Las matrices de covarianza y de correlación para estos datos, se reportan en las tablas 4a y 4b.

TABLA 4a. Matriz de covarianza para los cinco indicadores de la escala CP .

	<i>CP1</i>	<i>CP2</i>	<i>CP3</i>	<i>CP4</i>	<i>CP5</i>
<i>CP1</i>	3.6457				
<i>CP2</i>	2.7831	4.3877			
<i>CP3</i>	2.3934	2.2280	2.9864		
<i>CP4</i>	2.0706	2.3973	1.9109	3.7786	
<i>CP5</i>	1.8338	2.0782	1.6011	1.8813	3.3138

Fuente: SPSS IBM 21, con datos propios

TABLA 4a. Matriz de correlación para cinco indicadores de la escala CP .

	<i>CP1</i>	<i>CP2</i>	<i>CP3</i>	<i>CP4</i>	<i>CP5</i>
<i>CP1</i>	1.0000				
<i>CP2</i>	.6969	1.0000			
<i>CP3</i>	.7254	.6155	1.0000		
<i>CP4</i>	.5579	.5888	.5688	1.0000	
<i>CP5</i>	.5276	.5450	.5089	.5317	1.0000

Fuente: SPSS IBM 21, con datos propios

Usando la fórmula de la ecuación 6, se tiene:

$$\alpha = \frac{5}{5 - 1} \left(\frac{42.356}{42.356 + 18.112} \right) = .876$$

donde:

$$k = 5,$$

$$\sum \sigma_i^2 = 3.6457 + 4.3877 + 2.9864 + 3.7786 + 3.3138 = 18.112$$

$$\sum_{i=1}^k \sum_{\substack{j=1 \\ i \neq j}}^k \sigma_{ij} = 2(2.7831 + 2.3934 + 2.2280 + 2.0706 + 2.3973 + 1.9109 + 1.8338 + 2.0782 + 1.6011 + 1.8813) = 42.356.$$

La suma de los indicadores fuera de la diagonal, fue multiplicada por dos para mostrar ambas covarianzas, tanto abajo como arriba de la diagonal. Al repetir el procedimiento mencionado para la matriz de correlación y con el uso de la ecuación 7, obtenemos como resultado .876 para alfa. Observe que no existe diferencia entre ambos valores de alfa. Esto será generalmente el caso; sin embargo, si las varianzas de los indicadores son muy diferentes, entonces los 2 valores pueden no ser los mismos. El cálculo (con $r = .585$), es como sigue:

$$\frac{kr}{1 + (k - 1)r} = \frac{5 \times .5865}{1 + (5 - 1) \times .5865} = .876.$$

Coficiente alfa. Explicación 2

Pues bien, el coeficiente alpha de Cronbach, se define como la proporción de la varianza total de la escala que es atribuible a la variable latente X. Cuanto mayor sea este valor, querrá decir que X está mejor representado por Y en la escala porque está causando (explicando) la mayor parte de la varianza de esta (varianza común). La parte de la varianza total que no explica la variable o constructo latente es la causada por los errores de medición de cada indicador y se denomina varianza específica. Hay que hacer notar que, dados los supuestos que asumimos en el apartado anterior, cada término de error, provoca sólo varianza en cada ítem por separado y esos errores no están correlacionados unos con otros. En resumen, cada indicador (y por suma, el conjunto de la escala) sólo varía como función de:

- a. La fuente de variación que supone la variable o constructo latente (varianza común).
- b. La fuente de variación que provoca el error (varianza específica).

Veamos cómo nos ayuda la matriz de varianzas-covarianzas a recoger esta información. Como hemos dicho, la varianza total de la escala (σ_y^2) es la suma de todos los elementos de esa matriz. Asimismo, la suma de los elementos de la diagonal, nos proporcionará la suma de la varianza de los ítems individuales ($\sum \sigma_i^2$). Pero como las covarianzas recogen variación entre pares de ítems, y la varianza específica procede de cada ítem² por separado, la varianza específica ha de venir recogida necesariamente por $\sum \sigma_i^2$. Se había definido el coeficiente alfa como la parte de la varianza total que era explicada por la variable o constructo latente (varianza común). Como la varianza total es la suma de la varianza común y la varianza específica, la varianza común será la varianza total menos la varianza específica, luego podremos escribir:

$$\alpha = \frac{\sigma_y^2 - \sum \sigma_i^2}{\sigma_y^2} = 1 - \frac{\sum \sigma_i^2}{\sigma_y^2}$$

En la expresión anterior, sin embargo, hace falta introducir una última corrección para tener la expresión del alfa de Cronbach. El número total de elementos de la matriz de varianzas covarianzas es k^2 . El número de elementos de la matriz que son específicos (la diagonal) es k , mientras que los elementos comunes (fuera de la diagonal) son $k^2 - k$.

Tenemos, en la expresión anterior, una fracción con un numerador basado en k valores y un denominador con base en k^2 valores. Para ajustar los cálculos de tal forma que el ratio exprese las magnitudes relativas, más que el número de términos que hay en numerador y denominador, corregiremos la expresión anterior para contrarrestar el efecto de la diferencia por $k^2 / (k^2 - k)$, o lo que es lo mismo, por $k / (k - 1)$, de forma que ahora, a estará acotado entre 0 y 1 (ecuación 9).

ECUACIÓN 9. Alfa de Cronbach por varianza-covarianza la cual, corresponde a una versión reducida de la ecuación 6.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_y^2} \right)$$

Para resumir, la confiabilidad medida por la expresión anterior, nos indica qué proporción de la varianza total está provocada por la variable o constructo latente, corrigiendo este valor por el número de casos que intervienen en los cálculos. Si se desea trabajar en términos de correlaciones entre los ítems de la escala, en lugar de en términos de varianzas-covarianzas, la expresión anterior puede adaptarse de una manera sencilla. La suma de las varianzas individuales de cada ítem $\sum \sigma_i^2$ puede ponerse como el producto entre el número de ítems (k) y la media de las varianzas de los mismos v (por ejemplo, 10 ítems que suman 50 y 10 veces su media, que es 5, dan el mismo valor). Por lo que respecta al denominador, si llamamos c a la media de las covarianzas, como hay k varianzas y $k^2 - k$ covarianzas, podremos poner la expresión anterior del alfa de Cronbach como sigue:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{k v}{k v + (k^2 - k) c} \right)$$

Sustituyendo el 1 por su equivalente $[k v + (k^2 - k) c] / [k v + (k^2 - k) c]$ y operando, se llega a:

$$\frac{k}{k-1} \frac{k v + (k^2 - k) c}{k v + (k^2 - k) c}$$

Y simplificando a:

$$\alpha = \frac{k c}{v + (k - 1) c}$$

Dado que se está buscando una expresión que utilice correlaciones en lugar de varianzas y covarianzas. Si estandarizamos las variables implicadas en la expresión anterior, la media de las covarianzas c se convierte en la media de las correlaciones (ρ) y la media de las varianzas es igual a 1, por lo que el alfa de Cronbach se calcula con la ecuación 9.

ECUACIÓN 10. Alfa de Cronbach por correlación.

$$\alpha = \frac{k \rho}{1 + (k - 1) \rho}$$

donde:

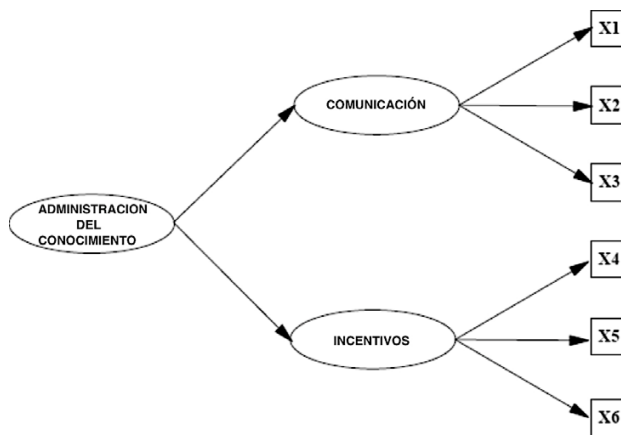
ρ es, como ya hemos indicado, la media de los coeficientes de correlación entre todos los ítems que conforman la escala.

A esta fórmula se le conoce como la fórmula de Spearman-Brown (Crocker y Algina, 1986). ¿Cuáles son los valores del alfa de Cronbach por debajo de los cuales no se considera como confiable a una escala? Siguiendo a Nunally y Bernstein (1994), es posible afirmar que este nivel depende de para qué vaya a utilizarse la escala. En etapas preliminares de desarrollo de una escala, un nivel de .7 puede ser suficiente y, tras las depuraciones oportunas de la escala, este valor no debe bajar nunca de .8. Si en función de los valores de la escala se van a tomar decisiones que afecten a los individuos (asignar empleados a capacitación diversa, según los resultados de su test de inteligencia, por ejemplo), el alfa no podrá ser inferior a .9 (Véase más adelante discusión al respecto de las alfas a utilizar).

Ejemplo de caso basado en explicación 2

Expuesto el desarrollo conceptual del análisis de confiabilidad mediante el alfa de Cronbach, veamos un ejemplo de aplicación por parte del constructo propuesto de administración del conocimiento (OECD, 2003) que resolveremos mediante SPSS. Suponga que se está realizando una investigación que pretende medir el grado de administración del conocimiento que se logra a partir de los incentivos y la comunicación con los empleados de una organización (figura 3).

FIGURA 3. Diagrama de trayectoria constructo administración del conocimiento .



Fuente: OECD (2003).

Así como la descripción el constructo e indicadores propuestos en la tabla 5.

FIGURA 3. Constructo con indicadores de la administración del conocimiento

Abajo le señalamos una serie de actividades que se suponen se tienen dentro de su firma y que se realizan o no con cierta regularidad. Para cada una de ellas, le solicitamos que indique con qué frecuencia se han realizado en los últimos seis meses. Tome en cuenta que los números significan: 1= Nunca; 2=Casi Nunca; 3= Rara Vez; 4= Ocasionalmente; 5= Con Alguna frecuencia; 6= Frecuentemente; 7= Con mucha frecuencia.

	Comunicación	1	2	3	4	5	6	7
X1	14. En su firma, los trabajadores comparten el conocimiento con documentación por escrito.							
X2	15. En su firma, los trabajadores comparten su conocimiento regularmente al actualizarse todas las bases de datos de sus productos							
X3	16. En su firma, los trabajadores comparten su conocimiento al realizar trabajo colaborativo en equipos virtuales.							
	Incentivos							
X4	17. En su firma, el conocimiento se comparte mediante recompensas con reconocimiento monetario e incentivos.							
X5	18. En su firma, el conocimiento se comparte mediante reconocimientos con incentivos no monetarios.							
X6	19. En su firma, se cuenta con un sistema que apoya el flujo del <i>Know-How</i> .							

Fuente: OECD (2003).

Supongamos que se ha pasado ese cuestionario a 50 gerentes de operación del sector electrónico de cierto país (empresas medianas y grandes) y que se quiere analizar la confiabilidad de la escala que mide la variable o constructo latente COMUNICACIÓN. Sin embargo, el investigador ha cometido un error y al calcular la confiabilidad ha incluido en la escala de COMUNICACIÓN el ítem X₄, que corresponde al constructo INCENTIVOS. A continuación, recogemos y comentamos la salida del SPSS donde se muestran los resultados del análisis de confiabilidad.

En primer lugar se ofrecen estadísticos descriptivos de la aplicación de la escala, con las medias, desviaciones típicas y el número de casos obtenidos por cada ítem en la aplicación a los 50 individuos. Ya puede apreciarse que la media del ítem X₄ es claramente distinta a las demás (tabla 6).

TABLA 6. Estadística descriptiva del caso.

	Media	Desviación típica	N
X1	5.18	1.395	50
X2	5.40	1.107	50
X3	5.52	1.216	50
X4	3.64	1.793	50

La siguiente información que se proporciona son las matrices de varianzas covarianzas y de correlación que, como se indicó al desarrollar el cálculo del alfa de Cronbach, sirven de base indistintamente para su cálculo. Si se observan las correlaciones del ítem X_4 con todos los demás, se observará que son muy bajas, incluso negativas (-.1 con X_1 ; .05 con X_2 y -.02 con X_3), avanzando ya que la escala, así construida va a tener problemas de confiabilidad (tablas 7a y 7b).

TABLA 7a. Matriz de covarianzas interítem del caso.

	X1	X2	X3	X4
X1	1.947			
X2	.763	1.224		
X3	1.211	.522	1.479	
X4	-.260	.106	-.054	3.215

Fuente: SPSS IBM 21, con datos propios.

TABLA 7a. Matriz de covarianzas interítem del caso.

	X1	X2	X3	X4
X1	1.000			
X2	.494	1.000		
X3	.713	.388	1.000	
X4	-.104	.053	-.025	1.000

Fuente: SPSS IBM 21, con datos propios.

La siguiente información a considerar, es posible que ayude a calcular manualmente el valor del alfa de Cronbach, tanto con la expresión tradicional de las varianzas-covarianzas, como mediante la expresión que implica correlaciones y valores estandarizados (*y que en el SPSS se denomina standardized ítem alpha*), (tablas 8a y 8b).

TABLA 8a. Estadísticos de resumen de los elementos .

	Media	Mínimo	Máximo	Rango	Máximo/mínimo	Varianza	N de elementos
Medias de los elementos	4.935	3.640	5.520	1.880	1.516	.765	4
Varianzas de los elementos	1.966	1.224	3.215	1.990	2.625	.782	4
Covarianzas inter-elementos	.381	-.260	1.211	1.471	-4.649	.278	4
Correlaciones inter-elementos	.253	-.104	.713	.818	-6.853	.097	4

Fuente: SPSS IBM 21, con datos propios.

TABLA 8b. Estadísticos de la escala .

Media	Varianza	Desviación típica	N de elementos
19.74	12.441	3.527	4

Fuente: SPSS IBM 21, con datos propios.

Mediante la ecuación 9, para calcular el alfa basta con conocer la varianza total de la escala, que aparece en el cuadro anterior bajo el título de varianza y que toma el valor 12.441 y la suma de las varianzas de los ítems que se pueden obtener de la diagonal de la matriz de varianzas-covarianzas que también hemos ofrecido:

$$\sum \sigma_i^2 = 1,9465 + 1,2245 + 1,4792 + 3,2147 = 7,8649$$

De tal forma que, sustituyendo en la ecuación 8, el alfa de Cronbach, es:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_y^2} \right) \quad \alpha = \frac{4}{3} \left(1 - \frac{7,8649}{12,4412} \right) = 0,4904$$

Si aplicamos la fórmula de Spearman-Brown para datos estandarizados que mostramos en la ecuación 9 la información a requerir proviene de la tabla 8a. Así, la media de los coeficientes de correlación aparece en la última fila de correlaciones interelementos bajo la columna media con el valor de .2535. Sustituyendo en la ecuación 0, se tiene:

$$\alpha = \frac{k\rho}{1+(k-1)\rho} \quad \alpha = \frac{4 \times 0,2535}{1+(4-1) \times 0,2535} = 0,5759$$

El ejercicio sólo es de comprobación ya que el SPSS, reporta los datos resultado de manera directa (tabla 9).

TABLA 9. Estadísticos de confiabilidad .

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
.490	.576	4

Fuente: SPSS IBM 21, con datos propios.

Hasta aquí, el investigador puede determinar sus propias conclusiones. Si se compara el valor del alfa de Cronbach (.4904) con el límite mínimo permitido de acuerdo con lo indicado en el desarrollo teórico (.7 si la escala está en fase de desarrollo o .8 si ya es una escala contrastada), se aprecia que la confiabilidad de la misma no es suficiente. El paso siguiente es analizar si la supresión de algún ítem

mejoraría estos resultados o si, por el contrario, no hay forma de lograr un nivel adecuado de confiabilidad. En nuestro caso, dado que hemos cometido el error de incorporar el ítem X4 que medía un constructo diferente, cabe esperar que suprimiéndolo de la escala la confiabilidad mejore. Véase la tabla 10 como medidas de sensibilidad a la eliminación de ítems o indicadores.

TABLA 10. Estadísticos total-elemento.

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
X1	14.56	7.068	.462	.575	.244
X2	14.34	8.433	.433	.257	.319
X3	14.22	7.604	.501	.513	.240
X4	16.10	9.643	-.037	.029	.777

Fuente: SPSS IBM 21, con datos propios.

La columna correlación elemento-total corregida nos muestra el coeficiente de correlación entre cada ítem y la suma de los restantes ítems que constituyen la escala. si uno de estos valores es mucho más bajo que los demás, esto ya aporta una evidencia de que probablemente no esté midiendo el mismo constructo. Como se observa, el valor correspondiente a X_4 , además de pequeño es negativo, con lo que es un firme candidato a ser eliminado de la escala. Esta conclusión queda reforzada por la información que proporciona la última columna del cuadro anterior alfa de Cronbach si se elimina el elemento que nos muestra qué valor tomaría el alfa de Cronbach si se rehiciesen los cálculos, pero eliminando cada uno de los ítems consecutivamente. Se comprueba que, si se eliminasen los ítems X_1 , X_2 o X_3 , el alfa sería mucho menor que el que hemos obtenido (.244; .319 y .240, respectivamente) pero, si eliminásemos X_4 , la confiabilidad sería muy superior, entrando en los límites aceptables (.777). Por lo tanto, es conveniente, eliminar el ítem X_4 en cuyo caso la escala para medir el constructo COMUNICACIÓN sería confiable. Por último, se muestra la salida correspondiente a SPSS que analiza la confiabilidad para los 3 primeros ítems. Obsérvese que el valor del alfa coincide con el pronosticado (.777).

Se propone, con los datos por último y como ejercicio, repetir el análisis de confiabilidad efectuado, pero ahora para la segunda variable latente (INCENTIVOS) a través de sus correspondientes ítems: X_4 , X_5 y X_6 . [Solución alfa de Cronbach = .7998; alpha mediante la fórmula de Spearman-Brown = .7992], (tabla 11).

TABLA 11. Secuencia de análisis de alfa de Cronbach de los 3 ítems del caso.

Resumen del procesamiento de los casos

		N	%
Casos	Válidos	50	100.0
	Excluidos ^a	0	.0
	Total	50	100.0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
.777	.773	3

Estadísticos de los elementos

	Media	Desviación típica	N
X1	5.18	1.395	50
X2	5.40	1.107	50
X3	5.52	1.216	50

Matriz de correlaciones inter-elementos

	X1	X2	X3
X1	1.000	.494	.713
X2	.494	1.000	.388
X3	.713	.388	1.000

Matriz de covarianzas inter-elementos

	X1	X2	X3
X1	1.947	.763	1.211
X2	.763	1.224	.522
X3	1.211	.522	1.479

Estadísticos de resumen de los elementos

	Media	Mínimo	Máximo	Rango	Máximo/mínimo	Varianza	N de elementos
Medias de los elementos	5.367	5.180	5.520	.340	1.066	.030	3
Varianzas de los elementos	1.550	1.224	1.947	.722	1.590	.134	3
Covarianzas inter-elementos	.832	.522	1.211	.688	2.317	.098	3
Correlaciones inter-elementos	.532	.388	.713	.325	1.838	.022	3

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
X1	10.92	3.749	.731	.565	.557
X2	10.70	5.847	.481	.247	.828
X3	10.58	4.698	.657	.511	.650

Estadísticos de la escala

Media	Varianza	Desviación típica	N de elementos
16.10	9.643	3.105	3

Fuente: SPSS IBM 21, con datos propios.

Coeficiente de alfa y dimensionalidad

Aunque el alfa está referido al grado de interrelación que existe entre un conjunto de indicadores diseñados para medir un solo constructo, existen varias descripciones adicionales del coeficiente alfa, siendo de las más relevantes, las mencionadas por Cortina (1993), tales como:

- a. α : es la media de todas las confiabilidades por mitades divididas,
- b. α : es la confiabilidad del límite inferior de una medición,
- c. α : es una medida de la saturación del primer factor,
- d. α : es igual a la confiabilidad bajo un supuesto equivalente de tau, y
- e. α : es una versión general del coeficiente KR-20 para indicadores dicotómicos.

Aunque estas descripciones han sido ampliamente utilizadas, la validez de algunas de ellas descansa en ciertas suposiciones (por ejemplo, el uso de α en su forma estandarizada o no normalizada, como se ha indicado).

Una conclusión que puede extraerse de las diversas descripciones de α en su relación con la dimensionalidad es la siguiente (Cortina, 1993: 100): “Es una función de la medida en que los ítems en una prueba tienen comunidades altas y, por tanto, una singularidad baja. También es una función de la interrelación, aunque hay que recordar que esto no implica una unidimensionalidad u homogeneidad”.

Por lo tanto, debe recordarse que es posible que un conjunto de indicadores estén interrelacionados pero no sean homogéneos. Como tal, el coeficiente α no es una medida de la unidimensionalidad y debe usarse para evaluar la consistencia interna sólo después de que se establezca la unidimensionalidad (Clark y Watson, 1995; Cortina, 1993; Hattie, 1985; Schmitt, 1996).

Como se indicó en el capítulo 2, la unidimensionalidad puede definirse como la existencia de un rasgo o constructo latente o subyacente de un conjunto de indicadores/medidas (Hattie, 1985). Este conjunto de indicadores también se ha denominado como medición congénica. La importancia de establecer la dimensionalidad antes de establecer otras propiedades psicométricas (por ejemplo, la consistencia interna y la validez nomológica) no debe subestimarse (Cortina, 1993; Hattie, 1985; Schmitt, 1996).

Para operacionalizar los constructos latentes, los investigadores usan a menudo puntuaciones compuestas, sumando o promediando indicadores diseñados para medir el constructo de interés. El cálculo y el uso de tales puntuaciones son significativos sólo si los indicadores tienen una unidimensionalidad aceptable.

El uso de escalas multidimensionales como si fueran unidimensionales (es decir, sumando o promediando compuestos de indicadores) puede resultar en ambigüedades interpretativas de las relaciones entre constructos en una prueba de teoría. Es decir, si un constructo es multidimensional, pero todas las puntuaciones de los ítems se suman/promedian entre las dimensiones en una sola puntuación compuesta y se correlacionan con una variable de criterio, tal correlación es ambigua. Neuberger *et al.* (1997) ofrecen una exposición interesante del costo de tratar una escala multidimensional como si fuera unidimensional a través de analogía experimental. Debe recordarse que un objetivo primario de la experimentación es crear una manipulación inconfundible de una variable independiente para evaluar con precisión su efecto sobre la variable dependiente. Si dos constructos están siendo manipulados por una experimentación diseñada para influir en un constructo, el efecto de un constructo sobre la variable dependiente no puede ser medido con precisión, ya que desentrañar su efecto de la variación no deseada del segundo constructo es problemático. Del mismo modo, los investigadores tienen el objetivo de desarrollar escalas que sí estén evaluando un constructo. La justificación detrás de la unidimensionalidad es que la “interpretación de cualquier medida –ya sea que represente un rasgo, un estado de ánimo, una capacidad o una necesidad– es más clara si sólo una dimensión subyace a la medida” (Neuberger *et al.*, 1997: 1022). Cuando sólo una dimensión subyace a una medida, la correlación de esa medición con un criterio, es más clara.

Cuando existe más de una dimensión, posiblemente sugiriendo que se está evaluando más de una variable rasgo/diferencia individual, la correlación de una medición con un criterio puede confundirse. En suma, el problema que surge de tratar una escala multidimensional como si fuera unidimensional (por ejemplo, sumando una puntuación total) es que más de una dimensión subyace a la puntuación total. Esto plantea las posibilidades de que cualquier efecto sobre una variable de criterio pueda atribuirse a la dimensión errónea o que todas las dimensiones sean necesarias para producir el efecto. Dado que el objetivo de las investigaciones es construir y probar teorías, se requiere un significado de constructo válido. Por lo

tanto, la unidimensionalidad es una condición necesaria para la consistencia interna, la validez del constructo y la prueba teórica.

Cabe señalar que hay trabajos de investigadores que han ofrecido técnicas para combinar múltiples dimensiones de un constructo que intentan hacer inconfundibles los efectos de combinar las dimensiones en una sola puntuación compuesta (Carver, 1989; Hull, Lehn y Tedlie, 1991). En general, estos enfoques intentan estimar el efecto de cada dimensión de un constructo sobre las variables de criterio a través de medios estadísticos (por ejemplo, regresión y enfoques de modelos de ecuación estructural). También hay que señalar que hay casos especiales para los cuales se puede justificar la suma de las puntuaciones a través de las dimensiones y la formación de una puntuación compuesta. Por ejemplo, si las dimensiones no tienen efectos diferenciales sobre las variables de criterio de interés, entonces puede ser necesario crear un compuesto global (sumando las puntuaciones a través de las dimensiones) (Carver, 1989; Richins y Dawson, 1992).

Coeficiente de alfa, longitud de escala, correlación y redundancia entre indicadores

También hay que recordar que α es una función de la longitud de la escala, el promedio de la correlación interítem (covarianza) y la redundancia de los indicadores. Así, consideramos:

1. *La longitud de la escala.* Las fórmulas para α expresadas en las ecuaciones 6 y 7 sugieren que a medida que aumenta el número de indicadores, α tenderá a aumentar. Debido a que la parsimonia también es una referencia en la medición, una pregunta importante es “¿qué es una longitud de escala apropiada en relación con el coeficiente α ?” La respuesta depende parcialmente del dominio y las dimensiones del constructo. Naturalmente, un constructo con un dominio de contenido amplio y múltiples dimensiones requerirá más elementos para aprovechar adecuadamente el dominio/dimensiones que un constructo con un dominio estrecho y una dimensión. Dado que la mayoría de las escalas se autoadministran se debe considerar la fatiga y/o la falta de cooperación de los encuestados, por lo que la brevedad de la escala es a menudo una preocupación (Clark y Watson, 1995; Cortina, 1993; DeVellis, 1991; Nunnally y Bernstein, 1994). Para resolverla, varios investigadores han sugerido el uso de la teoría de generalizabilidad en el diseño de escalas.
2. *La redundancia de los ítems y el promedio de la correlación interítem del coeficiente α .* Se recomienda ampliamente que la formulación de los indicadores sea simple y directa, que exploten el contenido del constructo y que el encuestado obtenga el significado deseado del investigador. También se recomienda que se utilicen varios elementos para aprovechar adecuadamente el dominio del constructo. Sin embargo, cuando el texto de los indicadores es demasiado similar, el coeficiente α (así como la validez del contenido y la dimensionalidad)

dad) puede ser artificialmente mejorado. Los indicadores que están redactados de forma demasiado similar aumentarán el promedio de la correlación interítem, que en efecto incrementa el coeficiente α , pero sin añadir sustancialmente a la validez del contenido de la medida (Boyle, 1991, Clark y Watson, 1995). Aunque es necesaria cierta similitud entre los indicadores de una escala para aprovechar el dominio, varios de éstos que son sólo variaciones ligeras de redacción de otros elementos son redundantes y contienen muy poca información nueva sobre el constructo (Clark y Watson, 1995). Los indicadores redundantes pueden contribuir a la paradoja de la atenuación (*paradox attenuation*) en la teoría psicométrica, por lo que el incremento del coeficiente α más allá de un cierto punto no aumenta la consistencia interna (Boyle, 1991; Loevinger, 1957).

Dada esta posibilidad, los investigadores deben tener cuidado en su interpretación de α , considerando su relación con el número de indicadores de una escala, el nivel de correlación interítem, redundancia de indicadores y dimensionalidad.

3. Como evidencia adicional de la interrelación de la longitud de la escala, el nivel promedio de la correlación interítem y la redundancia de la redacción de los indicadores (y la dimensionalidad), considere lo siguiente: un nivel ampliamente preconizado de adecuación para α ha sido .70, independientemente de los temas que acabamos de discutir. De nuevo, α nos da información sobre la medida en que cada indicador de un conjunto se correlaciona con otros ítems en ese conjunto, y a medida que aumenta el número de indicadores y el incremento promedio de la correlación interítems (*ceteris paribus*), α aumentará. Además, el número de indicadores en una escala puede tener un efecto pronunciado en los niveles más bajos de correlación entre indicadores. Por ejemplo, en su metanálisis, Peterson (1994) encontró que el nivel medio de α para una escala de 3 ítems con un promedio de correlación interítem de .47 fue de .73. Si el nivel de correlación entre indicadores de .47 se aplica a una escala de 9 ítems, esta escala mostraría un nivel α de .89. Sin embargo, para las escalas de 9 ítems revisadas, Peterson (1994) encontró un promedio correlación interítem de .31 y un nivel promedio de α de .80.

Cortina (1993) demostró además las relaciones entre el coeficiente α , el promedio de correlación interítem, la longitud de la escala y la dimensionalidad. Sus hallazgos se basan en escalas con 1, 2 o 3 dimensiones; 6, 12 o 18 indicadores; y el promedio de correlaciones interítem de .30, .50 y .70. Tres de sus hallazgos clave son los siguientes:

- a. El número de indicadores afecta en gran medida α . Si una escala tiene un gran número de indicadores, α puede estar por encima de .70 incluso con correlaciones interítem bajas. Además, para una medida unidimensional con un

- promedio de correlación interítem de .50, α será superior a .70 independientemente del número de indicadores en la medición.
- b. Para una escala que tenga más de 14 ítems, un α de .70 o más alto resultará aun si se combinan dos dimensiones ortogonales (no relacionadas) con un promedio de correlación interítem de .30. Si las dos dimensiones tienen el promedio de las correlaciones de indicadores que están por encima de .70, entonces α puede ser mayor que .85.
 - c. Dada una gran cantidad de elementos razonablemente correlacionados, una escala puede tener un α aceptable aunque contenga tres dimensiones ortogonales. En suma, α aumenta a medida que aumenta la correlación interítem, y disminuye como una función de la multidimensionalidad. Sin embargo, α todavía puede ser alto, a pesar de las correlaciones interítem bajas y la presencia de multidimensionalidad. Cuando se interpreta α , se debe tener una unidimensionalidad ya establecida, y luego considerar el número de indicadores y el promedio de la correlación interítem.

Sin embargo, maximizar el alfa es un objetivo encomiable en la construcción de la escala, un objetivo que debe ser mediado considerando la longitud de la escala, la el promedio de correlación interítem, la redundancia del texto del indicador y la dimensionalidad/complejidad de la escala. Aunque para nuestro conocimiento, no existen criterios estadísticos duros en cuanto a cuál es el número mínimo o máximo de elementos en una escala, cuál es un aceptable mínimo α , o cuál es un nivel aceptable del promedio de la correlación interítem, por lo que se aplican una o varias heurísticas existentes. Por ejemplo, Robinson et al. (1991) abogó por un nivel α de .80 o mejor, y el promedio de las correlaciones interítem de .30 o mejor, como ejemplo. Clark y Watson (1995) abogaron por un promedio de correlación interítem de .15 a .50 entre constructos, y para constructos estrechamente definidos, abogaban por un rango de .40 a .50. También abogaron por un nivel de coeficiente α de al menos .80 para una nueva escala. Una vez que se alcanza el punto de referencia .80, sin embargo, la adición de elementos es de poca utilidad para la consistencia interna y la validez del contenido, particularmente con un constructo estrechamente definido. Para tales constructos, 4 a 7 indicadores podrían bastar. Cabe decir que estas directrices y heurísticas representan sólidos consejos en la construcción de escala.

Índice de fiabilidad compuesta

Como se ha comprobado, el alfa de Cronbach permite analizar la confiabilidad de un conjunto de indicadores utilizados para medir un constructo dado. Pero, normalmente, en las investigaciones de las ciencias de la administración no aparece implicado un solo constructo, sino varios. El alfa de Cronbach de cada factor por separado NO tiene en cuenta la influencia sobre la confiabilidad del resto

de constructos. Por esta razón, Fornell y Larcker (1981) proponen el cálculo del índice de la fiabilidad compuesta (IFC) para cada factor que, interpretándose exactamente igual que el alfa de Cronbach, sí que tiene en cuenta las interrelaciones. Su fórmula, para el factor i , se aprecia en la ecuación 11.

ECUACIÓN 11. Índice de la fiabilidad compuesta

$$IFC_i = \frac{\left(\sum_j L_{ij} \right)^2}{\left(\sum_j L_{ij} \right)^2 + \sum_j Var(E_{ij})}$$

Fuente: Fornell y Larcker (1981).

donde:

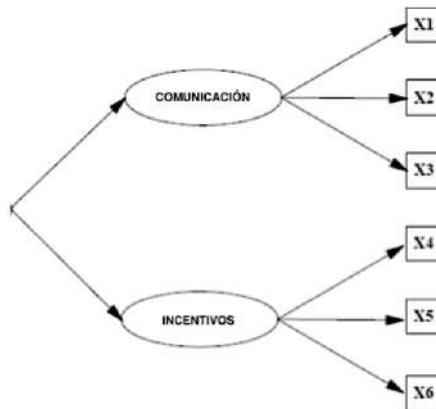
L_{ij} : es la carga factorial estandarizada de cada uno de los j indicadores que cargan sobre el factor i , y

$Var(E_{ij})$: es la varianza del término de error asociado a cada uno de los j indicadores del factor i .

Se puede calcular, también como: $Var(E_{ij}) = 1 - L_{ij}^2$

Para obtener esta información es necesario realizar el análisis factorial confirmatorio entre los instrumentos de medida. Tomando el ejemplo de la figura 3, el modelo de medida afecta a las variables latentes COMUNICACIÓN e INCENTIVOS, por lo que el análisis factorial confirmatorio a estimar sería el representado en la figura 4.

FIGURA 4. Diagrama de trayectorias para aplicar el CFA en el modelo de administración del conocimiento.



Fuente: OECD (2003)

La tabla 12 muestra la sintaxis de EQS que permite estimar este modelo y en la tabla 13 la parte de la salida necesaria para el cálculo de la confiabilidad compuesta que se obtiene ejecutando dicha sintaxis (que no es otra que los coeficientes de regresión estandarizados).

TABLA 12. Sintaxis EQS.

```

/TITLE
ADMINISTRACION DEL CONOCIMIENTO
/SPECIFICATIONS
VARIABLES= 6; CASES= 50; DATAFILE='E:\AFCS.ESS';
MATRIX=RAW; METHOD=ML; ANALYSIS= COV;

/EQUATIONS
V1=F1+E1;
V2=*F1+E2;
V3=*F1+E3;

V4=F2+E4;
V5=*F2+E5;
V6=*F2+E6;

/VARIANCES
E1=*;
E2=*;
E3=*;
E4=*;
E5=*;
E6=*;

F1 TO F2=*;

/COVARIANCES
F1, F2=*;

/LMTEST

/PRINT
FIT=ALL;

/END
    
```

Fuente: EQS 6.2, con datos propios.

TABLA 13. Información necesaria para calcular el IFC.

STANDARDIZED SOLUTION:				
X1	=V1	=	.963 F1	+ .270 E1
X2	=V2	=	.514*F1	+ .858 E2
X3	=V3	=	.741*F1	+ .671 E3
X4	=V4	=	.946 F2	+ .323 E4
X5	=V5	=	.656*F2	+ .755 E5
X6	=V6	=	.672*F2	+ .740 E6

Fuente: EQS 6.2, con datos propios.

TABLA 14. Recopilación de información que calcula el IFC.

Constructo e indicadores	Carga factorial estandarizada L_{ij}	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de fiabilidad compuesta $IFC > .7$ $\frac{\sum (L_{ij})^2}{[\sum (L_{ij})^2 + \sum Var(E_{ij})]}$
COMUNICACIÓN			
X ₁	.963	.073	.796
X ₂	.514	.736	
X ₃	.741	.451	
Total	2.218	1.259	
INCENTIVOS			
X ₄	.945	.107	.809
X ₅	.657	.568	
X ₆	.673	.547	
Total	2.275	1.222	

Fuente: EQS 6.2. con datos y adaptación propios.

Con la información de la tabla 14 y aplicando la ecuación 11 para el IFC, éste sería para ambos factores:

$$IFC_i = \frac{\left(\sum_j L_{ij}\right)^2}{\left(\sum_j L_{ij}\right)^2 + \sum_j Var(E_{ij})}$$

$$IFC_1 = \frac{\left(\sum_j L_{1j}\right)^2}{\left(\sum_j L_{1j}\right)^2 + \sum_j Var(E_{1j})} = \frac{(2,218)^2}{(2,218)^2 + 1,259} = 0,796$$

$$IFC_2 = \frac{(2,275)^2}{(2,275)^2 + 1,222} = 0,809$$

Se puede comprobar cómo ambos índices superan lo que en investigación debería ser entendido como un mínimo el valor de .7.

Índice de la varianza extraída

Fornell y Larcker (1981) presentan este índice IVE como la relación entre la varianza que es capturada por un factor en relación con la varianza total debida al error de medida de ese factor, esto es:

ECUACIÓN 12. Índice de varianza extraída.

$$IVE_i = \frac{\sum_j L_{ij}^2}{\sum_j L_{ij}^2 + \sum_j Var(E_{ij})}$$

Fuente: Fornell y Larcker (1981).

Nótese que la única diferencia con el IFC es que cada carga factorial estandarizada es primero elevada al cuadrado antes de ser sumadas. forner y larcker (1981) sugieren que es deseable que el constructo tenga valores del IVE iguales a .5 o superiores, es decir, que sea superior la varianza capturada por el factor que la debida al error de medida. este nivel mínimo suele ser muy conservador y es fácil encontrar en la literatura ejemplos de escalas aceptadas con ive inferiores. la tabla 13 ya nos ofrecía toda la información necesaria para obtener el IVE en conjunto con el recopilado de la información de la tabla 14. Los cálculos necesarios se reflejan en la tabla 15.

TABLA 15. Recopilación de información que calcula el IVE.

Constructo e indicadores	Cuadrado de la carga factorial estandarizada $(L_{ij})^2$	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de varianza compuesta $IVE \geq .5$ $\frac{\sum (L_{ij})^2}{[\sum (L_{ij})^2 + \sum Var(E_{ij})]}$
COMUNICACIÓN			
X ₁	.927	.073	.580
X ₂	.264	.736	
X ₃	.549	.451	
Total	1.741	1.259	
INCENTIVOS			
X ₄	.893	.107	.592
X ₅	.432	.568	
X ₆	.453	.547	
Total	1.778	1.222	

Fuente: EQS 6.2. con datos y adaptación propios.

Sustituyendo en la ecuación 12 del IVE:

$$IVE_1 = \frac{\sum_j L_{ij}^2}{\sum_j L_{ij}^2 + \sum_j Var(E_{ij})} = \frac{1,741}{1,741 + 1,259} = 0,580$$

$$IVE_2 = \frac{1,778}{1,778 + 1,222} = 0,592$$

Valores que, como se ha señalado, superan los mínimos recomendables.

Conceptos adicionales de la confiabilidad

Finalmente, en la tabla 16, se mencionan algunos para sus consideraciones.

TABLA 16. Conceptos de confiabilidad adicionales.

Confiabilidad	Enfoque de:	Cuando se aplica:	Algunos índices, son:
Relativa	Estabilidad	Una misma escala en dos ocasiones sucesivas	Coef. de correlación Coef. W. p.
	Equivalencia	En la misma muestra dos formas paralelas de una escala	Coef. de correlación entre dos formas paralelas
	Consistencia interna	La escala una sola vez	Spearman - Brown Rulon y Guttman Kuder - Richardson Host α de Cronbach
Absoluta	Variación interna	La escala una vez	Varianza
	Consistencia interna	La escala una vez y se mide el error de la prueba o cuando se aplica en dos pruebas diferentes	σ Dif med.

Fuente: Elaboración propia

Conclusión

Este capítulo ha discutido varias formas de confiabilidad, reSaltando las ventajas del coeficiente alfa de Cronbach. También discutió la relación entre el coeficiente alfa, la unidimensionalidad, la correlación entre los elementos, la duración de la escala y la redundancia de la redacción de los elementos. Destacamos la importancia de establecer la dimensionalidad antes de evaluar la consistencia interna, así como los efectos de la correlación entre las palabras y la redundancia de la redacción en la confiabilidad.

Como reconocimos previamente, la dimensionalidad y la confiabilidad son entradas necesarias para establecer la validez de una escala. El siguiente capítulo discute los diversos tipos de validez que se encuentran y los procedimientos para establecer estos tipos de validez en el desarrollo de la escala.

CAPÍTULO 4

Validez

¿Cuándo es válida una escala? Cuando lo que está midiendo realmente es la variable latente que se supone que tiene que medir. Esta definición puede sofisticarse mucho más, pero esa es la esencia de la validez. Siguiendo la definición de Bohrnstedt (1976) que apunta en esta misma línea: “validez es el grado en que un instrumento mide el concepto bajo estudio”. La validez, sin embargo, es un concepto multifacético y tiene diversas dimensiones que deben explicarse y analizarse por separado.

Este capítulo aborda cuestiones críticas relacionadas con la validez de constructo, tal como se aplican al desarrollo de la medición para constructos latentes o subyacentes. La validez de constructo se considera como el grado en el cual, una medición refleja verdaderamente el concepto que se está investigando o el grado en el cual las variables utilizadas para observar la covariación dentro y entre constructos pueden interpretarse en términos de constructos teóricos (Calder, Phillips y Tybout, 1982).

La validez de constructo puede y debe considerarse ampliamente como una referencia al grado en que las inferencias se pueden hacer legítimamente, provenientes de las mediciones de estudios de constructos teóricos en los que se basan esas operacionalizaciones. Similar a las opiniones expresadas por Haynes et al. (1999) y Trochim (2002), la validez de constructo representa la calidad global de un estudio de investigación o incluso un programa de estudios. En general, una medición tiene validez de constructo cuando refleja o muestra:

- a. Un grado de evaluación en magnitud y dirección de una muestra representativa de las características del constructo, y
- b. Un grado en que la medición no está contaminada con elementos del dominio de otros constructos o errores (Peter, 1981).

Es importante señalar que la validez de constructo de una medición no se evalúa directamente sino que se deduce de la evidencia en que las puntuaciones sustantivas de la medida, se comportan como se esperaba de las pruebas planeadas teóricamente, así como de la calidad de los procedimientos que se emplearon en el

desarrollo y la validación de la medición.

El resto del capítulo describe las diversas maneras de demostrar o proporcionar evidencia de validez de constructo. Estos tipos de validez pueden utilizarse en el desarrollo de la medición para generar y desarrollar ítems o indicadores válidos de constructos no observables y/o para proporcionar evidencia de la calidad de la medición. Usando clasificaciones similares a las propuestas por Haynes et al. (1999) y Trochim (2002), los diversos tipos o pruebas de validez discutidos en el capítulo se agrupan de la siguiente manera:

1. Validez de versión (*translation validity*)
 - Validez aparente (*face validity*)
 - * Validez del contenido (*content validity*)
2. Validez relacionada con el criterio (*criterion-related validity*)
 - Validez predictiva y post-dictiva (*predictive and post-dictive validity*)
 - Validez concurrente (*concurrent validity*)
 - * Validez convergente (*convergent validity*)
 - * Validez discriminante (*discriminant validity*)
 - Validez del grupo conocido (*known-group validity*)
- 3.* Validez nomológica (*nomological validity*)

A continuación, se presentarán los conceptos de cada tipo de validez antes señalado, para su conocimiento y exploración. Sin embargo, las que se abordarán a detalle son las marcadas con *, por su utilidad y practicidad. Así, definiremos que la validez del constructo se divide en dos: validez convergente (*convergent validity*) y validez discriminante (*discriminant validity*). Una escala tiene validez de constructo, cuando tiene ambas.

Validez de versión (*translation validity*)

La validez de contenido (*content validity*) y la validez aparente (*face validity*) reflejan la medida en que un constructo se traduce en la operacionalización del constructo (Trochim, 2002). Los términos a menudo se confunden y/o se utilizan indistintamente. Aunque la distinción entre la validez aparente y la validez de contenido frecuentemente no está clara, Rossiter (2001) argumentó que los dos conceptos de validez difieren en formas importantes y no deben ser confundidos.

Validez aparente

La evidencia de la validez aparente (*face validity*) se proporciona a partir de una evaluación post hoc en la que los ítems o indicadores de una escala miden adecua-

damente el constructo (Nunnally y Bernstein, 1994; Rossiter, 2001). Esta validez, puede ser juzgada después de que una medición ha sido desarrollada, a menudo antes de la aplicación en otro estudio, por los usuarios potenciales de dicha medición. Sin embargo, en exámenes posteriores del contenido de un indicador, para una medición, sólo ofrece pruebas incompletas de validez, porque los evaluadores sólo ven los indicadores que quedan y tienen que deducir lo que se omitió y por qué. Por lo tanto, la validez aparente puede considerarse como limitado en la validez de contenido, (Nunnally y Bernstein, 1994: 110). Además, algunos argumentan que la validez aparente es más similar a lo que un encuestado puede inferir acerca de lo que pretende medir un indicador (Nevo, 1985) y ha sido referido como “la mera suposición de que una medida tiene validez” (Kaplan y Saccuzzo, 1997: 1320). Por lo tanto, una alta validez aparente de un instrumento mejora su uso en situaciones prácticas al inducir la cooperación entre los encuestados a través de la facilidad de uso, el nivel de lectura adecuado, la claridad y los formatos de respuesta apropiados.

Validez de contenido

Las garantías de validez de contenido (*content validity*) se basan sobre esfuerzos de producción, a *priori*, teórica, de la generación de indicadores a medir. Específicamente, la validez del contenido representa el grado en que los elementos de un instrumento de medición son relevantes y representativos del constructo objetivo, para un propósito de evaluación particular (Haynes *et al.*, 1995: 238). La representatividad se refiere al grado en que los elementos son proporcionales a las facetas del constructo objetivo y al grado en que se ha muestreado todo el dominio de dicho constructo. Es decir, la medición debe ser coherente con el dominio teórico de la constructo en todos los aspectos, incluyendo la formulación de los ítems, los formatos de respuesta y las instrucciones (Haynes y otros, 1999; Haynes *et al.*, 1995; Netemeyer, Pullig y Bearden, 2002; Robinson *et al.*, 1991). El objetivo básico, entonces, es asegurar que los ítems reflejen las áreas de contenido abarcadas por el constructo objetivo. Como tal, la validez de contenido se manifiesta a partir de procedimientos en las etapas iniciales del desarrollo de la escala que generan ítems o indicadores representativos del dominio del enfoque del constructo. El desarrollo de una medición válida de contenido, se mejora durante las primeras etapas del desarrollo de la escala mediante la creación efectiva de un *pool de indicadores* y la posterior evaluación de los indicadores por parte de expertos.

En primer lugar, los ítems iniciales de dicho pool, deben elegirse de modo que muestren todas las áreas de contenido posibles que podrían ser abarcadas por el constructo de acuerdo con el conjunto completo de perspectivas teóricas que subyacen a dicho constructo (Clark y Watson, 1995). El grupo inicial de ítems debe

ser amplio en cobertura e incluir un gran número de indicadores o ítems potenciales a través de las dimensiones teóricas realizadas a *priori*. El grupo “*pool de indicadores*”, inicial debe incluir, sin duda, algunos ítems que posteriormente serán eliminados en los procedimientos de evaluación de seguimiento y análisis psicométricos. Un gran número de indicadores aumenta la probabilidad de que todas las dimensiones se representen adecuadamente. Es decir, el enfoque en la amplitud de la agrupación de elementos se aplica a todas las áreas o dimensiones del constructo posibles, de modo que las áreas individuales no estén subrepresentadas en la escala final. La validación de contenido es particularmente importante para constructos ambiguos o complejos. La validación de contenido se mejora mediante la definición precisa y la conceptualización del constructo, incluyendo la especificación de la dimensionalidad y de las definiciones individuales de las diversas dimensiones que comprende el constructo. Además, se garantiza la validez de contenido en la medida en que los expertos coinciden en que los ítems que sean de efecto o reflectivos (*reflective*) en toda la construcción global y que estos jueces estén de acuerdo en que los ítems son representativos del dominio y las facetas del constructo.

Los indicadores o ítems que se incluirán en un grupo, se pueden obtener usando cualquier número de fuentes. Las fuentes de indicadores con más frecuencia utilizadas en el desarrollo de escalas incluyen declaraciones, previamente empleadas de investigaciones que involucran el constructo, casos de muestras ya cerrados, de sujetos representativos y declaraciones generadas por investigadores basadas en el conocimiento y comprensión del dominio del constructo y sus fundamentos teóricos.

Como ejemplo de validez de contenido, se tiene el trabajo de “Los factores determinantes del modelo de negocios abierto” (OBM) de Mejía-Trejo (2017), el cual muestra la creación de un modelo empírico del OBM, que consiste en cinco principales factores/22 variables/139 ítems o indicadores distribuidos en administración del negocio (BMG, 10 variables/76 indicadores), estrategia (STR, 3 variables/14 indicadores), tecnología (TEC, 3 variables/24 indicadores), nuevos emprendimientos (NWE, 3 variables /7 indicadores) y orientación de la innovación abierta (OIO, 3 variables/18 indicadores). Conclusión. Aunque el modelo empírico final de OBM tiene un efecto positivo significativo entre sus variables, también mostró diferentes niveles de carga de factores, lo que significa oportunidades para mejorar el modelo para el sector de las TIC de la zona metropolitana de Guadalajara (ITSMZG), México. Se trató de un estudio documental para seleccionar las principales variables entre los especialistas de las ITSMZG que practican el proceso OBM mediante el proceso de jerarquía analítica (AHP. *analysis hierarchy process*) y el panel de Delphi a fin de contrastar los términos académicos con la experiencia de los especialistas. Es un estudio descriptivo, exploratorio, correlacional, transeccio-

nal, cualitativo-cuantitativo para obtener un cuestionario final en escala Likert, con confiabilidad a través de la prueba piloto (alfa de Cronbach > 0.7), aplicado entre enero 2015-mayo 2016 a una población total de 600 especialistas en el ITSMZG (150 profesores de IT; 150 representantes de consultores de firmas IT como “parte consultora”; 290 CEO PYME y 10 CEO de empresas grandes como parte de “toma de decisiones”, con 1 año en el mercado, 80% con licenciatura, 20% con posgrado, 20% mujeres y 80% hombres). Se diseñó un modelo de ecuaciones estructural de primer orden (SEM) como técnica de análisis factorial confirmatorio (CFA), mediante el software EQS 6.2 para analizar las variables subyacentes de OBM, y determinar un modelo final.

Estudio de caso: administración del conocimiento (OECD, 2003)

Al final, una escala no es sino un conjunto de indicadores que se supone reflejan todas las dimensiones del concepto que pretenden medir. Se define la validez de contenido como el grado en que la escala recoge todas esas dimensiones. Retomando el ejercicio mostrado en la figura 3: Diagrama de trayectoria constructo administración del conocimiento (capítulo 3), vemos que la escala propuesta por la OECD (2003), se compone de dos factores: COMUNICACIÓN e INCENTIVOS. La escala tendrá validez de contenido si los ítems seleccionados de entre todos los posibles para construirla, recogen adecuadamente esos dos factores.

Queda claro, de la explicación precedente, que es muy difícil determinar si una escala tiene validez de contenido, por cuanto que se basa en el juicio del experto que se supone que ha considerado todos los posibles factores y contenidos del concepto, fenómeno o comportamiento analizado. Un siguiente paso ahora es estimar el modelo mediante la utilización del software EQS, aunque no se trata de enseñar el manejo del programa, sí que es muy importante que se vea cómo la sintaxis refleja el conjunto de relaciones que se han puesto de manifiesto en la figura 3 (capítulo 3), y la inmediatez de las instrucciones.

Considerando la tabla 12: Sintaxis EQS (capítulo 3), para hacerla comprensible, expliquemos algunas instrucciones. En la primera línea, se indica únicamente que se ejecute el programa, tomando en cuenta que los datos son correlaciones, pero que deben transformarse en varianzas-covarianzas. Las ecuaciones (/EQUATIONS) se corresponde a la formulación de Bentler (Bentler y Wu, 1993). Tomemos las ecuaciones que definen el modelo, por ejemplo:

$$V3 = *F1 + E3$$

Si nos fijamos en la figura 3, vemos que V3 se obtiene multiplicando por una cierta carga factorial L el factor F1, sujeta a un error que hemos llamado E₃. Pues

bien * es como EQS indica que la carga factorial L debe ser estimada, ese parámetro aparece multiplicado por f1 y sumándosele E3 nos daría el indicador V3.

Bajo VARIANCES aparecen las varianzas de los dos factores que tienen que ser estimadas, por eso se observa que la varianza de F1 y la de F2 están con un asterisco (*).

/VARIANCES
F1 TO F2 = *

Mientras que las varianzas de los términos de error E₁ hasta E₆ también deben ser estimadas:

E1 TO E6 = *

Finalmente, bajo /covariances se le indica al programa qué parámetros de las covarianzas se fijan a una cantidad y cuáles hay que estimar. Como en nuestro caso todas deben ser estimadas, no aparecerá F1, F2 = 1, sino F1, F2 = *, es decir que el parámetro de covarianzas entre F1 y F2 debe estimarse. Una forma de no tener que escribir todas las combinaciones entre los 6 factores es utilizar el término TO:

/COVARIANCES F1 TO F6 = *;

Pues bien, con esas sintaxis EQS estima el modelo. El primer paso en la interpretación de los resultados pasa por determinar cuál es la *bondad de ajuste* del mismo (tabla 1).

TABLA 1. Estadísticos de bondad de ajuste proporcionados por EQS.

GOODNESS OF FIT SUMMARY FOR METHOD = ML			
INDEPENDENCE MODEL CHI-SQUARE	=	110.190 ON	15 DEGREES OF FREEDOM
INDEPENDENCE AIC =	80.190	INDEPENDENCE CAIC =	36.509
MODEL AIC =	-4.301	MODEL CAIC =	-27.597
CHI-SQUARE =	11.699	BASED ON	8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS		.16515	
THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS			10.841.
FIT INDICES			

BENTLER-BONETT	NORMED FIT INDEX =	.894	
BENTLER-BONETT	NON-NORMED FIT INDEX =	.927	
COMPARATIVE	FIT INDEX (CFI) =	.961	
BOLLEN'S	(IFI) FIT INDEX =	.964	
MCDONALD'S	(MFI) FIT INDEX =	.964	
JORESKOG-SORBOM'S	GFI FIT INDEX =	.931	
JORESKOG-SORBOM'S	AGFI FIT INDEX =	.820	
ROOT MEAN-SQUARE	RESIDUAL (RMR) =	.165	
STANDARDIZED	RMR =	.081	
ROOT MEAN-SQUARE	ERROR OF APPROXIMATION (RMSEA) =	.097	
90% CONFIDENCE INTERVAL	OF RMSEA (.000,	.206)

Fuente: EQS 6.2. con datos propios.

Interpretación

De su análisis se desprende lo siguiente.

La prueba Chi cuadrado analiza la hipótesis nula de que el modelo es no significativo. Esta hipótesis debería poder rechazarse ($p > 0.05$). Sin embargo, con grandes muestras, el estadístico Chi cuadrado tiene tendencia a ser significativo, incluso aunque el modelo de un buen ajuste (James, Mulaik y Brett, 1982). Por ello, además de analizar otros estadísticos que luego veremos, algunos autores proponen que el estadístico es aceptable si el ratio entre el valor de la Chi cuadrado y el número de grados de libertad es inferior a 2. En nuestro caso, la Chi cuadrado es significativa, pero el ratio es inferior a 2 ($11.699/8 = 1.46$). Parece indicar que el modelo de *administración del conocimiento* tiene un buen ajuste, pero podemos seguir analizando otros estadísticos.

Entre el conjunto de indicadores de ajuste que proporciona EQS, se recomienda analizar aquellos no normalizados, dado que tienen menor tendencia a ofrecer resultados sesgados en casos de pequeñas muestras (Bentler y Wu, 1993). Se observa que, tanto el “Non normed Fit Index – NNFI” (Bentler y Bonnet, 1980) como el “Comparative FIT Index – CFI” (Bentler y Wu, 1993), que cumplen este requisito, ya que tienen valores superiores a 0.9, indicando un ajuste razonable.

Con esto, ya hemos estimado el modelo de medida y hemos comprobado que ofrece un buen ajuste. Pues bien, sobre este modelo de medida, ya podemos analizar la validez convergente y la validez discriminante de las escalas implicadas. Para analizar la validez nomológica, será necesario estimar el modelo teórico.

Validez relacionada con el criterio

Los esfuerzos para demostrar la validez relacionada con el criterio (*criterion-related validity*) implican procesos en los que se emplean mediciones externas al instrumento de medición propuesto (Nunnally y Bernstein, 1994). Las fuentes de validez específicas que pueden utilizarse bajo la etiqueta esta etiqueta, incluyen la validez predictiva y pos-dictiva, la validez concurrente, la validez convergente y discriminante y la validez del grupo conocido.

Validez predictiva y post-dictiva (predictive and post-dictive validity)

El término validez predictiva a menudo se utiliza indistintamente como validez de criterio. En aplicaciones tradicionales que implican desarrollo de escala, la validez predictiva se refiere a la capacidad de una medida para predecir con eficacia algún criterio subsiguiente y temporalmente ordenado.

Un ejemplo, es el desarrollo de una prueba de conocimiento de la persuasión del consumidor para su uso con adolescentes (Friestad y Wright, 2001) puede usarse para ilustrar el problema. Así, la validez de tal prueba podría ser apoyada por el tamaño de las correlaciones entre una prueba de conocimiento de persuasión del consumidor e índices o una observación de los comportamientos efectivos del consumidor que se producen después de que la prueba se ha administrado.

La validez del criterio post-dictivo ocurre cuando la variable de resultado se mide antes de la variable independiente. Por ejemplo, los hogares de consumidores pueden ser observados por comportamientos que demuestren las necesidades únicas de consumidores (CNFU. *Consumers' need for uniqueness*). Posteriormente, una escala diseñada para medir el CNFU es administrada a un adulto de cada hogar. Una correlación significativa entre un índice que comprende los comportamientos observados y la escala del CNFU ofrecería alguna evidencia de validez post-dictiva para la escala de CNFU (Tian, Bearden y Hunter, 2001).

Validez concurrente

La evidencia de la validez concurrente (*concurrent validity*) de una medición es proporcionada por correlaciones considerables entre la medida de constructo en desarrollo y una medida de criterio recogida simultáneamente o “concurrentemente”. A modo de ejemplo, Bearden et al. (2001) aportaron pruebas de la validez concurrente de sus mediciones de su indicador “auto-confianza del consumidor” utilizando correlaciones entre las escalas de dicha variable y las mediciones de otra variable llamada “conocimiento subjetivo del producto” (Park, Mothersbaugh y Feick, 1994). Además, estos esfuerzos de desarrollo incluían también la demostración de la validez de la escala relativa, en las que las mediciones de confiabilidad que se proponían estaban más fuertemente correlacionadas que de las mediciones fueron realizadas del indicador “autoestima” (Rosenberg, 1965) y el procesamiento del indicador “autoconfianza” (Wright, 1975). En estas comparaciones posteriores, se obtuvieron pruebas de relativa validez concurrente a partir de diferencias significativas en las pruebas de correlaciones dependientes, lo que sugirió que las mediciones evaluadas estaban más fuertemente correlacionadas con el criterio concurrente (es decir, evaluaciones subjetivas de conocimiento) que las medidas competitivas.

Validez convergente

Se dice que una medición posee validez convergente (*convergent validity*) si las mediciones independientes del mismo constructo convergen o están altamente

correlacionadas. La evidencia de validez convergente es ofrecida por correlaciones significativas y fuertes entre diferentes medidas del mismo constructo.

Estudio de caso: administración del conocimiento (OECD, 2003)

La validez convergente existe cuando se emplean distintos instrumentos para medir un mismo constructo (distintos indicadores para una misma variable latente) y estos instrumentos están fuertemente correlacionados. Como veremos al desarrollar el ejemplo que vendrá a continuación, en este tema la validez convergente se determinará revisando los test t de las cargas factoriales. Si todas las cargas factoriales de las variables manifiestas que miden el mismo constructo son estadísticamente significativas, será una evidencia que apoya la validez convergente de estos indicadores (Anderson y Gerbing, 1988). Hair, Anderson, Tatham y Black (1998) recomiendan que, además de ser significativos, el promedio de las cargas sobre cada factor sean superiores a .7. También deberíamos analizar si el modelo mejoraría significativamente si se añadiera alguna relación entre una variable manifiesta diseñada para medir un factor dado y otro factor para el que no se ha diseñado. Los multiplicadores de Lagrange nos permiten verificar esta posibilidad.

Si analizamos las tablas 2, 3 y 4 para el modelo de medida observamos:

- a. Todas las cargas factoriales son significativas, como se desprende de sus respectivas t (3.073 para la carga entre V2 y F1; 3.810 para la carga entre V3 y F1...). Si este valor de t es superior a 1.96, el parámetro es significativo para $p < 0.05$, si es superior a 2.576 es significativo para $p < 0.01$ y si es superior a 3.291 para $p < 0.001$. Puede comprobarse en la misma Tabla 4.2 que el parámetro $p < 0.01$ es significativo para la variable V2 y de $p < 0.001$ para las variables V3, V5 y V6. Asimismo, si calculáramos el promedio de los mismos, veríamos que superan claramente .7
- b. El ajuste del modelo mejoraría significativamente si la variable V4 también cargara sobre el factor F1 (figura 4, primera línea).

Interpretación

Estos últimos resultados cuestionan la validez convergente (*convergent validity*) de la escala. El que la Chi cuadrado mejorara significativamente si se añadiera la relación entre V_4 y F_1 (se reduciría en un valor de 4.143), es coherente con el patrón de residuos elevados que hemos visto con anterioridad (RMSEA superior a 0.08). Ante una situación como ésta pueden hacerse varias cosas:

- ▶ Lo más fácil es añadir esa relación, pero esto es tanto como asumir que el modelo no posee validez convergente.
- ▶ Otra alternativa es asignar V4 a F1 y eliminarla de F2, pero vemos que la carga factorial sobre F1 es muy elevada, lo que hace esta opción también inviable. Por tanto, la opción más razonable es asumir que V4 no es un buen ítem de la escala y eliminarlo completamente. Por lo tanto, en la sintaxis eliminaremos la expresión: $V4 = * F1 + E4$.

A continuación reestimamos el modelo y volvemos a analizar la validez convergente a partir de las mismas salidas de bondad de ajuste (tabla 5), significatividad de las cargas factoriales (tabla 6) y valor de las cargas estandarizadas (tabla 7).

TABLA 2. Estimación de las cargas factoriales.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.			
X1	=V1	= 1.000 F1	+1.000 E1
X2	=V2	= .423*F1 .138 3.073@	+1.000 E2
X3	=V3	= .671*F1 .176 3.810@	+1.000 E3
X4	=V4	= 1.000 F2	+1.000 E4
X5	=V5	= .646*F2 .164 3.932@	+1.000 E5
X6	=V6	= .616*F2 .154 3.994@	+1.000 E6

Fuente: EQS. 6.2. con datos propios..

TABLA 3. Estimación de las cargas factoriales estandarizadas.

STANDARDIZED SOLUTION:			
X1	=V1	= .963 F1	+ .270 E1
X2	=V2	= .514*F1	+ .858 E2
X3	=V3	= .741*F1	+ .671 E3
X4	=V4	= .946 F2	+ .323 E4
X5	=V5	= .656*F2	+ .755 E5
X6	=V6	= .672*F2	+ .740 E6

Fuente: EQS. 6.2. con datos propios..

TABLA 4. Estimación de las cargas factoriales estandarizadas.

CUMULATIVE MULTIVARIATE STATISTICS					UNIVARIATE INCREMENT			
STEP	PARAMETER	CHI-SQUARE	D.F.	PROB.	CHI-SQUARE	PROB.	D.F.	HANCOCK'S SEQUENTIAL PROB.
1	V4, F1	4.143	1	.042	4.143	.042	8	.844

Fuente: EQS. 6.2. con datos propios..

TABLA 5. Indicadores de ajuste del modelo corregido.

```

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE          =          66.051 ON          10 DEGREES OF FREEDOM

INDEPENDENCE AIC =          46.051  INDEPENDENCE CAIC =          16.931
MODEL AIC =          -2.897  MODEL CAIC =          -14.545

CHI-SQUARE =          5.103 BASED ON          4 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS          .27691

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS          4.949.

FIT INDICES
-----
BENTLER-BONETT          NORMED FIT INDEX =          .923
BENTLER-BONETT NON-NORMED FIT INDEX =          .951
COMPARATIVE FIT INDEX (CFI) =          .980
BOLLEN'S          (IFI) FIT INDEX =          .982
MCDONALD'S          (MFI) FIT INDEX =          .989
JORESKOG-SORBOM'S          GFI FIT INDEX =          .961
JORESKOG-SORBOM'S          AGFI FIT INDEX =          .854
ROOT MEAN-SQUARE RESIDUAL (RMR) =          .093
STANDARDIZED RMR =          .050
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =          .075
90% CONFIDENCE INTERVAL OF RMSEA (          .000,          .237)
    
```

Fuente: EQS. 6.2. con datos propios..

Analizando los indicadores de ajuste vemos que la Chi cuadrado ha disminuido sustantivamente y que los errores son menores (RMSEA = 0.075). Pero en lo que nos interesa para analizar la validez convergente, todos los coeficientes de regresión son significativos, con lo que esta validez puede afirmarse.

Validez discriminante

La validez discriminante (discriminant *validity*) requiere que una medición no se correlacione demasiado con mediciones de las que se supone debe diferir (Churchill y Iacobucci, 2002: 413). En términos de desarrollo y validación de mediciones, la evidencia de validez convergente típicamente se proporciona a partir de las correlaciones entre la nueva medición que se está desarrollando y las mediciones existentes. Obviamente, los problemas para generar evidencia de validez convergente a menudo ocurren en el desarrollo de mediciones cuando no hay mediciones alternativas disponibles.

Estas instancias se encuentran frecuentemente en constructos que no han sido estudiados previamente o que han sido investigadas usando métodos de operacionalización *ad hoc* o desarrollados de manera inadecuada.

Cuando las escalas previamente validadas no están disponibles para proporcionar la variedad necesaria en la medición, a menudo se emplean medidas de otros campos, como la psicología (Richins, 1983).

Entre los métodos empleados más frecuentemente para investigar la validez convergente, así como la validez discriminante, se encuentra la matriz multitrait-multimétodo (MTMM. *Multitrait-multimethod matrix*) propuesta por Campbell y Fiske (1959). Como se resume en Cook y Campbell (1979: 61), como centro de la evaluación de la validez del constructo hay dos procesos:

- a. Probar la convergencia entre diferentes mediciones del mismo concepto y,
- b. Probar la divergencia entre mediciones relacionadas pero conceptualmente distintas en conceptos.

La matriz MTMM típica muestra una serie de coeficientes de correlación de Pearson entre mediciones de constructos no observables. Casos de esto, son:

1. Richins (1983: 77) ofreció un ejemplo de “comportamiento del consumidor” en el que el análisis incluía dos rasgos (es decir, “defensa” y “agresión del consumidor”, como términos mercadotécnicos) y dos métodos (es decir, las escalas tipo Likert desarrolladas y diferentes escalas que evalúan la “defensa” y la “agresión” como términos de la psicología).
2. Nunnally y Bernstein (1994: 93) ofrecieron otro ejemplo, con dos rasgos: “ansiedad” y “depresión” con dos métodos: los auto-informes de “ansiedad” y “depresión”, con puntuaciones de observación obtenidas de expertos externos (gráfica 1).

Como se muestra en las gráficas 1 y se describe a continuación, Churchill (1979: 70-71) describió un ejemplo de “gestión de ventas” que incluía tres rasgos “satisfacción laboral, conflicto de roles y ambigüedad de roles y 2: métodos (Likert y

TABLA 6. Estimación de las cargas factoriales del modelo corregido.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS			
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.			
X1	=V1	= 1.000 F1	+1.000 E1
X2	=V2	= .442*F1	+1.000 E2
		.137	
		3.221@	
X3	=V3	= .707*F1	+1.000 E3
		.176	
		4.024@	
X5	=V5	= 1.000 F2	+1.000 E5
X6	=V6	= .614*F2	+1.000 E6
		.955	
		.643	

Fuente: EQS. 6.2. con datos propios..

TABLA 7. Estimación de las cargas factoriales del modelo corregido.

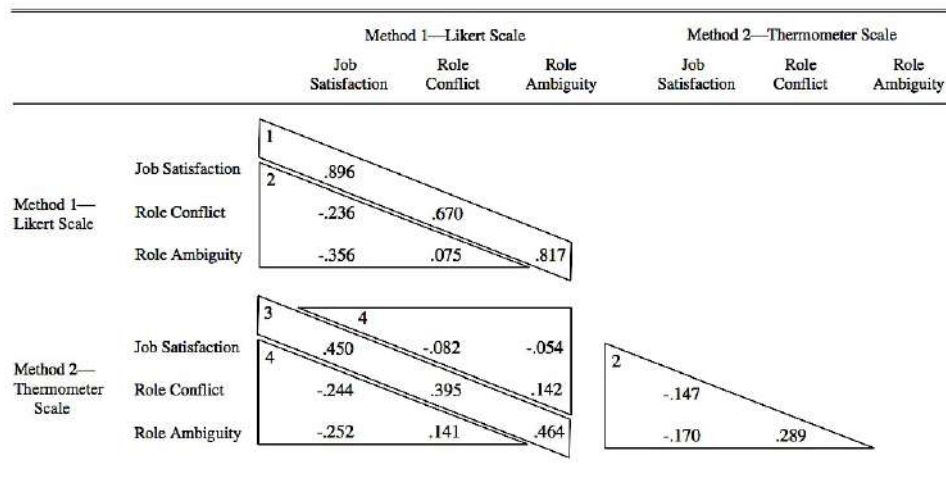
STANDARDIZED SOLUTION:			
X1	=V1	= .938 F1	+ .347 E1
X2	=V2	= .523*F1	+ .852 E2
X3	=V3	= .761*F1	+ .648 E3
X5	=V5	= .831 F2	+ .557 E5
X6	=V6	= .548*F2	+ .837 E6

Fuente: EQS. 6.2. con datos propios..

termómetro escalas). El patrón de correlaciones dentro de la matriz (es decir, la fuerza relativa y la importancia de las correlaciones incluidas en la matriz) se interpretan para evidencia de validez convergente y discriminante. (Kenny y Kashay [1992] y Widaman [1985] describieron los procedimientos de uso del análisis factorial confirmatorio para analizar datos multitrait-multimétodo).

De este segundo caso, en la gráfica 1, las estimaciones de confiabilidad de consistencia interna se representan en diagonal 1, en la esquina superior izquierda

GRAFICA 1. Matriz Multitrait Multi-Método.



Fuente: Adaptación de “a paradigm for developing better measures of marketing constructs”. Journal of Marketing Research, 16 (february), 1979, p. 17, Churchill, copyright © 1979 by the American Marketing Association.

de la tabla.

Las correlaciones incluidas en la diagonal 3, en la esquina inferior izquierda, proporcionan evidencia de validez convergente. En este ejemplo, las tres correlaciones en la diagonal 3 ofrecen soporte para la validez convergente, ya que las estimaciones son significativas y grandes en relación con otras correlaciones en la matriz.

La evidencia de la validez discriminante es ofrecida por las correlaciones en el bloque heterométrodo 4 en la esquina inferior izquierda (es decir, las correlaciones entre las mediciones de constructo para las cuales se emplean diferentes enfoques de medición) y los dos triángulos heterotrait-monométodo en las esquinas superior izquierda e inferior derecha (es decir, las correlaciones entre las medidas de constructo para las cuales se emplea un enfoque de medición único). la evidencia de la validez discriminante implica tres comparaciones (Churchill, 1979: 71):

- Las entradas en la diagonal de validez 3 deben ser mayores que las entradas en el bloque de heterométrodos 4 que comparten la misma fila y columna.
- Las correlaciones en la validez diagonal deben ser mayores que las relaciones en los triángulos de heterotrait-monométodo Este requisito más estricto sugiere que las correlaciones entre diferentes mediciones para un rasgo deben ser más altas que las correlaciones entre rasgos que tienen métodos en común.
- El patrón de las correlaciones culturales debe ser el mismo en todos los

triángulos de contorno (es decir, 2 y 4).

En el ejemplo, se cumplen las dos primeras condiciones para la validez discriminante.

Estudio de caso:
administración del conocimiento (OECD, 2003)

Por otro lado, es posible afirmar que la validez discriminante (*discriminant validity*) se consigue cuando, si diversos instrumentos de medida están diseñados para medir distintas variables latentes, entonces las correlaciones entre dichos instrumentos son bajas. Una escala tiene, pues, validez discriminante cuando no mide un constructo para el que no se diseñó. Así, volviendo al caso de la figura 3. Diagrama de trayectoria constructo administración del conocimiento de la OECD (2003), por ejemplo, si la variable latente F1 COMUNICACIÓN y la F2 incentivos, estuvieran fuertemente correlacionadas entre ellas, podría ocurrir que sirvieran para medir el mismo concepto (un caso extremo de una correlación 1 entre ambas), (tabla 8).

TABLA 8. Estimación de covarinzas del modelo corregido.

COVARIANCES AMONG INDEPENDENT VARIABLES	

STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH *.	
V	F
---	---
I F2 - F2	.293*I
I F1 - F1	.032 I
I	.883 I
I	I

Fuente: EQS. 6.2. con datos propios..

Donde se muestran las estimaciones de los coeficientes de las trayectorias (*paths*) entre los distintos factores, vemos que la relación entre el factor F1 y F2 no es muy alta (coeficiente 0.293). Podemos plantearnos si los indicadores de F1 servirían o no para medir F2, con lo que la validez discriminante de las dos escalas podría cuestionarse. Plantaremos tres procedimientos para analizar este hecho:

- ▶ Test de la diferencia entre las Chi cuadrado
- ▶ Test del intervalo de confianza
- ▶ Test de la varianza extraída

Test de la diferencia entre las Chi cuadrado

Se procede del siguiente modo:

- Estima mediante un análisis factorial confirmatorio el modelo de medida donde todas las variables latentes pueden covariar (ya lo hemos hecho).
- Crea un nuevo modelo de medida idéntico al anterior, pero donde el coeficiente entre los dos factores cuya validez discriminante suscita dudas, es fijado a 1 (es decir, serían el mismo factor, con lo que sus ítems serían perfectamente intercambiables y no habría validez discriminante).
- Se calcula un test para la diferencia entre las Chi cuadrado de los dos modelos. La validez discriminante quedará demostrada sólo si la Chi cuadrado es significativamente menor (mejor ajuste del modelo) para el primer modelo en el cual los dos constructos son vistos como factores distintos, aunque correlacionados (Anderson y Gerbing, 1988).

Dado que el primer modelo ya ha sido estimado, y su Chi cuadrado aparece en la tabla 5, modificaremos la sintaxis de EQS de ese modelo retocando el subcomando COVARIANCES, como se indica a continuación, para forzar que la covarianza entre los factores F1 y F2 sea 1 y estimamos con ella el modelo (tabla 9).

TABLA 9. Bondad de ajuste del modelo alternativo.

GOODNESS OF FIT SUMMARY FOR METHOD = ML			
INDEPENDENCE MODEL CHI-SQUARE	=	66.051 ON	10 DEGREES OF FREEDOM
INDEPENDENCE AIC =	46.051	INDEPENDENCE CAIC =	16.931
MODEL AIC =	-1.479	MODEL CAIC =	-16.039
CHI-SQUARE =	8.521	BASED ON	5 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS			.12976
THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS			7.718.
FIT INDICES			

BENTLER-BONETT	NORMED FIT INDEX =		.871
BENTLER-BONETT	NON-NORMED FIT INDEX =		.874
COMPARATIVE FIT INDEX (CFI)	=		.937
BOLLEN'S	(IFI) FIT INDEX =		.942
MCDONALD'S	(MFI) FIT INDEX =		.965
JORESKOG-SORBOM'S	GFI FIT INDEX =		.937
JORESKOG-SORBOM'S	AGFI FIT INDEX =		.811
ROOT MEAN-SQUARE RESIDUAL (RMR)	=		.342
STANDARDIZED RMR	=		.162
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)	=		.120
90% CONFIDENCE INTERVAL OF RMSEA	(.000,	.251)

Fuente: EQS. 6.2. con datos propios..

/COVARIANCES

F1, F2=1;

Pues bien, la Chi cuadrado del primer modelo era 5.103 con 4 grados de libertad, mientras que la del segundo modelo resulta ser 8.521 con 5 grados de libertad. La diferencia entre ambos estadísticos y sus grados de libertad es:

Chi cuadrado: 3.418 grados libertad: 1

Si buscamos en una tabla de la distribución Chi cuadrado vemos que, para un grado de libertad, el valor crítico es 2.706 para $p = 0.10$, 3.84 para $p = 0.05$; 6.635 para $p = 0.01$ y 10.827 para $p = 0.001$. Por lo tanto, dado que la diferencia es de 3.418, la diferencia entre los dos modelos es claramente significativa. En otras palabras, el modelo de medida estándar en el que los factores están correlacionados, pero son retenidos como distintos es significativamente mejor que el que los considera un único factor. ello afirmaría la validez discriminante de las escalas utilizadas para medir f_1 y f_2 . el proceso debería repetirse para cada par de posibles covarianzas.

Test del intervalo de confianza

Anderson y Gerbing (1988) proponen también el llamado este test, para analizar la validez discriminante entre dos escalas. Este test implica calcular un intervalo de confianza de ± 2 errores estándar entre la correlación entre los factores y determinar si este intervalo incluye al 1.0. Si no incluye al 1.0, la validez discriminante quedará confirmada.

De la tabla 4.8, se observa que el valor del coeficiente de correlación entre F_1 y F_2 es de 0.293 y su error estándar de 0.032. De esta forma el intervalo de confianza tendría estos extremos superior e inferior:

$$\text{Intervalo inferior: } 0.293 - 2 \times 0.032 = 0.229$$

$$\text{Intervalo superior: } 0.293 + 2 \times 0.032 = 0.357$$

El intervalo queda, consecuentemente, [0.229 – 0.357], no entra el 1, lo que confirma la validez discriminante de las dos escalas según el criterio expuesto. En resumen, los dos criterios expuestos confirman la validez discriminante de las escalas analizadas.

Test de la varianza extraída

Un tercer procedimiento para establecer la validez discriminante es aplicar este test (Fornell y Larcker, 1981). Según este procedimiento, calcularíamos el índice de varianza extraída por cada uno de los factores cuya validez discriminante se está

calculando (tal y como explicamos al tratar la fiabilidad). Este IVE se compara con el cuadrado de las correlaciones entre los dos factores. Podemos afirmar la validez discriminante si los IVE de los dos factores son mayores que el cuadrado de la correlación.

En nuestro ejemplo la correlación entre F1 y F2 es de 0.293, cuyo cuadrado es 0.085 (tabla 10).

TABLA 10. Recopilación de información que calcula el IVE.

Constructo e indicadores	Cuadrado de la carga factorial estandarizada $(L_{ij})^2$	Varianza del término de error $\text{Var}(E_{ij}) = 1 - (L_{ij})^2$	Índice de varianza compuesta IVE
COMUNICACIÓN			
X1=V1	.880	.120	.578
X2=V2	.274	.726	
X3=V3	.579	.421	
Total	1.733	1.267	
INCENTIVOS			
X5=V5	.691	.309	.496
X6=V6	.300	.568	
Total	.991	1.009	

Fuente: Elaboración propia.

De acuerdo con este criterio, como el índice de varianza extraída de F2 es superior al cuadrado de la correlación (.085), podríamos afirmar la validez discriminante. El resultado, sin embargo, está tan en el límite que la dirección en que apuntan los otros dos indicadores ofrecidos permitiría afirmar esta validez sin demasiado riesgo de error (0.496).

Validez del grupo conocido

Esta validez del grupo conocido (*known-group validity*), se refiere a grado en que una medida difiere según lo previsto entre los grupos, con un nivel de bajo y/o alto puntaje en un rasgo. La evidencia de apoyo de la validez de grupo conocido es típicamente proporcionada por diferencias significativas en las puntuaciones medias a través de muestras independientes. Tian *et al.* (2001: 56-57) aportaron varios ejemplos interesantes para validar su escala de necesidades únicas de consumidores (CNFU. *consumers' need for uniqueness*). Específicamente, la puntuación media de un indicador para una muestra heterogénea de encuestas por correo para su escala CNFU de 31 ítems fue de 2.60 (n = 621). En la tabla 11 se presentan tres muestras de comparación de grupos conocidos, el tipo de diferenciación de consumo que representó el grupo y los resultados de cada muestra. Las

muestras de comparación de grupos conocidos fueron las siguientes: artistas del tatuaje (*tattoo artists*), que reflejan el dominio de las muestras corporales de bienes de consumo ($M = 3.05$, $n = 39$, $t = 3.22$, $p < 0,01$); propietarios de mini-camiones (*owners of customized low rider*), reflejando la posesión de objetos únicos ($M = 2.99$, $n = 22$, $t = 3.22$, $p < 0,01$); y miembros de la *Sociedad de Anacronismo Creativo* (*Society for Creative Anachronism*), que representan la participación en grupos únicos que se visten de manera diferente y promulgan actuaciones medievales ($M = 2,91$, $n = 21$, $t = 2.49$, $p < 0.01$).

Otro ejemplo interesante de la validez de los grupos conocidos fue Saxe y Weitz (1982). Al validar su escala de orientación del agente de ventas (SOCO. *Salesperson Customer Orientation*), en el que encontraron diferencias significativas entre las muestras de posiciones de ventas que difirieron ampliamente en su estatus profesional.

Validez nomológica

Un medio para evaluar la validez de constructo de una medición, viene de determinar hasta qué punto la medida encaja "lícitamente" en una red de relaciones o una "red nomológica" (Cronbach y Meehl, 1955). Es decir, un aspecto de la validez del constructo implica el nivel en que una medición opera dentro de un conjunto de constructos teóricos y sus respectivas medidas (Nunnally y Bernstein, 1994: 91).

La validez nomológica (*nomological validity*) se basa en investigaciones de constructos y mediciones en términos de hipótesis formales derivadas de la teoría. Como tal, la validación nomológica es principalmente externa e implica investigar tanto las relaciones teóricas entre constructos diferentes como las relaciones empíricas entre las medidas de esos constructos. Así, las mediciones deben demostrar validez nomológica para ser aceptadas como constructos válidos (Peter, 1981: 135).

La evidencia de validez nomológica es proporcionada por la posesión de un constructo con distintos antecedentes de causa, efectos y / o condiciones de modificación, así como diferencias cuantitativas en el grado en que un constructo está relacionado con antecedentes o consecuencias (Iacobucci, Ostrom, Grayson, 1995; Tian *et al.*, 2001: 59). Ciertamente, la evidencia proporcionada por los métodos descritos en las secciones anteriores de este capítulo puede tener implicaciones para validez nomológica. Por ejemplo, las diferencias en las puntuaciones para una medición de la fuerza de ventas al consumidor a través de grupos de vendedores ineficaces y efectivos proporcionan algunas implicaciones para las teorías de la venta efectiva (es decir, validez nomológica), así como de validez del

TABLA 11. Resumen caso validez del grupo conocido.

Known-Group Validity Test	Unique Group				Comparison Group				Validity Support	
	Sample	n	Mean	SD	n	Mean	SD	Value of t		Value of p
Tattoo and body piercing artists	1	39	3.05	.70	621	2.60	.56	3.22	<.001	Supported
Owners of customized low rider autos	2	22	2.99	.45	621	2.60	.56	3.22	<.001	Supported
Members of medievalist reenactment group	3	21	2.91	.44	621	2.60	.56	2.49	<.01	Supported
Student art majors	4	22	3.06	.45	273	2.71	.50	3.15	<.01	Supported
Student purchasers of unique poster art	5	78	2.83	.43	273	2.71	.50	1.89	<.05	Supported

SOURCE: Adapted from "Consumers' Need for Uniqueness: Scale Development and Validation," *Journal of Consumer Research*, 28(June), p. 58, Tian, Bearden, and Hunter, copyright 2001 by the University of Chicago Press.

grupo conocido. Así, los modelos de ecuación estructural (SEM. *Structural Equation Modeling*), métodos basados en regresión y/o experimentales se usan frecuentemente en esfuerzos para proporcionar evidencia corroborativa de validez nomológica (Bollen, 1989; Hoyle, 1995). Como ejemplo, Bearden et al. (2001) emplearon la regresión logística para demostrar que un subconjunto de sus mediciones de “confianza del consumidor” moderó (como planteó de hipótesis) la relación entre el esquema de calidad de los precios al consumidor y la elección de opciones más altas o más bajas en un conjunto de opciones. Lastovicka et al. (1999) utilizaron una regresión múltiple para demostrar que su medida de “*frugalidad*” explicaba un índice de comportamientos “*frugales*” más allá de explicaciones alternativas para comportamiento frugal.

Como otro ejemplo, Tian (2001) describió un gran número de pruebas de diferencias en las relaciones que implican una serie de antecedentes de rasgos (por ejemplo, “individualismo colectivo”), resultados (por ejemplo, elección de diseños exteriores únicos vs. comunes) y moderadores situacionales por ejemplo, la popularización de diseños únicos) para medir la necesidad de unicidad de los consumidores.

Estudio de caso: administración del conocimiento (OECD, 2003)

Finalmente, una escala tiene validez nomológica (*nomological validity*) cuando el constructo que miden es capaz de sacar a la luz relaciones con otros constructos que, conceptual y teóricamente deberían existir. Nótese que para contrastar esta validez ya no nos basta con tener una escala que mida una variable latente, sino diversas escalas con relaciones teóricas entre ellas. Por este motivo el ejemplo que desarrollaremos a continuación envuelve a diversas escalas con estas relaciones teóricas. Si los constructos cuya relación se comprueba se han medido al mismo tiempo que la escala, hablamos de validez concurrente (*concurrent validity*), si se han medido en distintos momentos del tiempo, hablamos de validez predictiva (*predictive validity*).

Hemos de señalar que existen muchos programas estadísticos para abordar el análisis factorial confirmatorio y el análisis de sistemas de ecuaciones estructurales necesarios para el contraste de la validez. Quizá el más conocido sea LISREL, desarrollado por Jöreskog y Sörbom (1993). SPSS, por su parte, incorpora en sus últimas versiones el programa AMOS.

Al final, la elección entre uno y otro depende del gusto particular del investigador. Nosotros, atendiendo a estos gustos, solemos recurrir a EQS (Bentler y Wu, 1993) aunque, puestos a encontrar un software óptimo, recomendaríamos la utilización del programa CALIS de SAS (SAS Institute Inc, 1989; Hatcher, 1994). La

razón de esta recomendación es que permite el planteamiento del análisis utilizando la notación más cómoda para el investigador. Admite tanto la notación de Bentler (1980), como la de Jöreskog (1993). En cualquier caso, cualquiera de estos programas tiene la precisión suficiente para efectuar un buen análisis. En nuestro ejemplo hemos utilizado la versión 6.1 de EQS.

El modelo de medida, como hemos visto, es un modelo que satura las posibles relaciones entre las variables latentes estableciendo covarianzas entre todas ellas. El modelo teórico, es mucho más parsimonioso, es decir, espera llegar a los mismos resultados con muchas menos relaciones. El modelo teórico tendrá validez nomológica (*nomological validity*) si, efectivamente, no hay diferencias significativas entre los ajustes del modelo de medida y el teórico, dado que las escalas habrán sido capaces de establecer relaciones predictivas de otras variables tan sustantivas que, siendo menos, igualan la bondad del modelo.

Las tablas 12 y 13, nos muestran los estadísticos de bondad de ajuste que son necesarios para determinar la validez nomológica. Es posible comprobar que las respectivas Chi cuadrado del modelo teórico y de medida (que obtuvimos al analizar la validez discriminante) son:

TABLA 12. Bondad de ajuste del modelo teórico inicial.

GOODNESS OF FIT SUMMARY FOR METHOD - ML			
INDEPENDENCE MODEL CHI-SQUARE	=	110.190	ON 15 DEGREES OF FREEDOM
INDEPENDENCE AIC =	80.190	INDEPENDENCE CAIC =	36.509
MODEL AIC =	-4.301	MODEL CAIC =	-27.597
CHI-SQUARE =	11.699	BASED ON	8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS .16515			
THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS			10.841.
FIT INDICES			

BENTLER-BONETT	NORMED FIT INDEX =	.894	
BENTLER-BONETT	NON-NORMED FIT INDEX =	.927	
COMPARATIVE	FIT INDEX (CFI)	= .961	
BOLLEN'S	(IFI) FIT INDEX =	.964	
MCDONALD'S	(MFI) FIT INDEX =	.964	
JORESKOG-SORBOM'S	GFI FIT INDEX =	.931	
JORESKOG-SORBOM'S	AGFI FIT INDEX =	.820	
ROOT MEAN-SQUARE	RESIDUAL (RMR) =	.165	
STANDARDIZED	RMR =	.081	
ROOT MEAN-SQUARE	ERROR OF APPROXIMATION (RMSEA) =	.097	
90% CONFIDENCE	INTERVAL OF RMSEA (.000,	.206)

Fuente: EQS. 6.2, con datos propios.

TABLA 13. Bondad de ajuste del modelo teórico final.

GOODNESS OF FIT SUMMARY FOR METHOD = ML			
INDEPENDENCE MODEL CHI-SQUARE	=	66.051 ON	10 DEGREES OF FREEDOM
INDEPENDENCE AIC =	46.051	INDEPENDENCE CAIC =	16.931
MODEL AIC =	-2.897	MODEL CAIC =	-14.545
CHI-SQUARE =	5.103	BASED ON	4 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS			.27691
THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS			4.949.
FIT INDICES			

BENTLER-BONETT	NORMED FIT INDEX =		.923
BENTLER-BONETT	NON-NORMED FIT INDEX =		.951
COMPARATIVE FIT INDEX (CFI)	=		.980
BOLLEN'S	(IFI) FIT INDEX =		.982
MCDONALD'S	(MFI) FIT INDEX =		.989
JORESKOG-SORBOM'S	GFI FIT INDEX =		.961
JORESKOG-SORBOM'S	AGFI FIT INDEX =		.854
ROOT MEAN-SQUARE RESIDUAL (RMR)	=		.093
STANDARDIZED RMR	=		.050
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)	=		.075
90% CONFIDENCE INTERVAL OF RMSEA	(.000,	.237)

Fuente: EQS. 6.2, con datos propios.

Interpretación

La tabla 13 nos muestra los estadísticos de bondad de ajuste que son necesarios para determinar la validez nomológica. Se puede comprobar que las respectivas Chi cuadrado de los modelos teórico y de medida (que obtuvimos al analizar la validez discriminante (*discriminant validity*)) son:

Teórico: 11.699;
 Grados de libertad: 8;
 Medida: 5.103;
 Grados de libertad: 4

La validez nomológica (nomological validity) se establecería calculando un test de diferencias entre las Chi cuadrado, tal como hicimos para la validez discriminante (*discriminant validity*):

Diferencias entre las Chi cuadrado: 6.596;
 Diferencias en grados de libertad: 4;
 Valor crítico $p < 0.01$: 13.277

En otras palabras, el modelo teórico ofrece un ajuste significativamente mejor que el de medida (diferencia inferior al valor crítico), lo que permitiría constatar la validez nomológica de las escalas.

Respuesta socialmente deseable

La respuesta socialmente deseable (SDR. *Socially desirable responding*) es un tema complejo que los psicólogos han discutido durante años. En este apartado, se plantea la cuestión de reiterar su importancia y recordar a los lectores que la SDR justifica la consideración en investigación, en particular cuando el potencial de sesgo de respuesta afecta las relaciones entre constructos. En pocas palabras, el SDR puede ser visto como un estilo de respuesta o sesgo que refleja tendencias para proporcionar respuestas favorables con respecto a normas y prácticas (Nederhof, 1985). Mick (1996) definió la respuesta socialmente deseable como la tendencia de los individuos a hacerse ver bien con respecto a las normas culturales al responder a las preguntas de los investigadores. Como se discute a continuación, este aspecto del SDR es consistente con el concepto de *gestión de impresiones (impression management)* de Paulhus (1993), que resalta los intentos de los encuestados de dar forma a sus respuestas para reflejar una imagen más positiva. SDR puede afectar la medición de constructos, así como las relaciones entre ellos (Mick, 1996: 109-110). Es decir, el SDR puede aumentar las relaciones de tal manera que las correlaciones entre los constructos se deban a la variación compartida de los SDR. Este fenómeno se denomina “efecto espuria”. En el “efecto de supresión”, la correlación verdadera entre dos medidas está enmascarada por SDR. Un tercer efecto posible de SDR ocurre cuando la forma de la relación entre dos variables medidas se afecta. En estas últimas situaciones, SDR modera la relación entre los constructos. Mick (1996) y Ganster, Hennessey y Luthans (1983) describen los procedimientos para investigar estos problemas alternativos asociados con el sesgo de respuesta. Los métodos para hacer frente al sesgo SDR fueron resumidos recientemente por Tian, Bearden y Manning (2002). Los procedimientos se encuadran en dos categorías: métodos diseñados para evitar que los participantes en la encuesta respondan de una manera socialmente deseable y métodos para detectar y medir el sesgo de respuesta a la deseabilidad social (Nederhof, 1985; Paulhus, 1991).

Las técnicas empleadas por los investigadores para prevenir o disminuir la respuesta socialmente deseable en la investigación por encuestas, incluyen el uso de preguntas neutrales, ítems de elección forzada, la técnica de respuesta aleatoria, cuestionamiento indirecto y autoadministración del cuestionario.

- Las preguntas neutras, los elementos de elección forzada y la técnica de respuesta aleatoria no han producido pruebas convincentes de su efectividad para reducir la respuesta socialmente deseable en la investigación general en ciencias sociales (Nederhof, 1985).
- Los artículos de elección forzada, para los cuales se hace que los encuestados escojan entre dos ítems que sean aproximadamente equivalentes en su deseabilidad social pero que estén relacionados con diferentes temas, son problemáticos porque algunos individuos pueden percibir todavía una diferencia en la deseabilidad de los ítems (Nederhof, 1985).
- Los encuestados pueden ser antagonizados por el uso de técnicas de respuesta al azar, por lo que los encuestados al responder a uno de los dos elementos seleccionados al azar, con el investigador no saben qué tema se respondió.
- El cuestionamiento indirecto, una técnica proyectiva en la que se pide a los encuestados que respondan a preguntas estructuradas desde la perspectiva de otra persona o grupo, ha sido demostrado empíricamente como una técnica eficaz para reducir el sesgo de la respuesta de deseabilidad social (Fisher, 1993).
- La autoadministración en las encuestas por correo y las administraciones de masas anónimas tienden a disminuir la respuesta socialmente deseable en las investigaciones que involucran la autodescripción de los individuos (Nederhof, 1985; Paulhus, 1984). Sin embargo, la autoadministración y las técnicas relacionadas que proporcionan anonimato reducen sólo la respuesta socialmente deseable que proviene de la “gestión de impresiones”, no reducen el sesgo auto-engañoso (Nederhof, 1985).

La segunda categoría de métodos implica el uso de medidas directas de respuesta socialmente deseable para evaluar las tendencias para que los individuos se presenten favorablemente con respecto a normas y estándares. La medición de la respuesta socialmente deseable permite evaluar hasta qué punto la tendencia a informar excesivamente los comportamientos deseables o subnotificar los comportamientos indeseables, confunde la evaluación precisa de las variables de contenido y suprime, infla o modera las relaciones variables (Ganster *et al.*, 1983; Zerbe y Paulhus, 1987). Además, las medidas directas permiten investigar la respuesta socialmente deseable como una constructo psicológico por derecho propio (Mick, 1996, y Paulhus, 1991). Aunque se han desarrollado una variedad de medidas de respuesta socialmente deseables, éstas varían en su capacidad para capturar de manera independiente las dimensiones de la “gestión de impresiones” y el “autoengaño” de las respuestas socialmente deseables y en su aplicabilidad al estudio de varias poblaciones de encuestados (Paulhus, 1991).

Resumen

Este capítulo trató cuestiones relacionadas con la validez de constructo, aplicada al desarrollo de medidas para conceptos no observables. En general, la validez de constructo representa la medida en que una medida operacional refleja verdaderamente el concepto que se está investigando o la medida en que las variables operacionales utilizadas para observar la covariación en y entre los constructos se puede interpretar en términos teóricas. La validez de constructo de una medida no se evalúa directamente sino que se deduce de la evidencia de que las puntuaciones sustantivas de la medición que se comporten como se esperaba, con la evidencia procedente de pruebas teóricamente derivadas y de la calidad de los procedimientos empleados en el desarrollo y validación de la medida. Los diversos tipos o fuentes de evidencia sobre la validez de las medidas destinadas a evaluar la validez de constructo discutidas en el capítulo fueron las siguientes:

- Validez de contenido (*content validity*): Grado en que los elementos de un instrumento de medición son relevantes y representativos del constructo objetivo para un propósito de evaluación particular. Las garantías de validez de contenido se basan en la generación teórica a priori, la generación de artículos y el juicio. Validez de la cara: una evaluación de que los ítems de una escala miden adecuadamente la construcción.
- Validez aparente (*face validity*): Se puede juzgar después de que se haya desarrollado una medición, a menudo, antes de la aplicación en otro estudio, por los usuarios potenciales de la medición.
- Validez predictiva (*predictive validity*): Capacidad de una medida para predecir con eficacia algún criterio subsiguiente y temporalmente ordenado.
- Validez concurrente (*concurrent validity*): para la cual la evidencia es proporcionada por las correlaciones considerables entre la medida del constructo bajo desarrollo y una medida del criterio recogida simultáneamente o “concomitantemente”.
- Validez convergente (*convergent validity*): La medición en que las medidas independientes del mismo constructo convergen o están altamente correlacionadas.
- Validez discriminante (*discriminant validity*): El nivel en que las mediciones divergen de otras operacionalizaciones a partir de las cuales el constructo es conceptualmente distinto.
- Validez del grupo conocido (*known-group validity*): la medida en que una medición difiere según lo previsto entre los grupos que deben puntuar bajo y alto en un rasgo.

- Validez nomológica (nomological validity): El grado en que la medida encaja “legalmente” en una red de relaciones o una “red nomológica”; es decir, el grado en que una medida opera dentro de un conjunto de constructos teóricos y sus respectivas medidas. Además de estas fuentes de evidencia o tipos de validez específicos, el capítulo concluyó con una discusión de los problemas asociados con la respuesta socialmente deseable, así como una descripción de los métodos diseñados para evitar que los participantes en la encuesta respondan de una manera sesgada y métodos dirigidos a detectar y medir el sesgo de la respuesta de deseabilidad social (Nederhof, 1985; Paulhus, 1991), (tabla 14).

TABLA 14. Resumen de conceptos de tipos de validez.

Tipos de validez	Concepto
De contenido	Se aplica para estimar el grado con el que una escala es representativa del concepto que mide, o por el grado en que los ítems de la escala son representativos de la población de ítems que desarrollan un concepto. Por su naturaleza cualitativa, no se puede expresar mediante un coeficiente; tan sólo aproximaciones procedentes de juicios.
De construcción o de concepto	Se aplica porque una escala no es un conjunto de cuestiones o ítems inconexos, sino un instrumento diseñado para medir algo concreto. Pretende responder a la pregunta: qué concepto mide la escala. Los procedimientos básicos para probarla son las matrices multiconcepto- multimétodo y el análisis factorial confirmatorio.
Convergente	Se refiere a que dos o más escalas que tratan de medir el mismo concepto deben converger hacia una misma puntuación. La comparación se ha de realizar en el mismo lapso temporal.
Discriminante	Describe el hecho de que una escala sólo cuantifica las características del concepto que trata de medir y no cualquier otra característica de cualquier otro concepto. Se mide mediante la correlación: si las puntuaciones de dos escalas que miden el mismo concepto han de correlacionarse (validez convergente), las escalas usadas para medir conceptos distintos no deben hacerlo.
Nomológica	Trata de determinar si el instrumento de medida se comporta según lo esperado con respecto a otras construcciones con las cuales está teóricamente relacionada. Se evalúa mediante la confirmación de relaciones significativas entre construcciones según establezca la teoría.
Factorial	Se refiere al hecho de que la estructura de relaciones entre ítems y sus correspondientes conceptos o factores se mantiene en sucesivos análisis. Requiere la existencia de validez convergente y discriminante.
Predictiva, pragmática o relacionada con criterios	Se refiere al hecho de que una escala de medida sea capaz de predecir otras variables (criterios); éstas pueden ser cualquier tipo de variable obtenida con otras escalas y que pueden ser estimadas con la escala bajo estudio.

Validez

Tipos de validez	Concepto
Predictiva simple	Es la capacidad de la escala para pronosticar el valor futuro de otras variables (por ejemplo, intención de compra). Por lo tanto, los datos de la escala bajo estudio y de los criterios se recogen en momentos diferentes del tiempo. Para tener validez predictiva es necesario tenerla también de construcción: si algo no se puede medir con precisión, difícilmente se podrán hacer predicciones válidas.
Concurrente	Se refiere a la capacidad de la escala para predecir otras variables o criterios, contrastándose simultáneamente los datos recogidos del concepto bajo análisis y de la variable criterio. Normalmente, la evaluación de la validez concurrente se lleva a cabo incluyendo las medidas de las variables criterio en el mismo cuestionario donde se recogen los datos de la escala bajo estudio, por lo tanto, en el mismo momento del tiempo.

Fuente: Netemeyer *et al.* (2003).

CAPÍTULO 5

Definiendo el constructo y su contenido (etapas 1 y 2)

Los capítulos 2, 3 y 4 discuten los principios psicométricos de la dimensionalidad, la confiabilidad y la validez principalmente desde un punto de vista conceptual, así como también ofrecen ejemplos de algunas escalas desarrolladas más recientemente. Los capítulos 5, 6 y 7 ahora ofrecerán aplicaciones de estos principios desde un punto de vista empírico. Estos capítulos cubrirán los pasos y procedimientos recomendados en el desarrollo de la escala y la validación.

La primera parte de este capítulo analiza el paso crítico de la definición de constructo y la delimitación del dominio de contenido del mismo. Se resaltan así, los siguientes problemas:

- ▶ El papel de la teoría en la definición del constructo y el dominio de su contenido,
- ▶ La importancia de una revisión exhaustiva de la literatura y la definición a juzgar por expertos e individuos de poblaciones relevantes,
- ▶ El enfoque en los indicadores de efectos (reflexivos), y
- ▶ La importancia de una dimensionalidad, a priori, de constructo.

La segunda parte de este capítulo trata sobre el segundo paso en el desarrollo de escala: determinar y analizar un conjunto inicial de indicadores que soporten al constructo. Con este paso, cubriremos brevemente los supuestos teóricos del muestreo de dominios. También se discutirán temas en la generación de un grupo de indicadores, opciones de redacción de preguntas/declaraciones, formatos de elección de respuestas y evaluación de los indicadores de contenido y validez de aparente (*face validity*) por parte de expertos y encuestados potenciales de poblaciones relevantes.

La última parte de este capítulo usa ejemplos que ilustran y aplican muchos de los procedimientos recomendados para los dos primeros pasos en el desarrollo de la escala.

ETAPA 1

El CONSTRUCTO

La importancia de la definición de un constructo claro y preciso

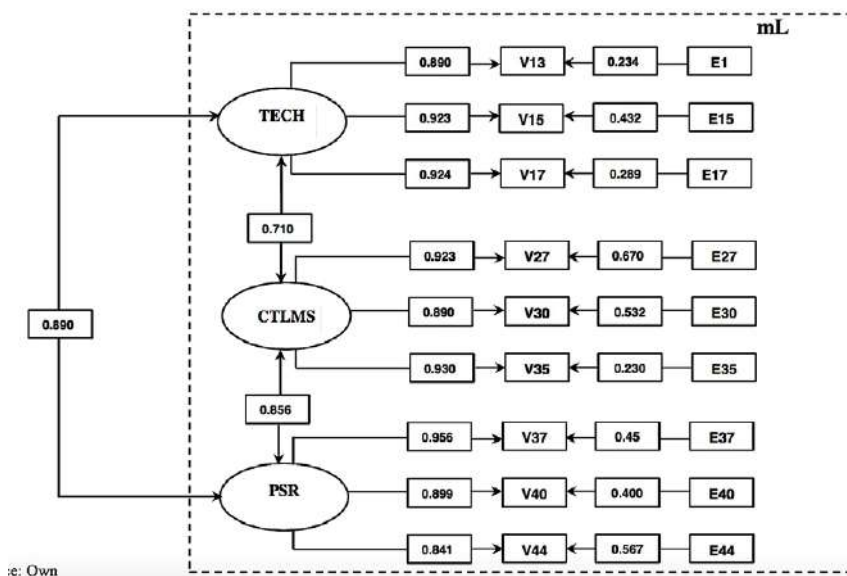
La importancia de un constructo bien definido no puede exagerarse, ya que la validez de lo que se mide se basará en gran parte en su definición y dominio de contenido. Definir claramente el constructo -sus facetas y dominios- es un primer paso esencial que muchos consideran el paso más difícil en el proceso de escalamiento (Churchill, 1979; Haynes et al., 1999; Haynes et al., 1995; Nunnally y Bernstein, 1994). Al definir los constructos, se debe tener cuidado en lo que se incluye y excluye en el dominio del constructo ya que las fuentes de invalidez pueden tener sus orígenes en el proceso de definición del mismo. Una medición puede ser demasiado estrecha y no incluir facetas importantes. Esto se ha referido como insuficiente representación del constructo (Messick, 1993). Si el dominio está demasiado definido, se pueden incluir factores extraños o dominios de constructos que no sean el objetivo. Aunque estos dominios extraños pueden estar altamente correlacionados con los dominios relevantes del constructo, crean lo que se ha denominado como constructo de varianza irrelevante (*construct-irrelevant variance*), que a su vez puede crear confusiones entre la constructo objetivo y sus variables de predictor y criterios (Neuberg et al., 1997). La razón detrás del dominio de una constructo es que la interpretación de cualquier medición es más clara cuando el dominio es claro, o sea, inconfundible. Cuando se incluyen factores extraños, posiblemente se sugiera que más de un rasgo/o variable de diferencia individual está siendo evaluada, por lo que la correlación de ese constructo con un criterio puede ser confundida. En resumen, cuando se incluyen factores extraños o dominios de otras constructos, más de una constructo subyace al puntaje total y la validez del constructo está amenazada. Por lo tanto, se considera ampliamente que el factor clave para el aspecto del contenido de la validez del constructo es la clara especificación de los límites del dominio del constructo que va a ser evaluado.

La importancia de una teoría sólida, revisión de la literatura y su análisis

En sus trabajos clásicos sobre medición y validez, Cronbach y Meehl (1955) y Loevinger (1957) declararon la importancia de la teoría en la medición. Para que las medidas de los constructos latentes tengan relevancia en las ciencias de la administración, los constructos deben basarse en un marco teórico. Incluso los

constructos altamente compactos basados en la teoría, son más útiles como antecedentes o consecuencias de otros comportamientos latentes cuando se insertan en la teoría. Como tal, la relevancia de un constructo latente para las ciencias de la administración, depende en parte de las teorías en las que está formulada: “¿qué es lo que hace que el constructo subyacente haga predicciones?; ¿qué predice el constructo subyacente?” Esto se le conoce como como la red nomológica de un constructo latente. Un buen ejemplo es el mostrado por Mejía-Trejo *et al.* (2015). (figura 1).

FIGURA 1. Modelo de 1 factor con relación total de las variables.



Fuente: Mejía-Trejo *et al.* (2015).

Al desarrollar su medida de aprendizaje móvil (*mobile learning*), basado en rasgos (atributos o variables) como la tecnología (TECH), la administración de contenidos y estilos de enseñanza aprendizaje (CTLMS. Contents teaching-learning management and styles); y el rol del profesor y estudiante (PSR. professor and student role). En éstas de postularon dimensiones e indicadores sumando 60, algunas de ellas como: la *facilidad del uso de la tecnología*, *las redes sociales*, *la percepción profesor-estudiante del valor-costo*, *las políticas de uso*, etc. Lo anterior constituye una red de constructos que conforman al modelo, que es fundamental para guiar el desarrollo de la escala y evaluar la validez.

En esencia, no puede haber validez de constructo de una medición, sin una

teoría bien especificada. Esta teoría bien pensada comienza con la conceptualización/definición del constructo basada en una revisión exhaustiva de la literatura. Tal revisión de la literatura idealmente sirve para varios propósitos importantes (ver el apartado Importancia de la revisión de la literatura del estado del arte en el marco teórico, capítulo 1) ya que:

1. Una revisión de la literatura debe alertar al investigador sobre las teorías en las que el constructo puede resultar útil como una variable independiente o dependiente. Un manejo más preciso de lo que es el constructo, sus límites, dimensiones y dominio de contenido, se puede descubrir a través de una revisión de la literatura. Una revisión de la literatura también puede revelar intentos previos para medir el constructo y las fortalezas y debilidades de tales intentos. Los problemas con los intentos previos de medir el constructo se pueden descubrir y, por lo tanto, evitarse en el propio esfuerzo.
2. Una revisión exhaustiva puede descubrir importantes constructos relacionados y mediciones que se usarán en la validación convergente y/o discriminante.
3. Finalmente, dado que el desarrollo y la validación de la escala es una tarea que lleva tiempo y, en ocasiones, es costosa, una revisión exhaustiva de la literatura debe ayudar a responder la siguiente pregunta: ¿se necesita una escala? Si ya existen buenas medidas de un constructo, el valor de una nueva medición puede ser pequeño en relación con los costos involucrados en su desarrollo. Una nueva medida, para ser útil, debe mostrar alguna ventaja teórica o empírica sobre una medida existente del mismo constructo. Para que una nueva escala tenga validez incremental sobre las mediciones existentes, debe capturar el constructo específico con más precisión o de manera más eficiente (por ejemplo, es más corto, más barato, más fácil de usar o más fácil de responder) que las medidas existentes (Clark y Watson, 1995; Haynes *et al.*, 1999). Otra forma útil de mejorar la precisión y la exhaustividad de la definición del constructo es la revisión por parte de expertos. Aunque tales revisiones tradicionalmente se han asociado más con la evaluación de los ítems, pueden ayudar a delimitar las facetas del constructo antes de la generación de indicadores.

El enfoque de indicadores reflectivos (de efecto) vs. indicadores formativos (de causa)

Como se dijo en el capítulo 1, este texto se centra en las medidas que se consideran indicadores de efecto (*reflexivos o reflectivos*) de los constructos más que indicadores formativos (*causales o formativos*). Para los indicadores de efecto, las

puntuaciones sobre los propios indicadores, son hipotéticos para ser manejados por el constructo latente. Es decir, las diferencias individuales en el constructo cuentan para las diferencias individuales en la aprobación del indicador. Con los indicadores formativos, las diferencias individuales en los indicadores cuentan para las diferencias individuales en la aprobación del constructo. La siguiente discusión, permitirá entenderlo mejor:

El constructo “estrés de la vida en el último año” puede ser interpretado como un constructo de indicador formativo, en el cual los indicadores son diferentes fuentes de estrés, incluyendo un cambio de trabajo, la muerte de un ser querido, el nacimiento de un niño y una enfermedad. Es contraintuitivo que los cambios en el estrés de la vida produzcan la muerte de un ser querido o el nacimiento de un niño; Lo opuesto tiene sentido (Smith y McCarthy, 1995). Asimismo, como segundo ejemplo, el estatus socioeconómico (SES) se conceptualiza mejor como un constructo de indicador formativo. Las personas tienen SES alto debido a su riqueza o educación; Ellos no se vuelven ricos ni educados como resultado del alto SES (Nunnally y Bernstein, 1994). Por lo tanto, los indicadores “*forman*” el puntaje del constructo. Otras dos diferencias importantes entre el efecto (reflectivos o *reflectives* y los indicadores formativos (causales o *normatives*) se encuentran tanto en las áreas conceptuales como en las metodológicas, ya que:

- ▶ Para los indicadores que se conceptualizan para formar la constructo latente, se cree que la lista de indicadores debe ser completa. Es decir, con indicadores formativos, se necesita un censo de indicadores, no solo una muestra, para medir completamente el constructo. La exclusión de cualquier indicador cambia la composición de la constructo latente y por lo tanto, cada indicador es importante para la validez del constructo (Bollen y Lennox, 1991; MacCallum y Browne, 1993; Neuberger et al., 1997). Este no es necesariamente el caso con los indicadores de efecto (reflectivos o *reflectives*). Con éstos, los indicadores deben representar una muestra razonable de elementos que tocan el dominio del constructo (Nunnally y Bernstein, 1994).
- ▶ El hecho de que los indicadores de un constructo formativo (causal o *formatives*) se combinen para producir un índice general no implica necesariamente que todos los puntajes del indicador individual estén intercorrelacionados, y si están o no correlacionados es irrelevante para la confiabilidad de la medida (Bollen y Lennox, 1991; Smith y McCarthy, 1995). Los indicadores formativos (causal o *formatives*) no necesitan ser consistentes internamente, por lo que los métodos de confiabilidad basados en la consistencia interna no se aplican.
- ▶ Con los indicadores de efecto (reflectivos o *reflectives*), la interrelación entre ítems y, por lo tanto, la consistencia interna, es motivo de preocupación para la

confiabilidad de la medida. No solo existen diferencias conceptuales entre los indicadores formativos (causal o *formatives*) y los de efecto (reflectivos o *reflectives*), también difieren los métodos utilizados para desarrollar tales mediciones. (Para una excelente revisión de estos métodos, vea Diamontopoulos y Winklhofer, 2001.) Aquí nuevamente, la importancia de una definición de constructiva bien definida y pensada es muy útil. La teoría y una revisión exhaustiva de la literatura pueden ayudar a determinar si el constructo es una medida de indicador formativa (causal o *formatives* o de efecto (reflectivos o *reflectives*).

El papel de la teoría al especificar la dimensionalidad

Como se señaló en los capítulos 1 y 2, la dimensionalidad se refiere a la homogeneidad de los ítems o indicadores. La unidimensionalidad puede definirse como la existencia de un rasgo, atributo, variable o constructo latente subyacente a un conjunto de indicadores/medidas (Hattie, 1985). Dado que el dominio de contenido de un constructo se puede hipotetizar como unidimensional o multidimensional, su estructura empírica debe reflejar su dimensionalidad hipotética. Para un solo constructo unidimensional, cada elemento/indicador se refleja por su constructo latente, o factor de primer orden. Dicho conjunto de elementos también se ha denominado una “medición congénere” (*congeneric measure*), es decir, una por la que una sola dimensión subyace al conjunto de indicadores. Cuando un constructo está compuesto por múltiples facetas que están relacionadas, pero distintas, el constructo puede clasificarse como multidimensional.

La escala de confianza del consumidor de seis factores de Bearden *et al.* (2001) (*the six factor consumer self-confidence scale*) representa una escala multidimensional. Por el contrario, Lastovicka *et al.* (1999) La escala de la frugalidad del consumidor (*the consumer frugality scale*), las de aprendizaje móvil e innovación de Mejía-Trejo *et al.*, 2015, así como conocimiento del consumidor e innovación de Mejía-Trejo *et al.*, 2016 (*mobile learning and innovation customer knowledge management and innovation*) son unidimensionales.

Con constructos multidimensionales, cada dimensión se considera un factor de primer orden representado por una variable latente separada, donde cada indicador se usa para operacionalizar su respectiva dimensión hipotética. Dentro de sus respectivas dimensiones, los indicadores deben mostrar evidencia de unidimensionalidad (es decir, cualidades congéneres -*congeneric qualities*- Se dice que un conjunto de medidas es “congénere” si cada medida en el conjunto pretende evaluar el mismo constructo, excepto los errores de medición (Jöreskog, 1993). En la medida en que varios factores de primer orden incluyan una variación común,

un factor de orden superior (también llamado factor de segundo orden) podría ser conceptualizado para capturar la varianza compartida a través de los factores de primer orden. Los factores de primer orden se consideran indicadores de efecto (reflectivos o *reflectives*) o dimensiones de constructo de orden superior, y el constructo de mayor orden cuenta para la covarianza (correlación) entre los factores de primer orden. Un constructo de orden superior sugiere que las dimensiones (factores de primer orden) miden el mismo concepto jerárquico, a excepción del error aleatorio y la especificidad de medición (Bagozzi y Heatherton, 1994). La importancia de especificar teóricamente y probar empíricamente la dimensionalidad antes de otras propiedades psicométricas (por ejemplo, consistencia interna y validez nomológica) no debería subestimarse (Cortina, 1993; Gerbing y Anderson, 1988; Hattie, 1985). Para operacionalizar los constructos latentes, los investigadores a menudo usan puntajes compuestos, sumando o promediando entre elementos diseñados para medir el constructo de interés. El cálculo y el uso de tales puntuaciones son significativos sólo si los ítems tienen una unidimensionalidad aceptable. Como se señaló en el capítulo 3, las escalas multidimensionales, cuando se usan como si fueran unidimensional (es decir, compuestos de elementos sumados o promedios), pueden dar lugar a ambigüedades interpretativas de las relaciones entre constructos en una prueba de teoría. Si una construcción es multidimensional pero todos los puntajes de los ítems se suman/promedian entre dimensiones en un solo puntaje compuesto y se correlacionan con una cierta variable criterio, dicha correlación es ambigua (Neuberg *et al.*, 1997). Aquí nuevamente, el papel de la teoría y una revisión exhaustiva de la literatura deberían ayudar a especificar *a priori* la dimensionalidad del constructo de interés.

ETAPA 2

GENERANDO Y ANALIZANDO LOS INDICADORES DE CONTENIDO

La muestra y su contenido

Una vez que el constructo se ha definido y delimitado con precisión, comienza la tarea de generar indicadores del dominio del constructo. Como modelo de medición de error, se sugiere que una medición se componga de una muestra de indicadores de un gran dominio hipotético de elementos. Existe una gran agrupación (“pool”) de indicadores o elementos que podrían aprovechar el dominio de la constructo (y/o) para llegar a una medición de escala final, por lo que debe

seleccionarse una muestra de elementos de este dominio con propiedades psicométricas deseables, que:

- En teoría, este grupo se extrae de lo que algunos han denominado un “universo” o pool de indicadores, pudiendo basarse en expertos.
- En la práctica, los indicadores para un grupo inicial se extraen de la literatura existente o generados por los autores de la escala.

Los expertos del tema, también pueden contribuir con indicadores y variables, Siempre que el grupo de indicadores inicial sea lo suficientemente grande en relación con el dominio de contenido del constructo, los principios de muestreo del mismo todavía son aplicables (Nunnally y Bernstein, 1994).

Por otro lado, la suficiencia de contenido existe cuando:

- El contenido de la medición es una muestra aleatoria representativa del dominio a medir,
- Los indicadores finales elegidos (por ejemplo, la muestra de ítems) son considerados representativos del constructo, y cuando
- Tienen las mismas medias y varianzas que los de la totalidad del conjunto del pool de indicadores.

Como se señaló en el capítulo 4, la validez del contenido (content validity) es relevante para el muestreo de dominios porque representa parcialmente el grado en que uno puede generalizar de un conjunto determinado de indicadores en el dominio del constructo.

El error de medición considerado con el muestreo de dominio, entonces, es el error producido al usar una muestra de los indicadores en lugar de todo el dominio de ellos.

Generando el pool de indicadores

Al generar un conjunto de indicadores, un objetivo importante es muestrear sistemáticamente todas las áreas de contenido del constructo. Se deben tener en cuenta varias cuestiones relevantes para la validez de versión (*translation validity*), ya que:

1. Los indicadores de la escala generados deben aprovechar el dominio de contenido del constructo y mostrar la validez del contenido (*content validity*). Como se discutió en el Capítulo 4, este término se ha definido de muchas maneras, y la mayoría de las definiciones enfatizan que los indicadores de una medición son una muestra apropiada del dominio teórico de la constructo (Nunnally y Bernstein, 1994; Peter, 1981). La mayoría de las definiciones son

consistentes en que la validez del contenido (*content validity*) refleja “el grado en que los elementos de un instrumento de medición son relevantes y representativos del constructo específico para un propósito particular de evaluación” (Haynes et al., 1995: 238). Los elementos se refieren al contenido individual de los indicadores, formatos e instrucciones de respuesta para los encuestados. La representatividad se refiere al grado en que los elementos son proporcionales a las facetas (dominios) del constructo objetivo y al grado en que todo el dominio del constructo ha sido muestreado. Es decir, los elementos deberán aparecer coherentes con el dominio teórico del constructo en todos los aspectos, incluidos los formatos de respuesta y las instrucciones.

2. En la generación de indicadores, también se debe considerar la validez aparente (*face validity*). Una escala de alta validez aparente mejora la cooperación de los encuestados debido a su facilidad de uso, nivel de lectura adecuado y claridad, así como a sus instrucciones y formatos de respuesta. Por lo tanto, desde una perspectiva práctica, la validez aparente (*face validity*) puede estar más referida con lo que los entrevistados de poblaciones relevantes infieren con respecto a lo que se está midiendo, y la validez de contenido (*content validity*) es referida con la validez aparente (*face validity*), así como a lo que el investigador cree que está construyendo (Haynes et al., 1995; Nunnally y Bernstein, 1994).
3. Incluso, enfocados en ambas validaciones, aparente (*face validity*) y de contenido (*content validity*), se deben considerar otros dos aspectos en la construcción del pool de indicadores. Clark y Watson (1995) abogaron por que el desarrollador de la escala vaya más allá de su propia visión al generar una agrupación de indicadores inicial que contenga, en última instancia, los que sólo estarán relacionados tangencialmente. Por lo tanto, es mejor incluir indicadores en exceso al dominio del constructo en lugar de hacerlo en carestía o cortedad. También se debe tener cuidado para garantizar que cada área de contenido del constructo tenga una muestra adecuada de indicadores. Aunque es difícil de lograr en la práctica, las áreas de contenido más amplias deben estar representadas por un conjunto de ítems más grande. Con estos problemas en mente, la generación de indicadores puede comenzar con una reflexión cuidadosa sobre:
 - a. Cuál debería ser la fuente de los dichos indicadores,
 - b. Problemas de redacción de su redacción, y
 - c. Cuántos artículos deben servir como grupo inicial. Para estos problemas, las respuestas definitivas no existen, pero algunas pautas prácticas son evidentes.

Fuentes de indicadores

En este término, se sugiere que los desarrolladores de escala vayan más allá de su vista del constructo objetivo y consulten varias fuentes. Cabe destacar, que muchas escalas existentes tienen ítems con validez de contenido (*content validity*) y validez aparente (*face validity*) que aprovechan uno de los dominios del constructo. Observar y analizar cómo los estudios previos han operacionalizado el constructo (o constructos relacionados), puede ser una fuente valiosa en la generación de indicadores. Por ejemplo, para medir el conflicto trabajo-familia (WFC. *work-family conflict*) y el conflicto familia-trabajo (FWC. *family-work conflict*) Netemeyer et al. (1996) seleccionaron indicadores de las mediciones existentes de WFC, así como mediciones de *sobrecarga de roles de trabajo* (*work role overload*), y Bearden et al. (2001) convirtió indicadores de varios estudios de “confianza/experiencia del consumidor” (*consumer confidence/expertise*) para su pool inicial de indicadores en evaluar las dimensiones de autoconfianza del consumidor (*consumer self-confidence*).

Otra fuente de generación de ítems es la población de interés (Haynes et al., 1995; Haynes et al., 1999). Aquellos que pueden estar respondiendo a la escala (es decir, miembros de la población) pueden ofrecer hallazgos de lo que podría ser el constructo y cómo medirlo, mejorando así la validez aparente (*face validity*).

Las entrevistas estructuradas y no estructuradas con miembros de la población pueden proporcionar información sobre la redacción de los ítems y los formatos de respuesta. Varias escalas desarrolladas en la literatura de administración de la mercadotecnia, han utilizado este enfoque tanto para ayudar a definir el constructo como para generar ítems (por ejemplo, Bearden et al., 2001; Lastovicka et al., 1999, Mejía-Trejo et al., 2016).

Dos fuentes más de indicadores son: los expertos en el campo el desarrollador de escalas. Los académicos, investigadores, profesionales (consultores) familiarizados con los constructos objetivo, pueden ser excelentes fuentes para indicadores. Los expertos en el campo pueden sugerir, escribir o modificar indicadores coherentes. Obviamente, el desarrollador de la escala también es una fuente de indicadores. Esto será particularmente cierto para constructos nuevos o para las que se dispone de muy poca literatura existente. Incluso con otras fuentes que contribuyen al pool inicial de indicadores, la mayoría de los desarrolladores de escala tendrán que escribir varios de sus propios elementos para generar un grupo inicial adecuado.

Escritura de los indicadores

Una segunda consideración en la generación de indicadores implica la escritura de los mismos. Se deben considerar varias cuestiones:

- La claridad de la redacción,
- La redacción redundante,
- La redacción en sentido positivo o negativo, y
- La elección de formatos de respuesta.

Claridad de la redacción

La claridad del indicador es sumamente importante. De hecho, Nunnally y Bernstein (1994) sugieren que la regla de oro de la escritura del indicador se resume en una sola palabra: “claridad”. Un buen indicador debe ser inequívoco hasta el punto de que todos los encuestados comprendan su significado de la misma manera. Se aplican algunas reglas simples, como:

- a. Use un lenguaje común al nivel de lectura de la población objetivo.
- b. Mantenga el indicador corto y simple independientemente del nivel de lectura de la población de interés. Aunque no se quiere sacrificar el significado de un elemento por razones de brevedad, la longitud y la complejidad tenderán a disminuir la claridad de los elementos.
- c. Evita la jerga o el argot de moda. El significado de dicho idioma está fechado rápidamente o está sujeto a una interpretación diferente a medida que pasa el tiempo.
- d. Evite las declaraciones de doble contenido que, de hecho, abordan dos problemas (o áreas de contenido) en una declaración.
- e. Escriba indicadores que todos se adhieran ya que de lo contrario, serán de poca utilidad. Tales indicadores producen una pequeña variación de ítems y, por lo tanto, una pequeña variación de escala. Como lo señalaron Clark y Watson (1995), indicadores que todos respaldan de la misma manera de manera positiva (por ejemplo, “a veces, soy más feliz que otras ocasiones”) o de la misma manera, en negativo (por ejemplo, “siempre estoy furioso”) agrega poco a la validez de contenido (*content validity*) de un constructo.

La redacción redundante

En la fase de generación de indicadores, se necesita cierta redundancia de redacción entre ellos, acorde con el contexto que aborda. Sin embargo, el desarrollador

de la escala debe ser consciente, sin embargo, que la redundancia que es inútil. Crear un nuevo indicador cambiando solo una palabra que no es crítica para el significado del mismo, al no ser importante para el contenido del indicador, refleja una redundancia inútil.

Por otro lado, variando la selección de palabras y la estructura gramatical para crear nuevos indicadores, de modo que el dominio de contenido de la constructo se esté aprovechando de manera diferente, refleja una redundancia útil.

DeVellis (1991: 56) ofrece un buen ejemplo para el aprovechamiento del “grado en que un padre ayudará a su hijo”, siendo los los indicadores propuestos:

“Haré casi cualquier cosa para asegurar el éxito de mi hijo” y

“Ningún sacrificio es demasiado grande si ayuda a mi hijo a alcanzar el éxito” tienen redundancia útil porque ambas declaraciones expresan el contenido del constructo de diferentes maneras. Sin embargo, al evaluar la actitud hacia los amantes de las mascotas, las declaraciones “En mi opinión, los amantes de las mascotas son amables” y “Yo estimo, que los amantes de las mascotas son amables” comparten vocabulario y gramática comunes que no aprovechan el constructo, lo que refleja una redundancia inútil. Sin embargo, en general, en la fase de generación de indicadores del desarrollo de la escala, es necesaria cierta redundancia de los mismos, pero no debe reflejar diferencias puramente triviales de redacción/gramática.

En la escala final, cuando la redacción de los indicadores es demasiado similar y refleja solo diferencias triviales, algunas propiedades psicométricas en realidad pueden estar comprometidas. Los indicadores que son redactados de manera muy similar aumentarán la media de correlación interitem, lo que de hecho aumenta la consistencia interna sin contribuir sustancialmente a la validez de contenido de la medida. Estos indicadores, también contribuyen a errores de medición correlacionados en el análisis factorial confirmatorio que, de hecho, amenazan la dimensionalidad de una medición (Floyd y Widaman, 1995). Aunque se necesita cierta similitud entre los indicadores de una escala para aprovechar el dominio, los que tengan ligeras variaciones producen escasa información relevante para el constructo, que cualquier indicador individual (Clark y Watson, 1995). Dado lo anterior, los indicadores con redundancia inútil pueden contribuir a la “*paradoja de atenuación*” (*attenuation paradox*) en la teoría psicométrica, por lo que el aumento de las estimaciones de consistencia interna más allá de un cierto punto, no mejorarán la validez.

La redacción en sentido positivo o negativo

La elección de utilizar todos o algunos de los indicadores redactados positivamente o negativamente, también es de interés. Varios desarrolladores de escala han escrito indicadores que reflejan: niveles bajos de, lo contrario de, o la ausencia del constructo objetivo. El objetivo principal de un procedimiento de este tipo es también “mantener la honestidad del entrevistado” y así evitar el sesgo de respuesta en forma de complacencia, afirmación o tendencia decir sí a todo (*acquiescence, affirmation, yea-saying*). La escala de *necesidad de cognición* (*need for cognition scale*) de Cacioppo y Petty (1982) es un ejemplo. Ellos utilizaron indicadores que reflejan el nivel más alto del constructo, como:

“Prefiero los problemas complejos a los simples”, y los indicadores que reflejan un bajo nivel de constructo, tales como:

“Sólo pienso tan duro como tengo que hacerlo”. Sin embargo, y de acuerdo a la experiencia, se tiene que los indicadores redactados negativamente no muestran una confiabilidad tan alta como los elementos redactados positivamente o incluso, pueden ser confusos para los encuestados. Tales indicadores, también pueden contribuir a métodos de modelos analíticos de factor a factor, porque los indicadores redactados positivamente tienden a cargar altamente a un factor y los elementos redactados negativamente tienden a cargarse altamente en otro factor (véase Herche y Engellend [1996] para una discusión y ejemplos). Con esa posibilidad en mente, el investigador debe sopesar las ventajas y desventajas potenciales de usar indicadores redactados negativamente en el *pool* de indicadores.

La elección de formato de respuesta

Con respecto a los formatos de respuesta, se deben considerar dos cuestiones básicas: a) Puntos de escala dicotómicos vs. multicótomos y b. Redacción de los puntos de escala de respuesta.

Aunque existen muchos tipos de formatos de respuesta, incluidas listas de verificación (*checklists*) y análogos visuales (visual analog), las dos formas dominantes son los formatos de escala dicotómicos y multicótomica. La mayoría de los formatos dicotómicos utilizan puntuación verdadero-falso o sí-no, y la mayoría de los usos multicótomos usan tipo Likert o diferenciales semánticos (o alguna variación de) con escalas con tres o más puntos de escala. Dado el dominio de su uso, nuestra discusión se limita a estos dos tipos de escalas.

Para más información sobre el uso de listas de verificación (*checklists*”, elecciones forzadas (*forced choice*) y escalas análogas visuales (*visual analog scales*), véase DeVellis [1991]; Green, Goldman y Salovey [1993], y Clark y Watson [1995]).

Existen desventajas y ventajas tanto en los formatos dicotómicos como en los multicótomos. Los formatos dicotómicos han sido criticados por:

- Su tendencia a tener distribuciones de respuesta altamente desequilibradas, esto es, todos los individuos responden siempre a “verdaderos” o todos los individuos siempre responden a “falso” (Comrey, 1988). Una evaluación cuidadosa de los indicadores durante la etapa de desarrollo de la escala, puede ayudar a eliminar tales ítems, y el formato dicotómico generalmente lleva menos tiempo para que el entrevistado responda, permitiendo que más ítems sean respondidos en un periodo de tiempo más corto.
- Su relación con la escala de varianza. Una propiedad de medición deseable de una escala es que tenga suficiente varianza en las puntuaciones de la escala global para una muestra dada. Dado que cualquier indicador para una escala dicotómica produce una covarianza limitada con otro indicador, debido al formato binario, la varianza de la escala global será limitada. Para una escala compuesta de indicadores dicotómicos, necesitará un mayor número de ítems para producir una variación de escala similar a la de ítems que usan escalas de tipo Likert y diferenciales semánticas.

Las escalas tipo Likert y los diferenciales semánticos pueden usar una serie de opciones y formatos de respuesta. Las escalas tipo Likert generalmente piden a los encuestados que indiquen su nivel de acuerdo con una declaración declarativa (*diferentes niveles desde totalmente en desacuerdo hasta totalmente de acuerdo en intervalos impares*); el grado en que lo que se expresa en la declaración es el “cierto de una creencia, actitud o característica del entrevistado” o la frecuencia de un comportamiento (es decir: desde nunca hasta siempre). Con las escalas de tipo Likert, también es posible etiquetar cada punto de escala, así como los puntos finales de escala. Por ejemplo, para una escala de 5 puntos, cada punto de escala podría etiquetarse de la siguiente manera: *1 = muy en desacuerdo, 2 = algo en desacuerdo, 3 = no está de acuerdo ni en desacuerdo, 4 = está de acuerdo y 5 = muy de acuerdo*. Dicho etiquetado sirve para dar al encuestado una mejor idea del respaldo que está haciendo. Dependiendo de lo que se está evaluando, los puntos de escala pueden tomar muchas palabras diferentes. Por lo tanto, el investigador debe tener cuidado al elegir etiquetar puntos de escala o puntos finales (Clark y Watson, 1995; Comrey, 1988; DeVellis, 1991; Nunnally y Bernstein, 1994). Un indicador correcto elaborado en escala tipo Likert debe expresar la opinión, la actitud, la creencia, el comportamiento o la característica que se evalúa en términos claros y usar la redacción apropiada para los puntos de escala. Una evaluación cuidadosa de los indicadores puede ser muy útil a este respecto.

Los indicadores diferenciales semánticos utilizan puntos extremos basados en adjetivos que son de naturaleza bipolar (por ejemplo, amistoso-hostil, bueno-malo, frío-calor) o de naturaleza unipolar (por ejemplo, amigable -no amigable, bueno- no bueno, caliente, no caliente) (Osgood y Tannenbaum, 1955). Aquí, el encuestado califica el estímulo (por ejemplo, objeto o persona) en una serie de escalas diferenciales semánticas que evalúan algún atributo latente del estímulo. Tanto para los indicadores de escala tipo Likert como para los diferenciales semánticos, los puntajes en los ítems que constituyen la escala se suman para formar un puntaje general. Una ventaja principal de las escalas multicotómicas es que crean más variaciones de escala en relación con una escala dicotómica con un número similar de indicadores. Sin embargo, con las escalas multicotómicas, se debe tener en cuenta la cantidad de puntos de escala y si se debe usar una cantidad impar o incluso de cuántos intervalos deban ser (3-5-7-9-11). La mayoría de las escalas multicotómicas usan entre 5 y 9 puntos de escala, con algunos tan bajos como 3 y tan altos como 11 puntos de escala. Lo más común es utilizar formatos de 5 o 7 puntos considerados como suficientes, ya que proporcionar más alternativas de respuesta puede no mejorar la confiabilidad o la validez de la escala. Si el investigador quiere proporcionar una “etiqueta” para cada escala, es más fácil y probablemente más significativo, tanto para el desarrollador de la escala como para el encuestado, si se usan formatos de 5 o 7 puntos. Más alternativas pueden requerir más esfuerzo en favor del demandado forzándolo a hacer distinciones más finas. Esto, a su vez, puede producir respuestas aleatorias y más variaciones en el error de escala.

Otra consideración es usar un número impar o incluso de puntos de escala. Con un número impar, al encuestado se le ofrece una escala de punto medio o una respuesta “neutral”. Dicha respuesta, en efecto, expresa una opción de “no opinión”, “no estoy seguro” o “ni de acuerdo ni en desacuerdo” dependiendo de lo que se evalúa. El desarrollador de la escala debe ser cuidadoso al elegir una redacción adecuada para dicho punto medio si elige etiquetar cada punto de la escala.

Según lo declarado por Clark y Watson (1995: 313), una escala de punto medio de “no puedo decir” puede confundir el punto medio de la escala con una incertidumbre de lo que el artículo significa para el encuestado. Por otro lado, un número par de puntos de escala obliga al encuestado a tener una opinión, o al menos a comprometerse débilmente con lo que se expresa en un ítem (es decir, expresar algún nivel que refleje una preferencia por el endoso o no endosando un indicador), que puede no tener. Aunque algunos investigadores creen que ni un número impar ni un par de puntos de escala son superiores, parecería que para algunos indicadores una respuesta neutral es una respuesta válida, de modo que una cantidad impar de puntos de escala es apropiada.

En general, se ha demostrado que si las escalas dicotómicas y multicotómicas se construyen de manera confiable y válida arrojarán resultados similares. Se recomienda fuertemente que los expertos evalúen cuidadosamente el formato, la redacción de los puntos de escala y el número de puntos de escala, y que se pruebe el piloto antes de otros pasos de construcción de escala.

Número de indicadores del pool inicial

Como se indicó anteriormente, es mejor tener un conjunto de indicadores que incluya en exceso en el dominio del constructo o no no inclusivo. Consistente con esta vista, se recomienda un gran conjunto de indicadores. Mucho dependerá del número y la complejidad de las facetas del dominio del constructo. Si éste se define de manera estrecha y la escala resultante es corta (5-10 indicadores), un *pool* tan pequeño como 20 a 30 indicadores puede ser suficiente.

De hecho, DeVellis (1991) recomienda que para los constructos estrechos, bastará con una agrupación que sea el doble del tamaño de la escala final. Para constructos más amplios y polifacéticos, se necesitarán muchos más indicadores para servir como grupo (*pool*) inicial. Algunos investigadores abogan por un grupo de 250 *indicadores* como ejemplo para la generación de ítems en constructos multifacéticos (Robinson *et al.*, 1991: 12-13). Aún así, uno debe considerar otros asuntos, tales como la redundancia de indicadores, un nivel deseado de consistencia interna y cooperación de los encuestados. Es posible que una agrupación (*pool*) sea tan grande que dificulte la cooperación de los encuestados. En resumen, no hay reglas fáciles y rápidas para el tamaño de un grupo (*pool*) de indicadores inicial. Los constructos de una sola faceta estrechamente definidas requerirán menos ítems que las complejas construcciones multifacéticas. Sin embargo, en general, se prefiere un número mayor, ya que la inclusión excesiva es más deseable que la falta de inclusión.

Evaluando indicadores: validez de contenido y validez aparente

El establecimiento de la validez de contenido (*content validity*) y la validez aparente (*face validity*) varía según la precisión del constructo y el grado en que los expertos concuerden con el dominio y las facetas de la construcción. Cuanto más amplia sea la discrepancia en cuanto al constructo, más difícil será la validación de contenido. Se dice que dicha validez se encuentra amenazada si:

- a. Los indicadores que reflejan cualquiera de los dominios (facetas) se omitieron de la medición,

- b. Los indicadores que miden dominios (facetas) fuera de la definición del constructo se incluyen en la medición,
- c. Un puntaje agregado en el constructo refleja de forma desproporcionada un dominio (faceta) sobre otros, y
- d. El instrumento fue difícil de administrar y responder a las poblaciones objetivo.

Guía para validez de ambas

Varios autores ofrecen pautas integrales para juzgar el contenido y la validez de los indicadores en un grupo (pool) inicial (Clark y Watson, 1995; Haynes et al., 1999; Haynes et al., 1995; Nunnally y Bernstein, 1994). En el capítulo 4, se abordaron estas pautas por lo que se agregan más, como:

1. El investigador debe tener todos los elementos de los indicadores evaluados con validez aparente y validez de contenido). Los ítems en sí, los formatos de respuesta, el número de puntos de escala y las instrucciones para el encuestado deberían ser evaluados por la representatividad mediante la intervención de expertos provenientes de la población de estudio a través de procedimientos cualitativos y cuantitativos. Algunos académicos recomiendan que las valoraciones de categorización de tres puntos deben ser utilizadas por al menos cinco expertos (para la validez del contenido) y cinco expertos de la población objetivo (para la validez aparente) para evaluar el grado en que los ítems “representan” la definición de la constructo y los dominios (es decir, “no representativos”, “algo representativo”, “claramente representativo”).

Algunos abogan por escalas de evaluación de 5 o 7 puntos que piden a cinco expertos o más que califiquen cada elemento en términos de representatividad (o, idoneidad), especificidad y claridad, y luego retienen ítems con altos niveles de acuerdo entre expertos (Haynes et al., 1995). Como regla empírica práctica, se prefieren más expertos (cinco más), ya que la detección de indicadores deficientes o marginales será más confiable si se tiene en cuenta a más evaluadores expertos. En este sentido, tanto los procedimientos cualitativos como los cuantitativos pueden ayudar a identificar indicadores que deban ser refinados, modificados o eliminados.

Los procedimientos cualitativos incluyen expertos que escriben o verbalizan comentarios sobre indicadores específicos, así como grupos de enfoque (focus group) y/o entrevistas personales con miembros de poblaciones relevantes para evaluar la validez aparente.

- Los procedimientos cuantitativos incluyen estadísticas de categorización que evalúan el nivel de acuerdo entre los expertos (Perreault y Leigh, 1989).
2. Los resultados de las evaluaciones deben notificarse cuando se publiquen nuevos instrumentos. Esto permite al lector juzgar el grado en que los procedimientos fueron adecuados y tomar decisiones sobre la adecuación de la escala para su uso y/o ajustes necesarios al instrumento.
 3. Finalmente, algunos psicométricos abogan por que una vez que una agrupación (*pool*) de ítems ha sido juzgada, modificada y/o recortada por expertos de la población objetivo, se esté probando un indicador piloto en una muestra más grande ($n = 100-200$) de una población relevante orden (Clark y Watson, 1995; Haynes et al., 1999). Aquí, los elementos pueden evaluarse inicialmente para la consistencia interna, medias, varianzas, las intercorrelaciones con otros indicadores y la estructura del factor. Los indicadores que no se comporten empíricamente como se esperaba, pueden ajustarse para redacción, etiquetas de escala, etc., como recopilación de datos inicial. Algunos indicadores pueden requerir una eliminación basada en estos análisis, y si es necesario, se pueden redactar y juzgar más indicadores, de modo que se retenga un conjunto de ítems suficiente para los estudios de desarrollo de escala que siguen.

Guía para el desarrollo de escalas

El último punto a tratar en el capítulo, es la propuesta de una serie de pasos que todo investigador debería seguir para desarrollar un instrumento de medida (escala) adecuado de una variable latente.

Determinar qué es lo que se pretende medir

Aunque puede parecer que todo investigador sabe qué es lo que quiere medir, muchas veces se encuentra con que sus ideas son más vagas de lo que creía a la hora de elaborar preguntas. La teoría es la mejor ayuda para esta fase, una escala sólo será correcta si todas las dimensiones del concepto que se quiere medir son conocidas por el investigador, y sólo el conocimiento profundo de la literatura puede garantizar que esta fase se desarrolla correctamente.

Generar un listado de ítems

Una vez que el propósito de la escala está claro, el investigador ha de comenzar a construirla. El primer paso es generar un amplio conjunto de ítems de entre los que

deberán salir los que conformen definitivamente la escala.

Cada uno de estos ítems tiene que reflejar la variable subyacente que se pretende medir. Asimismo, es mejor generar varios ítems para cada una de las dimensiones del constructo que recurrir a una sola pregunta, por cuanto que así puede calcularse su fiabilidad y, además, cuando se vayan sumando las respuestas aflorará el contenido común a ellos y se minimizará el efecto de las particularidades irrelevantes de cada uno por compensación. Así “haría casi cualquier cosa por asegurar la felicidad de mis hijos” y “ningún sacrificio es mucho si ayuda a conseguir la felicidad de mis hijos” podría ser una redundancia útil por cuanto expresan una idea similar de manera distinta.

En cuanto al número de ítems, es imposible dar una cifra de cuál debería ser este en el listado inicial. En todo caso debe recordarse que la fiabilidad de la escala es función de la correlación que exista entre los ítems. Como en esta fase no podemos calcular esas correlaciones, es deseable tener ítems suficientes para efectuar sustituciones que mejoren las correlaciones al desarrollar posteriormente la escala. No es extraño comenzar con un listado de ítems tres o cuatro veces superior a la longitud final que se espera dar a la escala.

Referente a las características de un buen ítem, es imposible hacer un listado exhaustivo de qué es lo que hace que un ítem sea bueno o malo. En todo caso, debería tenerse en cuenta lo siguiente:

1. Deben evitarse los ítems excesivamente largos, por cuanto la longitud incrementa la complejidad y disminuye la claridad. Sin embargo, tampoco debe sacrificarse la claridad en aras a la brevedad eliminando preposiciones y conjunciones. En general un ítem del tipo 'Con frecuencia tengo problemas para expresar mis puntos de vista' es mejor que 'Debo decir que una de las cosas con las que parece que tengo un problema la mayor parte de las veces es la transmisión de mi punto de vista al resto de las personas'.
2. La complejidad sintáctica y de léxico de las frases es también muy importante. Deben evitarse la sucesión de negaciones “No estoy a favor de que las empresas no sigan dando fondos a los grupos antinucleares” es mucho más confuso que “Estoy a favor del apoyo de las empresas a grupos antinucleares”.
3. Deben evitarse también los ítems con doble argumentación. Así “estoy en contra de la discriminación racial porque es un crimen contra Dios” es un ejemplo de doble argumentación. Si se está en contra de la discriminación racial por motivo que nada tienen que ver con la religión ¿qué debe contestarse?
4. Combinación de ítems formulados en positivo y en negativo. Muchos investigadores abogan por combinar ítems que suponen la presencia del constructo

con ítems que suponen su ausencia. Por ejemplo, la escala de Rosenberg (1965) de autoestima RSE incluye frases como “creo que tengo bastantes buenas cualidades” junto a otras como “ciertamente, muchas veces me siento inútil”. El motivo de esta combinación es evitar el sesgo afirmativo del encuestado, es decir, la tendencia manifiesta de estar de acuerdo con las afirmaciones independientemente de su contenido. La combinación permite detectar a las personas que han contestado con este sesgo, dado que habrá dado puntuaciones de acuerdo altas tanto a aquellas variables que indican un alto grado de autoestima, como las que la suponen baja. Desgraciadamente, este tipo de formulaciones tienen un aspecto negativo: provocan confusión en los encuestados.

Determinar el formato de medida

En la literatura sobre las ciencias sociales existen muchas formas de hacer preguntas, nosotros presentaremos las más habituales.

Escala de Thurstone

Las escalas Thurstone formulan las preguntas generando una serie de ítems que suponen la presencia en distintos grados del constructo que se pretende medir. Es también habitual que estén diseñadas de tal forma que la diferencia de nivel del constructo entre cada par de afirmaciones sea la misma. Una hipotética escala Thurstone para medir las aspiraciones de los padres respecto a los logros académicos de sus hijos, sería la mostrada en la tabla 1.

Como señalan Nunally y Bernstein (1994) es mucho más fácil explicar una escala Thurstone que construirla, por la dificultad que entraña generar ítems que aporten grados diferenciales de presencia del atributo.

TABLA 1. Escala de Thurstone. Un ejemplo de caso.

1. Que mi hijo alcance el éxito es lo único que puede compensar mis esfuerzos como padre	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
2. Ir a una buena universidad y conseguir un buen trabajo es importante, pero no esencial, para la felicidad de mi hijo	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
3. La felicidad no tiene nada que ver con conseguir objetivos materiales o educacionales	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
4. Lo que habitualmente se considera un éxito es un obstáculo para la verdadera felicidad	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>

Fuente: Elaboración propia

Escala de Guttman

En una escala de Guttman, los ítems están generados de tal forma que responder afirmativamente a uno de ellos, supone responder afirmativamente a todos los anteriores. Si se pregunta: ¿Fuma?, ¿Fuma usted más de 10 cigarrillos?, ¿Fuma usted un paquete diario?... Responder afirmativamente a una de estas preguntas supone que también se posee el nivel inferior del atributo medido. Aunque tanto la escala Thurstone con la Guttman están formadas por ítems que gradúan la presencia del atributo, en la primera se busca una afirmación que fije el nivel de atributo poseído, mientras que la segunda se busca el punto de transición entre las respuestas afirmativas y las negativas (tabla 2).

TABLA 1. Escala de Thurstone. Un ejemplo de caso.

1. Que mi hijo alcance el éxito es lo único que puede compensar mis esfuerzos como padre	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
2. Ir a una buena universidad y conseguir un buen trabajo es importante para la felicidad de mi hijo	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
3. La felicidad es más probable si una persona ha conseguido sus objetivos educacionales y materiales	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>
4. Lo que habitualmente se considera un éxito no es un obstáculo para la verdadera felicidad	Acuerdo	<input type="checkbox"/>
	Desacuerdo	<input type="checkbox"/>

Fuente: Elaboración propia

La dificultad de las escalas Guttman está en lograr formular los ítems de tal forma que responder afirmativamente a uno suponga hacerlo también a los anteriores. En el ejemplo anterior se supone que si se contesta afirmativamente a la expresión 2 se debe haber contestado igual a la 3 y 4. Sin embargo, si un entrevistado viera el éxito como un fenómeno complejo que puede ser a la vez una ayuda y un obstáculo para la felicidad, podría dar un patrón de respuestas no esperado.

Escala con ítems del mismo peso (Likert y diferencial semántico)

En los dos tipos de escalas mostradas, las dificultades de generación suelen superar sus ventajas y no son muy utilizadas. Lo más habitual es generar ítems que sean detectores equivalentes del fenómeno medido, y no de niveles de éste. Una de sus ventajas es que las respuestas pueden recogerse de muchas formas distintas, lo que permite al investigador buscar la más adecuadas para su propósito. Inconvenientes y ventajas, se tienen como:

a. ¿Cuántas categorías de respuesta? La mayor parte de los ítems de una escala, consisten en dos partes: una afirmación y una serie de opciones de respuesta. Estas opciones pueden ir desde un gran número de alternativas (valorar de 1 a 100 un acuerdo, por ejemplo), a unas pocas (Muy de acuerdo, de acuerdo, indiferente, en moderado desacuerdo, muy en desacuerdo). No debemos olvidar que la misión de una escala es detectar la variabilidad en las respuestas, y fracasará en esta tarea si no puede discriminar entre los distintos niveles de presencia del atributo si las alternativas de respuesta son muy limitadas.

Una segunda cuestión, es si el entrevistado tiene capacidad para discriminar de manera significativa entre los distintos niveles de respuesta, lo que depende claramente de qué se está midiendo. Si se le pregunta sobre su ideología política, puede tener necesidad de matizar su respuesta y requerirá de una escala de bastantes puntos (digamos 10). Pero si se le pregunta acerca de su acuerdo o desacuerdo con la afirmación “Fumar perjudica la salud” esta necesidad de matización no será tan acuciante y es razonable graduar la respuesta en menor número de alternativas. ¿Han de ser estas alternativas un número par o impar? De nuevo depende de lo que se pregunte. El número impar supone la existencia de un punto neutral. Si el investigador considera que, dadas las características de la pregunta, el entrevistado puede buscar conscientemente la indefinición, será recomendable un número par de alternativas para forzar que tome partido.

En otras preguntas, sin embargo, el encontrar el punto de indefinición puede ser muy útil. Veamos el ejemplo que plantea DeVellis (1991). En un estudio para determinar cuál de dos riesgos prefiere la gente asumir (aburrimiento frente a peligro), el investigador puede ir variando la peligrosidad de la actividad hasta dar con el punto de inflexión donde el entrevistado, duda entre aburrimiento y peligro. Esa actividad sería un indicador de la propensión al riesgo del entrevistado (tabla 3).

b. Las escalas de Likert son las más utilizadas en el desarrollo de escalas. El ítem se presenta como una afirmación, seguida por alternativas de respuesta que suponen diversos niveles de acuerdo en ella. A estas escalas se le pueden aplicar las consideraciones anteriores sobre el número par o impar de alternativas y la cantidad de estas. Estas escalas plantean básicamente un problema. Si las afirmaciones son muy neutras, existe una gran tendencia al acuerdo y, además, pueden estar recogiendo, de hecho, más la ausencia de opinión que la opinión. Por ejemplo, no es lo mismo pedir el acuerdo o desacuerdo con afirmaciones como “los médicos no hacen caso normalmente a lo que les dicen sus pacientes” o “muchas veces los médicos no prestan demasiada atención a sus pacientes” que hacerlo sobre “de vez en cuando, los médicos pueden

olvidar algo de lo que les han dicho sus pacientes”. En resumen, un buen ítem likert debería manifestar la opinión de una manera clara.

- c. Diferencial semántico. En general, el diferencial semántico va asociado a un estímulo (como un grupo de personas, por ejemplo, los vendedores de coches). Una vez identificado el estímulo se presentan una lista de pares de adjetivos que representan los puntos opuestos de un continuo (honrado vs. no honrado) como se muestra en la tabla 4.

TABLA 1. Escala de Thurstone. Un ejemplo de caso.

<p>Señale su preferencia relativa por las actividades A o B entre las alternativas señaladas a continuación.</p> <p>Actividad A: Leer un libro de filosofía de la ciencia (ningún peligro).</p> <p>Actividad B: Viajar en un vuelo comercial (poco peligro)</p> <p>1 = Claramente prefiero A</p> <p>2 = Prefiero A</p> <p>3 = No prefiero una ni otra</p> <p>4 = Prefiero B</p> <p>5 = Claramente prefiero B</p> <p>Actividad B: Viajar en una avioneta 1,2,3,4,5</p> <p>Actividad B: Saltar en paracaídas, teniendo paracaídas de reserva 1,2,3,4,5</p> <p>Actividad B: Saltar en paracaídas sin paracaídas de reserva 1,2,3,4,5</p> <p>Actividad B: Saltar de una avión sin paracaídas para intentar llegar a una colchoneta 1,2,3,4,5</p>
--

Fuente: Elaboración propia

TABLA 4. Escalas con ítems en diferencial semántico. Un ejemplo de caso.

Vendedores de automóviles		
Honrados	-----	No honrados
Apacibles	-----	Ruidosos

Fuente: Elaboración propia

El listado de ítems debe ser revisado por expertos

Esta fase tiene mucho que ver con la difícil tarea de lograr la validez de contenido de la escala que no puede constatarse de ningún otro modo. Tras dar al panel de expertos la definición que el investigador ha hecho del constructo que tiene que medir y del conjunto de ítems que componen la escala, solicitará al panel que realice las siguientes tareas:

1. Los expertos confirmarán o invalidarán la definición que el investigador haya hecho del constructo que quiere medir.
2. A continuación el panel de expertos debería valorar el nivel de relevancia que creen que cada ítem tiene para medir el fenómeno que se pretende. Estas valoraciones permitirán seleccionar la muestra definitiva de ítems eligiendo entre los más significativos.
3. Cuando el constructo tiene distintas dimensiones, los expertos deberán revisar la asignación hecha de afirmaciones a dimensiones.
4. Expertos también deberán valorar la claridad y precisión de la formulación concreta de cada ítem, apuntando, si es necesario, formas alternativas de redactarlos. Esto hace referencia tanto a problemas de claridad estrictamente literaria, como a que puedan por esta falta de claridad reflejar factores extraños al constructo que se pretende medir.
5. También se pretende que los expertos determinen si se ha dejado fuera de la escala alguna de las dimensiones del fenómeno.

Considerar la inclusión de ítems de validación

Además de los ítems que pretenden medir la variable subyacente, el investigador debe considerar la incorporación de dos tipos de ítems adicionales:

- Ítems para detectar defectos o problemas de la escala. A veces los entrevistados pueden no estar respondiendo a las preguntas con la motivación que el investigador presupone. Una de las motivaciones más habituales es presentar una imagen socialmente aceptable, y no su verdadera forma de ser. Suele recomendarse para ciertos tipos de estudios la inclusión de una escala que mida esta tendencia, como la de Strahan y Gerbasi (1972).
- *Ítems para evaluar la validez de constructo de la escala.* Si la teoría apunta que el constructo que se está midiendo está relacionado con otros constructos, la inclusión de escalas de medición de algunos de ellos puede ayudar a analizar la validez de la nueva escala desarrollada, si se confirman esas relaciones.

Administrar la escala a una muestra de prueba

La muestra tiene que ser lo suficientemente grande como para eliminar la varianza que aporta un individuo como tal, haciendo que no sea representativa de una población determinada, sino de sí misma. Nunally y Bernstein (1994) apuntan la cifra de 300 personas como un número adecuado, aunque DeVellis (1991) señala que se han desarrollado escalas con éxito con un menor número.

Evaluando los ítems

Con los resultados de administrar la escala a una muestra de prueba, debe analizarse como han funcionado cada uno de los ítems de tal forma que se pueda seleccionar finalmente a los más adecuados.

- *Análisis inicial del funcionamiento de cada ítem.* Cuánto mayores sean las correlaciones entre los ítems, mayor será la fiabilidad de la escala. Por ello serán candidatos a abandonar la escala aquellos ítems que, perteneciendo a una misma dimensión del constructo, guarden poca correlación con los demás integrantes de la misma.
- *Formulación inversa.* Si algún ítem guarda una elevada correlación negativa con los demás, deberá formularse gramaticalmente en sentido inverso para mantenerlo en la escala. Si esta forma de enunciarlo se ha hecho intencionadamente, deberá cambiarse la codificación de la respuesta.
- *Analizar la varianza de los ítems.* Si todos los individuos responden igual a un ítem, la varianza será cero. En este caso (o en casos de varianzas muy pequeñas) el ítem no ha sido capaz de discriminar entre los individuos con distintos niveles del *constructo que se está midiendo, y debería suprimirse.*
- *Analizar la media de los ítems.* Es deseable que la media de los ítems esté entorno al centro de la escala de respuesta. Si están en la zona del acuerdo, indicará que el enunciado de la afirmación era demasiado neutro.
- *Cálculo del alpha de Cronbach.*

Optimizar la longitud de la escala

Ya se ha indicado que la fiabilidad de una escala viene influenciada por el número de ítems de la misma. Por un lado, escalas cortas reducen los problemas de su administración a los entrevistados, pero las largas son más fiables. Encontrar el punto de equilibrio debe ser el objeto de esta etapa. Si una escala es poco fiable, entonces la brevedad no es una virtud. Dicho de otro modo, se pueden intentar

eliminar ítems para facilitar la administración, pero nunca permitiendo que los alpha de Cronbach sean inferiores al .7 o .8 recomendados.

Sin embargo, como vimos en el ejemplo del epígrafe 3, eliminar malos ítems puede mejorar significativamente la fiabilidad de la escala. En ese mismo ejemplo, mostrábamos cómo detectar y eliminar esos ítems. Debe mantenerse, sin embargo, un cierto margen de seguridad con el valor de los alpha, porque no debe olvidarse que la escala se ha administrado a una muestra de prueba y estos valores pueden caer cuando se administren a la muestra definitiva. Si la muestra de prueba es lo suficientemente grande, debería considerarse su división en dos mitades, de tal forma que una de ellas se utilice para calcular los alpha, evaluar los ítems y eliminar los inadecuados, mientras que la segunda mitad se podrá utilizar para validar estos resultados.

Resumen

En este capítulo, hemos ofrecido varias ilustraciones de los dos primeros pasos importantes en el desarrollo de la escala: el paso 1 de la definición de construcción y el dominio de contenido, y el paso 2 de generar y evaluar los indicadores de medición. Se discutieron varios aspectos importantes relacionados con el primer paso, incluido el papel de la teoría en la definición de constructo y el dominio de contenido, la importancia de una revisión y definición exhaustivas de la literatura y la importancia de una dimensionalidad de construcción a *priori*. También se destacaron los temas relevantes para el segundo paso, incluyendo la generación de un grupo de elementos suficientemente grande para aprovechar el dominio de contenido, los diversos aspectos de la escritura de artículos y el uso de expertos para calificar el la validez aparente y de contenido. Los capítulos 6 y 7 se centrarán en estudios empíricos para desarrollar, perfeccionar y finalizar una escala. Por último, se hace una recopilación de recomendaciones de cómo desarrollar una escala.

CAPÍTULO 6

Diseñando la escala. Análisis factorial exploratorio (etapa 3)

El capítulo 5 se centró en la definición de constructos y en la generación, evaluación y refinamiento de un conjunto de indicadores inicial, es decir, los pasos 1 y 2 en el desarrollo de la escala. Este capítulo se centra en el diseño y la realización de estudios iniciales para su desarrollar y reducir aún más el conjunto de indicadores o ítems, es decir, el Paso 3 en el desarrollo de la escala. Específicamente, abordamos los siguientes objetivos: prueba piloto de un grupo de indicadores como reducción de indicadores y procedimiento de prueba de validez inicial, y realización de múltiples estudios para el desarrollo de la escala. Dentro de este segundo objetivo, destacamos:

- a. La importancia de incluir constructos para evaluar los diferentes tipos de validez,
- b. El análisis factorial exploratorio (EFA. *Exploratory Factor Analysis*) sobre múltiples conjuntos de datos para refinar la escala y examinar una estructura teórica *a priori* de factores iniciales, y
- c. El análisis de indicadores y su confiabilidad.

Finalmente, ofrecemos ejemplos de escalas recientemente desarrolladas para ilustrar cómo se pueden usar el EFA y análisis de elementos y confiabilidad en el desarrollo de la escala.

Prueba piloto

Como se indicó en el capítulo 5, se debe generar y evaluar un conjunto de indicadores para determinar el contenido y la validez aparente (“*face validity*”) antes de que se administren a las muestras del sujeto de estudio. Como también se indicó en el Capítulo 5, el conjunto o (pool) de indicadores inicial debería ser lo suficiente-

mente grande, con el tamaño apropiado variable dependiendo de la dimensión y la complejidad de la definición de constructo. Con un gran conjunto de indicadores, será útil reducir el grupo para los estudios de desarrollo que siguen y obtener algunas estimaciones iniciales de confiabilidad y validez. En este sentido, las pruebas piloto pueden ser útiles. Dado que la escala se administrará a más muestras para refinar, un estudio piloto puede reducir el número de elementos en un grupo inicial a un número más manejable al eliminar elementos que no cumplen ciertos criterios psicométricos. Las pruebas piloto también pueden servir para el propósito de probar inicialmente la validez. Posteriormente, una red nomológica puede evaluarse parcialmente mediante la inclusión de medidas/escalas de constructos relevantes con una muestra piloto. Cuatro cuestiones a tener en cuenta en las pruebas piloto:

Tamaño de la muestra. Considere el tamaño de la muestra de los encuestados en la prueba piloto. Algunos sugieren que las muestras piloto deberían estar desde $n = 300$ (DeVellis, 1991), mientras que otras sugieren que $n = 100$ a 200 serán suficientes (Clark y Watson, 1995). Un problema a tener en cuenta con el tamaño de muestra es cuántos indicadores hay en el grupo inicial. Para las agrupaciones con un pequeño número de elementos (por ejemplo, 20 elementos o menos) asociados con constructos estrechamente definidos, desde $n = 100$ a 200 parece razonable, ya que los análisis iniciales de elementos y factores pueden realizarse con dichos tamaños de muestra. Sin embargo, para un pool más grande (y para constructos complejas y/o multidimensionales), se prefieren tamaños de muestra más grandes desde $n = 300$.

Composición de la muestra. Para las pruebas piloto, las muestras de conveniencia (por ejemplo, estudiantes universitarios) pueden ser suficientes, pero es preferible usar una muestra de una población relevante de interés. Dado que el objetivo de la prueba piloto es reducir el número de indicadores a un número más manejable para los estudios de desarrollo más grandes, la representatividad de la muestra es un plus. Un elemento que se desempeñe bien (o mal) con una muestra de una población relevante será evaluado con más confianza como candidato para la inclusión (o supresión) para las siguientes muestras de desarrollo. Por lo tanto, se prefieren las muestras de poblaciones relevantes.

Estimaciones iniciales de la confiabilidad de los indicadores. Los elementos pueden evaluarse inicialmente para la consistencia interna, los medios, las variaciones, la correlación interitem media y la estructura del factor. Los indicadores que no se comporten empíricamente como se esperaba pueden ajustarse para la redacción,

las etiquetas de escala y, de otro modo, para la retención en los esfuerzos de recopilación de datos de desarrollo.

Algunos indicadores pueden requerir una eliminación basada en estos análisis, y si es necesario, se pueden redactar y juzgar más de ellos, de modo que se retenga un conjunto de indicadores suficiente para los estudios de desarrollo de escala que siguen.

Número y el tipo de escalas relacionadas con la validez a incluir. Se puede hacer una evaluación inicial de algunos aspectos de la validez a través de una muestra piloto. Aquí, algunas mediciones de constructo clave pueden ser muy útiles para obtener estimaciones iniciales de validez (es decir, parte de la red nomológica del constructo). Como se dijo anteriormente, una revisión exhaustiva de la literatura debe alertar al desarrollador de la escala sobre los constructos y las medidas complementarias que la construcción focal debe relacionarse con (y diferenciarse de). Por lo tanto, deberían incluirse tres o cuatro mediciones de constructos clave para examinar las estimaciones iniciales de validez. Como nota final sobre las pruebas piloto, es mejor tener un grupo de elementos que incluya en exceso el dominio de contenido del constructo que uno que sea escaso. Dado que el objetivo de las pruebas piloto es reducir indicadores, una pregunta importante es “¿Cuántos artículos se deben retener de la prueba piloto?” Por lo tanto, si un indicador se desempeña de manera marginal en los análisis de elementos y factores en una prueba piloto, pero se juzga que tiene contenido y validez nominal, lo mejor es retener ese elemento para otras muestras. Este será particularmente el caso si la muestra piloto fuera de conveniencia y no necesariamente representativa de la población de interés.

Análisis factorial exploratorio (EFA. *Exploratory Factor Analysis*)

Damos continuidad a lo mencionado en el apartado Análisis factorial exploratorio, de dimensionalidad para realizar una breve discusión de lo que implica a nivel de reducción de dimensiones y complementar lo citado en Mejía (2017c: 6-71) y darle al lector, mayores argumentos para la toma de sus decisiones de análisis.

El análisis factorial exploratorio (EFA) pueden usarse para dos propósitos principales en el desarrollo de la escala:

- a. Para reducir el número de ítems en una escala, de modo que los ítems restantes maximicen la varianza explicada en la escala y maximicen la confiabilidad de la escala, y
- b. Para identificar las posibles dimensiones subyacentes en una escala.

Estos dos usos están relacionados, y en el caso del desarrollo de escala, pueden usarse de manera complementaria. Típicamente, el análisis de factores por componentes principales (PCA. *principal components factor analysis*) se usa para tal fin en lugar del análisis factorial común (*common factor analysis*), (Floyd y Widaman, 1995; Hair, Anderson, Tatham y Black, 1998). Con PCA, los componentes (o factores) se estiman de tal manera que representan las varianzas entre los indicadores de la escala de la manera más sencilla (*económica*) posible, con el menor número posible de componentes significativos (*dimensiones*). PCA analiza la matriz de correlación entre las variables observadas (indicadores) con las de la diagonal principal. Al hacerlo, PCA maximiza todas las varianzas de los indicadores, independientemente de si es común a un factor (*componente*) o a un único un indicador. Aquellos ítems para los que se explica con una baja varianza se consideran candidatos para la eliminación (es decir, para reducir la cantidad de elementos en la escala).

Por el contrario, el análisis factorial común (*common factor analysis*, en el desarrollo de la escala) se asocia más a menudo con la búsqueda de dimensiones subyacentes para un conjunto de elementos. También utiliza una matriz de correlaciones (o covarianzas) entre los puntajes observados en los ítems para identificar un conjunto de variables (*factores*) latentes que explican las correlaciones entre ellos. Sin embargo, el análisis factorial común usa las estimaciones de comunalidad de los indicadores en la diagonal principal. Además, la varianza en un elemento dado se divide en lo que es común a un factor o variable latente (basado en la varianza compartida con otros elementos en el análisis) y la varianza que es única para un elemento dado, una combinación de factores específicos varianza y varianza de error aleatorio. El análisis factorial común puede usarse para identificar el (los) constructo(s) teórico(s) cuyos efectos se reflejen por las respuestas a los ítems en una escala. El análisis factorial común también puede ofrecer información sobre qué elementos eliminar o retener, similar a la de PCA. Muchos autores informan que las soluciones derivadas de PCA y análisis de factores comunes tienden a ser bastante similares. Este es particularmente el caso cuando la cantidad de indicadores supera los 30 o los las comunales superan los .60 para la mayoría de los indicadores (Hair *et al.*, 1998). Otros han cuestionado el hallazgo de que PCA y análisis de factores comunes arrojan resultados similares. Las diferencias pueden ser más pronunciadas con un pequeño número de indicadores y bajas comunales. De hecho, en tal caso, PCA y análisis de factores comunes pueden ofrecer resultados divergentes (Floyd y Widaman, 1995). Por lo tanto, algunos autores sugieren que se prefieren los análisis de factores comunes sobre el PCA, ya que la mayoría de las aplicaciones de desarrollo de escala buscan comprender un constructo en términos del número de factores latentes que lo subyacen, así como

también reducir el número de elementos en una escala. Además, dado que el análisis factorial confirmatorio () es una herramienta ampliamente utilizada en la finalización de la escala, los análisis de factores comunes basados en el EFA pueden generalizarse mejor al CFA que la PCA (Floyd y Widaman, 1995). Ver más al respecto en Mejía-Trejo (2017c: 18).

Número de factores a extraer

Existen varios criterios para extraer factores. Hair et al. (1998), Floyd y Widaman (1985) y Sharma (1996) ofrecen exposiciones muy claras y concisas de estos criterios. Solo los revisaremos brevemente aquí. Aunque existen algunas pruebas estadísticas para la extracción de factores, generalmente están restringidas a técnicas de estimación más alineadas con el CFA (por ejemplo, estimación de máxima probabilidad, a discutir en el capítulo 7). Para el EFA, los criterios psicométricos y las reglas clave se aplican con mayor frecuencia. Como se señaló en el capítulo 2, la regla *eigenvalor* >1 (criterio Kaiser-Guttman o criterio de raíz *latente latent root criterion*) a menudo se utiliza como criterio psicométrico. Cada componente (factor) tiene un eigenvalor propio que representa la cantidad de varianza contabilizada por el componente, donde la suma de todos los eigenvalores es igual al número de indicadores analizados. Un eigenvalor <1 indica que ese componente tiene menos varianza que cualquier elemento individual. Por lo tanto, con la reducción de datos como objetivo, un componente con eigenvalor <1 no se considera significativo. Cabe señalar que el criterio de Kaiser-Guttman puede subestimar el número de componentes en algunas circunstancias y puede no ser confiable. Cliff (1988) demostró que la regla del eigenvalor >1 tiene defectos, y probablemente se utilice mejor como una guía y no como un criterio absoluto para extraer factores. A menudo se recomiendan otras reglas generales para la extracción de factores, tales como:

1. La primera regla clave es el *scree test* o *gráfico de sedimentación*. Esta prueba, traza los eigenvalores y muestra la pendiente de una línea que los conecta. Los factores son retenidos hasta donde la pendiente de esta línea se aproxima a cero, y en ese punto se produce un “codo” agudo. Eliminar un factor muy por debajo de este codo mostrará escasa pérdida de la varianza explicada. Tal procedimiento puede ser particularmente el caso de PCA, pero con análisis de factores comunes, quizás valga la pena examinar los eigenvalores <1 porque el juicio del “codo” es subjetivo. De hecho, Floyd y Widaman (1995) sugirieron que para el análisis factorial común, cuando dos o más factores están cerca del punto de corte del codo, soluciones alternativas con diferentes números de factores deben ser examinados como un esfuerzo para evitar desechar un

factor útil. Así, se tiene a Horn (1965) quien propuso el uso del análisis paralelo (parallel analysis) para identificar y precisar el “codo” (para más detalles consulte el capítulo 2, apartado Análisis factorial exploratorio). Dicho análisis calcula los eigenvalores de las variables para un tamaño de muestra dado, suponiendo que las correlaciones entre las variables, son el resultado únicamente del error de muestreo. Es decir, este análisis proporciona una estimación de los eigenvalores para los indicadores que no tienen factores comunes. (La ecuación 3 del capítulo 2 se puede usar para determinar estos valores propios).

2. Una segunda regla clave para la retención de factores de retención se refiere a la cantidad de indicadores que cargan sustancialmente en un factor. Lo que se considera sustancial es algo abierto para el debate, pero las carga factoriales en desde .40 y superiores se han clasificado como sustanciales (Floyd y Widaman, 1995), y las cargas superiores a .50 se han considerado como “*muy significativas*” (Hair *et al.*, 1998). Los factores con una sola carga sustancial serán de poca importancia porque sólo se tiene en cuenta la varianza del factor específico asociado con ese elemento, y se ha sugerido que al menos tres indicadores que cargan factorial sean necesarios para identificar un factor (Comrey, 1988). Sin embargo, muchos autores sugieren que el tamaño de la muestra debe considerarse al juzgar no solo el tamaño de la carga factorial sino también su estabilidad. Las reglas generales para las técnicas de efa varían desde un tamaño de muestra mínimo de 100 hasta un tamaño de 200 a 300; Otra recomendación es un tamaño de muestra de 5 a 10 encuestados por indicador (Clark y Watson, 1995; Comrey, 1988; Floyd y Widaman, 1995; Hair *et al.*, 1998).
3. Una tercera regla clave implica que la cantidad de varianza se explica por un factor extraído en relación con la varianza total explicada por la solución de factor completa. Algunos defienden que el número de factores extraídos debe representar entre 50% y 60% de la varianza en los ítems y que para que un factor sea significativo, al menos el 5% de la varianza total explicada debería ser atribuible a ese factor (Hair *et al.*, 1998).
4. Una cuarta regla clave que parece apropiada para el desarrollo de la escala implica un criterio *a priori*. La mayoría de los desarrolladores de escala tendrán alguna idea sobre la cantidad de factores (*dimensiones*) que subyacen a un conjunto de indicadores. Por lo tanto, restringir la solución del factor a un número de factores preespecificado (consistente con la teoría) puede ofrecer información valiosa sobre la cantidad de varianza que los factores explican y la fuerza de las cargas de los indicadores sobre los factores respectivos (Floyd y Widaman, 1995; Hair *et al.*, 1998). Este enfoque también puede proporcionar información sobre el nivel de carga cruzada de un indicador a un factor que no

debe cargar (teóricamente) y, por lo tanto, revelar un elemento que puede ser un candidato para su eliminación. En la práctica, no se debe confiar en una sola regla o criterio psicométrico al decidir la cantidad de factores que se deben extraer. La mayoría de los criterios de extracción de EFA deben usarse junto con otros criterios, y el desarrollo de escala debería usar una teoría a priori y el sentido común como guías para decidir la cantidad de factores a extraer. Recomendamos que se realicen análisis factoriales que restrinjan una solución a un número de factores teóricamente a priori, con esos resultados comparados con soluciones no restringidas a un número a priori. Nuevamente, los desarrolladores de escala deberían tener alguna idea sobre la cantidad de factores que subyacen a un conjunto de indicadores. Restringir la solución a ese número puede ofrecer valiosa información de diagnóstico.

Métodos de rotación

Para hacer que los factores sean más interpretables (y que la retención y eliminación de indicadores sea más significativa), los factores se “rotan” después de la extracción. Un objetivo básico para los desarrolladores de escala es buscar una estructura simple después de la rotación. Una estructura simple se produce cuando cada indicador se carga con la mayor cantidad posible de factores o, más preferiblemente, tiene una carga sustancial en un solo factor. La rotación puede especificarse como ortogonal o como oblicua, donde la rotación ortogonal mantiene los factores no correlacionados y la rotación oblicua permite que los factores se correlacionen. VARIMAX es la forma más común de rotación ortogonal para EFA y mostrará una estructura simple en la mayoría de los casos. Sin embargo, dado que un objetivo de EFA para el desarrollo de la escala es buscar el grado al cual se correlacionan múltiples escalas/dimensiones, se recomiendan métodos de rotación oblicua (como PROMAX). La rotación oblicua revelará (en la mayoría de los casos) los factores teóricos más significativos. Ambos métodos pueden ayudar a revelar qué elementos conservar y eliminar para futuros estudios, lo que nos lleva a nuestro próximo problema en la aplicación de EFA como una herramienta de desarrollo de escala.

Retención de indicadores

Dado que EFA puede usarse como un método para reducir el número de indicadores en una escala, surgen preguntas sobre cuántos elementos deben eliminarse y qué criterios usar para eliminarlos. Como se indicó anteriormente, la obtención de una estructura simple es el objetivo de EFA logrado al buscar cargas que son

sustanciales (.40 y superiores). Este objetivo debe usarse con EFA en múltiples conjuntos de datos (es decir, al menos dos). Además, los desarrolladores de escala también deben buscar cargas extremadamente altas, ya que los indicadores con tales cargas pueden ser indicativos de redundancia de redacción que no agrega sustantivamente a la coherencia o validez interna de una escala. Por lo tanto, en general, abogamos por la retención de indicadores a través de múltiples EFA con cargas no inferiores a .40 pero no superiores a .90. En esta etapa, sin embargo, advertimos contra la eliminación de ítems que puedan no cumplir con este criterio, pero se considera que tienen validez aparente (*face validity*) y/o contenido (*content validity*). Además, la supresión y retención de ítems en los primeros estudios de desarrollo de la escala, deberían considerar simultáneamente la confiabilidad y las estadísticas tales como las correlaciones de indicadores a total corregidas (*corrected item-to-total correlations*), las correlaciones entre elementos promedio (*average interitem correlations*) y las varianzas de ítems (*item variances*).

Prueba inicial de indicadores y de confiabilidad

En las etapas más tempranas del desarrollo de la escala, es aconsejable examinar una serie de estadísticas de consistencia interna (*internal consistency*) y las basadas en ítems (ítem-based) junto con la EFA para usarlas en la toma de decisiones de conservar y/o eliminar indicadores. Estas estimaciones de consistencia interna incluyen: correlaciones interitem promedio de coeficientes alfa (coefficient alpha average interitem correlations), correlaciones de indicadores a total corregidas (“corrected item-to-total correlations”), varianzas de ítems (ítem variances) y redundancia de redacción de ítems (item-wording redundancy). Se han defendido varias heurísticas o reglas generales para los ítems como candidatos para la retención/eliminación. Robinson et al. (1991) abogó por una correlación interitem promedio de .30 o mejor como ejemplo. Clark y Watson (1995) recomendaron correlaciones interitem promedio de .15 a .50 a través de constructos; por constructos estrechamente definidos, abogaron por un rango de .40 a .50. También propugnaron un nivel de coeficiente alfa de al menos .80 para una nueva escala y sugirieron que retener ítems con mayores variaciones en relación con otros ítems ayudaría a aumentar la varianza de la escala global. Bearden y Netemeyer (1998) abogaron por correlaciones de elemento a total corregidas (*corrected item-to-total correlations*) de .50 y superiores y niveles alfa de .80 o superiores (sujeto a consideraciones de longitud de escala). Estas heurísticas tienen sentido, pero también se debe recordar que el coeficiente alfa está relacionado con la longitud de la escala, la correlación interitem media (*average interitem correlation, o covarianza*), la

redundancia de ítems y la dimensionalidad. Así que, considere primero, la longitud de la escala. Se observó en el capítulo 3 que a medida que aumenta el número de indicadores, alfa tenderá a aumentar. Debido a que la parsimonia también es una preocupación en la medición y la mayoría de las escalas son autoadministradas, la brevedad de la escala es a menudo una preocupación. La redundancia de elementos y la correlación interitem media (*average interitem correlation o covarianza*) también afectan el coeficiente alfa. Aunque se defiende que se necesitan varios indicadores para aprovechar adecuadamente el dominio del constructo, cuando la redacción de los elementos es demasiado similar, el coeficiente alfa (así como la validez y la dimensionalidad del contenido) puede no mejorarse. Los indicadores que están redactados de manera similar aumentarán la correlación entre elementos promedio y el coeficiente alfa; sin embargo, tales elementos pueden solo contribuir a la paradoja de atenuación (*attenuation paradox*) en la teoría psicométrica, por lo que aumentar el alfa coeficiente más allá de cierto punto no aumenta la consistencia interna (ver capítulo 3). Con estas consideraciones en mente, los investigadores deben ser cuidadosos en su interpretación de alfa y considerar su relación con el número de indicadores en una escala, el nivel de correlación interitem y la redundancia de indicadores. Por último, y como se discutió en los Capítulos 2 y 3, también debe recordarse que es posible que un conjunto de indicadores esté interrelacionado pero no sea unidimensional. El coeficiente alfa no es una medida de unidimensionalidad y debe usarse para evaluar la consistencia interna sólo después de que se haya establecido la unidimensionalidad (Clark y Watson, 1995; Cortina, 1993).

Consideraciones finales

Aunque el EFA y las estadísticas tanto de consistencia interna (*internal consistency*) como las basadas en ítem (*ítem-based*) pueden usarse en los primeros estudios de desarrollo de escala como medios para decidir qué ítems retener y eliminar, siempre es mejor retener muchos ítems en esta etapa para estudios posteriores. Es decir, los indicadores que no cumplen ciertos criterios estadísticos o reglas generales, pero que tienen validez aparente (*face validity*) /o contenido (*content validity*) deben mantenerse para la próxima “ronda” de estudios. Si continúan funcionando mal, siempre se pueden eliminar al derivar la forma final de la escala.

Caso de estudio

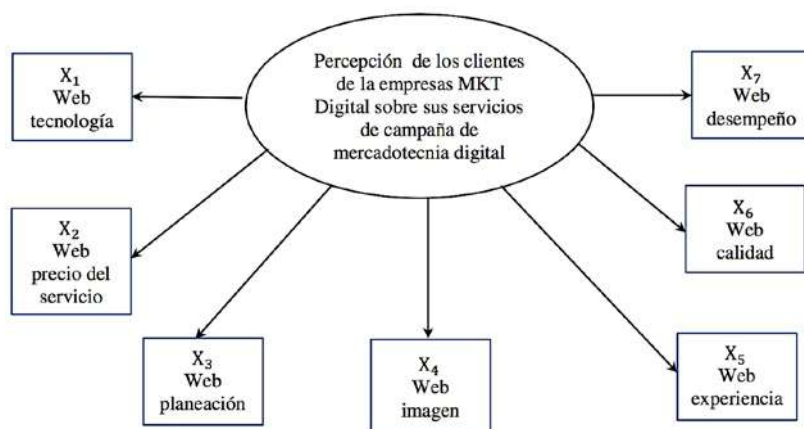
Se remite al lector a leer el caso de estudio presentado en la obra *Las Ciencias de la Administración y el Análisis Multivariante. Proyectos de investigación, análisis y*

FIGURA 1. Base de datos y diagrama de trayectoria. Caso de estudio inicial.

Base de datos.

Descripción de variables	Clasificación de variables	Tipo
V1.- Tipo de consumidor	Clasificación de variables	No métrico
V2.- Tipo de Industria		No métrico
V3.- Tamaño de la firma		No métrico
V4.- Región		No métrico
V5.- Sistema de contratación		No métrico
X1.- Web tecnología	Variables de efectividad	Métrico
X2.- Web precio del servicio		Métrico
X3.- Web planeación estratégica		Métrico
X4.- Web imagen		Métrico
X5.- Web experiencia del usuario		Métrico
X6.- Web calidad		Métrico
X7.- Web desempeño		Métrico
X8.- Web analítica		No métrico
X9.- Web contrataciones de clientes		Métrico
X10.- Web desempeño		Métrico
X11.- Web certificación		No métrico
X12.- Web plan de implementación		No métrico
X13.- Web seguridad		No métrico
X14.- Web situación de compra		No métrico
V15.- Probabilidad de recomendación	Variables de resultado	Métrico
V16.- Probabilidad de compra futura		Métrico
V17.- Nivel actual de compra		Métrico
V18.- Futuras asociaciones/alianzas		No métrico

DIAGRAMA TRAYECTORIA



Fuente: Mejía-Trejo (2017c: 43).

discusión de resultados de Mejía Trejo (2017c: 42-71), el cual tratamos aquí brevemente, como sigue:

Se utilizará la base de datos de WEB Diseño.sav, de la compañía MKT Digital, la cual contiene datos de 100 clientes a los que ha servido en sus campañas digitales., con la siguiente estructura de base de datos, como caso de estudio inicial. Ver Figura 1.

EFA: Condiciones iniciales

A continuación, se muestran varias consideraciones del caso, que convalidan de manera gradual (ver el tratamiento estadístico en Mejía Trejo (2017c), de todas y cada una de las consideraciones), la validez aparente (*face validity*) y la validez de contenido (*content validity*), como sigue:

A nivel diseño: la base de datos contiene 100 observaciones con un total de 23 variables, de las cuales se trabajará con 7/14 de las que corresponden al grupo de Efectividad. La base de datos WEB diseño.sav contiene las percepciones que tienen los 100 principales clientes. Asimismo, se recolectó la información de percepciones a través de *focus groups* y entrevistas a profundidad aplicados a los principales clientes de la empresa MKT Digital. Los gerentes de mercadotecnia digital, contestaron los 13 atributos utilizando una escala de 0 a 10, siendo 10 Excelente y 0 Pésimo, con 1 decimal. La información contempla aspectos como: tecnología, precio, planeación estratégica, imagen, experiencia del usuario, calidad, desempeño implicados en una campaña digital. Con el resultado del análisis factorial, se espera detectar el mejor agrupamiento de estas variables que le permitan al sector de mercadotecnia digital, ser más competitivo y prestar mejor servicio al cliente.

A nivel valores perdidos: La existencia de valores perdidos en un estudio de las Ciencias Sociales es algo prácticamente inevitable. Las consecuencias para la investigación dependerá del patrón que sigan estos datos ausentes, cuántos son y por qué están perdidos. El patrón de los valores perdidos es más importante que su cuantía; pues si su distribución es aleatoria en la matriz de datos no pueden causar mucho daño al análisis; sin embargo, si responden a un patrón determinado sí. Valores perdidos menores o iguales a 10 por ciento del total de los casos por lo general son ignorados y sustituidos por la media o la moda, según sea el caso, excepto cuando los valores perdidos se concentran en una pregunta determinada o un grupo de preguntas del cuestionario.

A nivel de valores atípicos u outliers: Son aquellos casos para los que una, dos o múltiples variables de una investigación determinada toman valores extremos que

los hace diferir del comportamiento del resto de la muestra, y permiten al investigador sospechar que han sido alterados o generados por mecanismos distintos al resto de los datos. ¿Por qué es importante detectar los valores atípicos? Por las consecuencias que generan: Distorsionan los resultados al oscurecer el patrón de comportamiento del resto de casos y obtenerse conclusiones que, sin ellos, serían completamente distintas y, pueden afectar gravemente a una de las condiciones de aplicabilidad más habituales de la mayor parte de las técnicas multivariantes: la normalidad.

A nivel de normalidad: La condición básica que debe asumirse en el análisis multivariante es la normalidad, y se refiere a que todos los datos de las variables métricas deben de seguir una distribución normal. Si la variación de la distribución normal es demasiado amplia, todos los resultados del análisis multivariante serán inválidos, porque la normalidad es un requisito esencial para los estadísticos F y t. La prueba de normalidad para una sola variable es fácil de realizar y existen diversos métodos para estimarse, pero la normalidad multivariante es más complicada de realizar y existen relativamente pocos métodos para estimarla.

A nivel homoscedasticidad: Debe definirse de manera distinta según se estén analizando datos no agrupados (caso de una regresión lineal múltiple), o datos agrupados (caso de un análisis de la varianza de un factor). En el primer caso la hipótesis de homoscedasticidad puede definirse como el supuesto de que cada uno de los valores que puede tomar la distribución se mantiene constante para todos los valores de la otra variable continua.

En el caso de datos agrupados la homoscedasticidad implica que la varianza de la variable continua es más o menos la misma en todos los grupos que conforman la variable no métrica que es la que determina los grupos.

En resumen, se puede decir que la homoscedasticidad es la igualdad de varianzas entre las variables independientes:

Linealidad: el supuesto de linealidad es fundamental para todas aquellas técnicas que se centren en el análisis de las matrices de correlación o de varianzas-covarianzas, como el análisis factorial, regresión lineal o los modelos de ecuaciones estructurales. La razón es sencilla: el coeficiente de correlación de Pearson sólo podrá captar una relación si ésta es lineal. Si la relación existe y es intensa pero, por ejemplo, es curvilínea, el coeficiente de correlación de Pearson tomará un valor relativamente bajo y el investigador puede interpretarlo como ausencia de relación cuando, de hecho, ésta existe sólo que no es lineal.

Cuando la técnica empleada tiene una variable dependiente, como ocurre en el caso de la regresión lineal múltiple, existen diversos procedimientos para contras-

tar la linealidad de las relaciones basadas en el análisis de los residuos o residuales.

Métodos: Gráficos: de dispersión entre variables y Estadísticos: Coeficientes de correlación bivariados.

A partir de que la base de datos WEB diseño.sav, cumple los niveles requeridos de: cantidad, valores perdidos, atípicos, normalidad, homocedasticidad y linealidad, se planteó el problema 5: determinar la percepción que tienen los gerentes de las empresas del sector de mercadotecnia digital acerca de los servicios ofrecidos a sus consumidores medidos en 7 atributos (X_1 a X_7) por las siguientes razones:

- Entender el agrupamiento de estas percepciones sobre los 7 atributos generan una matriz de correlaciones que serán agrupadas en términos de lo que los gerentes de las empresas del sector de mercadotecnia digital tienen sobre los servicios ofrecidos a sus clientes.
- Reducir con el EFA las siete variables en el menor número de factores: Si las siete variables se pueden representar en un pequeño número de factores o dimensiones, entonces se eligió a la técnica multivariable correcta y las otras técnicas ya no serían correctas.

Nuevamente, son tomadas en cuenta las siguientes consideraciones:

- a.¿Cómo deben medirse las variables a utilizar? Deben ser métricas en general, pero pueden introducirse no métricas (variables Dummies), si bien no gozan de las mismas propiedades.
- b.¿Cuál es el tamaño de la muestra? No menos de 50 observaciones, aunque lo mejor es contar con más de 100 observaciones. No menos de cinco observaciones por cada variable, si bien el ratio óptimo es 10 a 1. En el caso de WEB Diseño.sav se tienen 100 observaciones y 7 variables, lo que da un ratio adecuado de 14 a 1. (100/7).
- c.No se considera determinante la falta de normalidad, homocedasticidad pero sí la linealidad. Así que se debe cumplir:
 1. Hay que asegurarse que existen suficientes correlaciones significativas entre las variables (incorrelación implica un solo factor por cada variable). Por linealidad 50%+1.
 2. Estas correlaciones deben ser importantes ($> a .3$) La diagonal de la matriz de correlación anti-imagen debe ser superior a .5.
 3. El test de esfericidad de Bartlett debe ser significativo (.01 ó 0.05 error / 99% o 95% confiabilidad).
 4. El KMO debe ser $> .5$ (lo ideal es $> a .7$)
 5. Su p debe ser inferior a los niveles críticos .05 o .01. Debe saberse, sin embargo, que es un test muy sensible a incrementos en el tamaño de la muestra. Cuando esta se incrementa es más fácil que encuentra correlaciones significativas.

6. Método de extracción con valores a 1; Varianza extraída superior al 60% por gráfico de sedimentación.

Deben cumplirse si no es así, el análisis factorial no será posible.

EFA: Resultados

Se sugiere ver de Mejía Trejo, (2017c: 53-71), donde se observa el procedimiento a detalle SPSS de análisis factorial a través de reducción de dimensiones- selección de variables métricas-estadísticos descriptivos- solución inicial con análisis de matrices de correlación y anti-imagen-método de extracción y rotación: Varimax con prueba de esfericidad El caso de estudio concluye en que hay problemas con las variables, por lo que se recomienda revisar las que se reportan en la matriz anti-imagen, que son: X_1 , X_2 , X_5 . La que se recomienda remover es la que tiene un valor bajo en la matriz anti- imagen: (.288). No significa que la variable no sirva. Al remover se requiere realizar TODO el proceso anterior para volver a revisar (figura 12.21. p.57-59, *op.cit*).

Una vez que se remueve X_5 y comprobado que el modelo con las variables restantes es correcto, se procede a hacer la agrupación de variables, a través del método de obtención de factores por componentes principales.

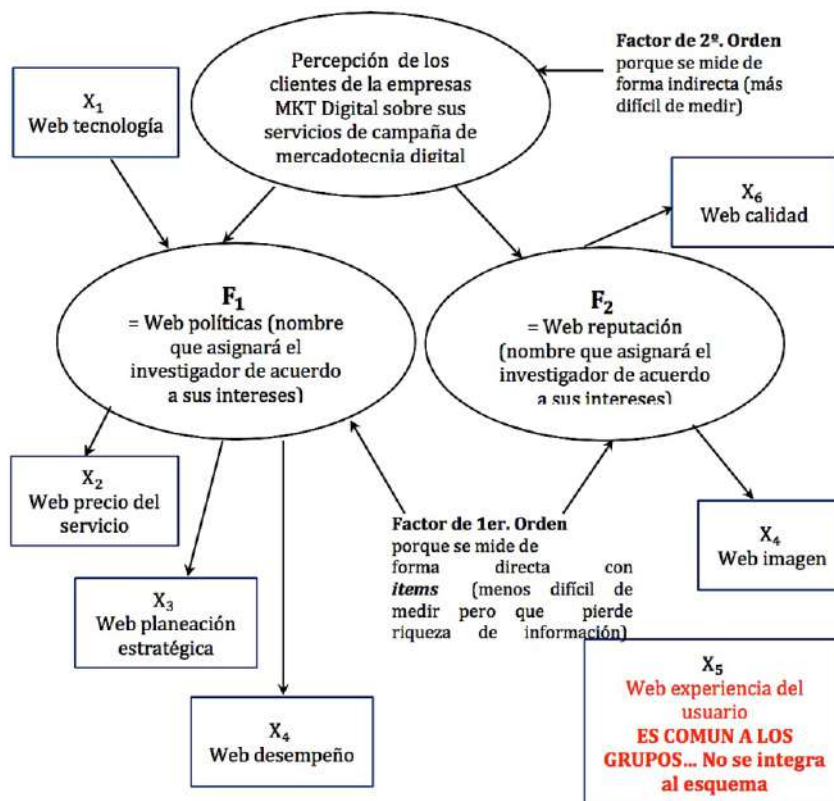
Importante considerar la prueba final de confiabilidad por el método de las mitades de las páginas 65-71 (*op cit.*).

Así, se tienen el caso de estudio final, mostrado en la figura 2.

EFA: Discusión y conclusión

- ▶ Se genera un nuevo modelo con reagrupación de variables (ítems) y que en vez de ser explicado con 7 variables se explica con 2 nuevos factores y 6 variables.
- ▶ Se excluye 1 variable que no aporta valor al modelo.
Los dos nuevos factores generados reciben el nombre que el investigador decida de acuerdo con los intereses del proyecto.
Los factores finales se dividen en: factores de 2o. Orden, dado que se mide de forma indirecta (la literatura sugiere evitar trabajar con este tipo de factores ya que es más difícil de medir, sin embargo, es capaz de verificar relaciones de estructura, dado que permite ver qué variable es más importante de los factores de 1er. Orden y/o cual de éstos aporta más al modelo) y factores de 1er. Orden porque se mide de forma directa con ítems (la literatura sugiere su uso ya que es menos difícil de medir pero, pierde riqueza de información).
- ▶ Nota: Si se deseara un factor de 1er. Orden volverlo variable (item) haciendo una media de sus variables para incorporarlo al modelo de factor de 1er.

FIGURA 2. Caso de estudio final.



Fuente: Mejía-Trejo (2017c: 63).

Orden a variable. Esto haría que el factor de 2o. Orden se vuelva incluso de 1er Orden.

- La variable X₅ se vuelve común (no se puede agregar) a los dos grupos y No se incluye en el modelo final, se considera que es básico, vital, primordial que brinde. Además, No puede haber factores con una sola variable (incorrelacionada), si se deseara tenerlo como constructo adicional. Lo que se elimina, es muy importante su discusión en el artículo a elaborar. Pudiera X₅ incluirse como factor 2o. Orden midiéndose en términos de % NO con escala de Likert. Posible estudiar con ecuaciones estructurales en 2o. Orden (estudios aún escasos al momento).
- Ya es posible publicar en revistas reconocidas y de impacto como estudio empírico. Posibles títulos:

- Construcción de un modelo empírico de percepción de los gerentes...
- La Construcción de un modelo teórico es 100% académico
- Siguiendo paso: buscar la teoría que justifique las nuevas relaciones los nuevos factores encontrados ya que estadísticamente está bien.
- ▶ Por análisis comparativo, se tienen resultados similares de cada una de las mitades respecto al modelo general.
- ▶ Lo anterior se refleja en : matriz de correlaciones, KMO y prueba de Bartlett, matriz anti-imagen, gráfico de sedimentación, comunalidades, varianza total explicada y matriz de componentes rotados Por lo tanto el modelo se presenta con mínimo error.
- ▶ Es posible aplicar ecuaciones estructurales para dar mayor explicación a las variables subyacentes,
- ▶ Para validar los resultados de un análisis factorial pueden emplearse dos métodos principales:
 - Uno de ellos es llevar a cabo un análisis factorial confirmatorio (CFA), mediante sistemas de ecuaciones estructurales (LISREL, AMOS, EQS).
 - Otro procedimiento puede ser separar la muestra en dos mitades aleatoriamente y llevar a cabo un análisis factorial con cada una de ellas.
- ▶ Si el análisis de las cargas factoriales no difiere sustancialmente, podemos concluir que los resultados son robustos y estables.

Resumen

Este capítulo abordó los siguientes problemas en el paso 3 del proceso de desarrollo de escala: prueba piloto de un conjunto de indicadores como reducción de ítems y validez inicial del procedimiento de prueba y realización de múltiples estudios para el desarrollo de la escala. En el segundo tema, destacamos:

- a. La importancia de incluir constructos para evaluar los diversos tipos de validez,
- b. El EFA sobre múltiples conjuntos de datos para refinar la escala y examinar una estructura de factores teóricos a priori, así como los
- c. Indicadores y análisis de confiabilidad.

Finalmente, ofrecemos un ejemplo de una escala desarrollada recientemente con el EFA y los análisis de ítems y confiabilidad pueden usarse en el desarrollo de la escala. El capítulo 7 se centra en los procedimientos de finalización de la escala.

CAPÍTULO 7

Finalizando la escala. Análisis factorial confirmatorio y ecuación estructural (etapa 4)

El análisis factorial confirmatorio hace referencia a las pautas de exploración de las relaciones entre varias variables. Estas pautas se representan por lo que se denominan como ponentes principales o, más comúnmente, factores. Como las variables ponderan fuertemente sobre un factor, se hacen descriptores de las dimensiones subyacentes. Sólo con el examen de las ponderaciones de las variables sobre los factores se identifica el carácter de la dimensión subyacente. Así, hay una similitud entre los objetivos del análisis factorial y el modelo de medida del SEM. Los factores son, en términos de los modelos medidos, las variables latentes. Cada variable actúa como indicador de cada factor (dado que cada variable tiene una ponderación para cada factor). Utilizado de esta forma, el análisis factorial es primordialmente una técnica exploratoria porque el investigador tiene un control limitado sobre las variables consideradas indicadores del constructo latente (es decir, aquellas variables ponderadas sobre cada factor). El SEM, sin embargo, tiene un papel confirmatorio porque el investigador tiene un control completo sobre la especificación de los indicadores de cada constructo. Además, el SEM permite un test estadístico de calidad del ajuste para la solución confirmatoria del factor, propuesta que no es posible con los componentes principales o el análisis de factor. El análisis de factor confirmatorio (CFA) es particularmente útil en la validación de las escalas para la medida de los constructos específicos.

ECUACIÓN 1. Representación de la ecuación estructural.

$$\begin{aligned} Y_1 &= X_{11} + X_{12} + X_{13} + \dots + X_{1n} \\ Y_2 &= X_{21} + X_{22} + X_{23} + \dots + X_{2n} \\ Y_m &= X_{m1} + X_{m2} + X_{m3} + \dots + X_{mn} \end{aligned}$$

(métrica) (métrica, no métrica)

El modelo de ecuaciones estructurales tienen una gran oportunidad de aplicación en las ciencias de la administración, ya que se tienen trabajos documentados en diversos campos que incluyen la mercadotecnia, la organización, la psicología, la educación, la psicología, la sociología, la gestión, la innovación, la salud, el comportamiento organizacional, la biología e incluso la genética. Las razones de su atractivo para áreas tan diversas son dobles: 1) proporciona un método directo de tratar con múltiples relaciones simultáneamente a la vez que se da eficacia estadística, y 2) su capacidad para evaluar las relaciones exhaustivamente y proporcionar una transición desde el análisis exploratorio al confirmatorio. Esta transición corresponde a los mayores esfuerzos en todos los campos de estudio hacia el desarrollo de perspectivas más sistemáticas y holísticas de los problemas. Tales esfuerzos exigen la capacidad de contrastar una serie de relaciones que constituyen un modelo a gran escala, un conjunto de principios fundamentales o una teoría entera. Estas son tareas para las que el modelo de ecuaciones estructurales se ajusta perfectamente.

Es una metodología estadística que parte de un enfoque confirmatorio (es decir, hipótesis-prueba) para el análisis de una teoría estructural, con base en un fenómeno a explicar algún fenómeno. Típicamente, esta teoría representa procesos “causales” que generan observaciones en múltiples variables (Bentler, 1988). En nuestro caso, una “causa” es un efecto directo de una variable sobre otra dentro del contexto de un modelo completo. Su magnitud y dirección están dadas por el coeficiente de regresión parcial. Si el modelo completo contiene todas las influencias relevantes en una variable dependiente dada, sus precursores causales son correctamente especificados. En la práctica, sin embargo, los modelos pueden omitir los predictores clave y pueden ser mal especificados, por lo que pueden ser inadecuados como un “modelo causal” en sentido estricto.

Por cierto, a lo largo de esta obra, los términos: *variables latentes, no observadas y no medidas se usan como sinónimos para representar un constructo o factor hipotético*; los términos: *variables observadas, manifestadas y medidas también se usan indistintamente*.

El término SEM transmite dos aspectos importantes del procedimiento:

- a. Que los procesos causales en estudio están representados por una serie de ecuaciones estructurales (es decir, regresión), y
- b. Que estas relaciones estructurales pueden modelarse gráficamente para permitir una conceptualización más clara de la teoría en estudio.

El modelo hipotético puede, por tanto, ser probado estadísticamente en un análisis simultáneo de todo el sistema de variables para determinar el grado en que es consistente con los datos. Si la bondad de ajuste (*goodness-of-fit*) es adecuada, el

modelo valida relaciones postuladas entre las variables; si es inadecuado, las relaciones son rechazadas. Varios aspectos del SEM que lo diferencian de otras generaciones anteriores de procedimientos multivariantes (tabla 1).

TABLA 1. Aspectos del SEM.

Ítem	Descripción
1	Como se señaló anteriormente, parte de un enfoque confirmatorio más que exploratorio para el análisis de datos (aunque aspectos de este último pueden abordarse). Además, al exigir que el patrón de relaciones se especifique <i>a priori</i> , el SEM se presta bien al análisis de datos con fines inferenciales. Por el contrario, la mayoría de otros procedimientos multivariantes son esencialmente descriptivos por naturaleza (por ejemplo, el EFA), por lo que la prueba de hipótesis es difícil, si no imposible.
2	Mientras que los procedimientos multivariantes tradicionales son incapaces de evaluar o corregir errores de medición, el SEM proporciona estimaciones explícitas de estos parámetros de varianza de error. De hecho, los métodos alternativos (por ejemplo, los que están basados en la regresión o el modelo lineal general (<i>general linear model</i>) suponen que el error en las variables explicativas (es decir, independientes) desaparece. Por lo tanto, aplicando esos métodos cuando hay un error en las variables explicativas equivale a ignorar el error que puede conducir, en última instancia, a inexactitudes serias, especialmente cuando los errores son considerables. Tales errores se evitan cuando se utilizan los análisis SEM correspondientes.
3	Mientras que los análisis de datos que usan los primeros métodos se basan en mediciones observadas sólo aquellos que usan procedimientos SEM pueden incorporar variables no observadas (es decir, latentes) y observadas.
4	Finalmente, no existen métodos alternativos ampliamente aplicados para el modelado de relaciones multivariantes, aunque sí para el SEM.

Fuente: Byrne (2006).

Dadas estas características altamente deseables, el SEM se ha convertido en una metodología muy popular para la investigación no experimental en las ciencias de la administración, en la que los métodos para probar las teorías no están bien desarrollados y las consideraciones éticas hacen que el diseño experimental sea inviable (Bentler, 1980). El SEM puede utilizarse de manera efectiva para abordar numerosos problemas de investigación que involucran investigación no experimental. Aquí se planteará un problema para su seguimiento en EQS 6.1 (Bentler, 2005; Bentler Wu, 2002), aunque existen casos diversos para su uso.

Todas las técnicas del SEM se distinguen por dos características:

- ▶ La estimación de relaciones de dependencia múltiples y cruzadas, y
- ▶ La capacidad de representar conceptos no observados en estas relaciones y tener en cuenta el error de medida en el proceso de estimación.

La diferencia más obvia entre el SEM y otras técnicas de relaciones multivariantes es el uso de relaciones distintas para cada conjunto de variables dependientes. En los términos más sencillos, el SEM estima una serie de ecuaciones de regresión múltiple distintas, pero interrelacionadas mediante la especificación del modelo estructural utilizado por el programa estadístico. En primer lugar, el investigador

utiliza la teoría, la experiencia previa y los objetivos de investigación para diferenciar qué variables independientes predicen cada variable dependiente. En nuestro ejemplo de los capítulos anteriores queríamos, en primer lugar, predecir la imagen del establecimiento. Además utilizar la imagen del establecimiento para predecir la satisfacción y utilizar a la vez las dos variables para predecir la fidelidad al establecimiento. Por tanto, algunas variables dependientes se convierten en variables independientes en relaciones ulteriores, dando lugar a la naturaleza interdependiente del modelo estructural. Además, muchas de las mismas variables afectan a cada una de las variables dependientes, pero con efectos distintos. El modelo estructural expresa estas relaciones entre variables dependientes e independientes, incluso cuando una variable dependiente se convierte en variable independiente en otras relaciones.

Las relaciones propuestas se trasladan a continuación a una serie de ecuaciones estructurales (parecidas a las ecuaciones de regresión) para cada variable dependiente. Esta característica sitúa al SEM en un lugar aparte de las técnicas discutidas previamente que trataban variables de dependencia múltiples (análisis multivariante de la varianza y de la correlación canónica) en la que sólo se permite una única relación entre las variables dependientes e independientes.

Sin embargo, es esencial primero revisar algunos conceptos clave asociados con la metodología.

CFA. Variables latentes vs. observadas

La estimación de relaciones múltiples interrelacionadas no es el único elemento de la modelización de ecuaciones estructurales. El SEM también tiene la habilidad de incorporar variables latentes al análisis. Una variable latente es un concepto supuesto y no observado que sólo puede ser aproximado mediante variables medibles u observables. Las variables observadas, que recogemos a través de varios métodos de obtención de datos (es decir, estudios de mercado, experimentos, observación) se conocen como *variables manifiestas*. Entonces, ¿por qué queríamos utilizar una variable latente no medida en lugar de datos exactos (variables directas, observables o manifiestas) que ofrecen los encuestados? Aunque esto pueda parecer que no tiene sentido o un enfoque de “caja negra” tiene justificaciones tanto prácticas como teóricas, al mejorar la estimación estadística, mejorar la representación de conceptos teóricos y tener en cuenta el error de medida.

En las ciencias de la administración los investigadores a menudo están interesados en estudiar constructos teóricos que no pueden ser observados directamente. Estos fenómenos abstractos se denominan variables latentes o factores. Ejemplos de variables latentes, por ejemplo, en psicología: son el autoco-

nocimiento y la motivación; en educación: capacidad verbal y expectativa de docentes; en economía: capitalismo y clase social. Como las variables latentes no se observan directamente, se deduce que no se pueden medir directamente. Así, el investigador debe definir operacionalizar la variable latente de interés en términos del comportamiento que se cree representarlo. Como tal, la variable no observada está vinculada a una que es observable, lo que hace que sea posible que sea medida. La evaluación del “comportamiento”, entonces, constituye la medición directa de una variable observable aunque también la medición indirecta de una variable no observada (es decir, el constructo subyacente). El término “comportamiento” se usa aquí en el sentido más amplio para incluir puntajes en un instrumento de medición particular. Por lo tanto, la observación puede incluir, por ejemplo, respuestas de autoinforme a una escala de actitud, puntajes en una prueba de logro, puntajes de observación *in vivo* que representan alguna tarea física o actividad, respuestas codificadas a preguntas de la entrevista, etc. Estas puntuaciones medidas (es decir, mediciones) se denominan observadas o variables manifiestas; dentro del contexto de la metodología SEM, sirven como indicadores del constructo subyacente que se supone que representan. Dado este necesario proceso de enlace entre variables observadas y variables latentes no observadas, ahora debería quedar claro por qué los metodólogos instan a los investigadores a ser circunspecto en su selección de medidas de evaluación. Mientras que la elección de los instrumentos psicométricos tal selección se vuelve aún más crítica cuando se presume que la medida observada representa un constructo subyacente.

CFA. Variables latentes exógenas vs. endógenas

Al trabajar con modelos SEM, es útil distinguir entre variables latentes que son exógenas y aquellos que son endógenos. *Las variables latentes exógenas* son sinónimo de variables independientes; ellos “causan” fluctuaciones en los valores de otras variables latentes en el modelo. Cambios en los valores exógenos de las variables no son explicadas por el modelo; más bien, se les considera influenciados por otros factores externos a él, por ejemplo, las variables de clasificación, como el género, la edad y el nivel socioeconómico. *Las variables latentes endógenas* son sinónimo de variables dependientes y, como tales, son influenciadas por las variables exógenas en el modelo, ya sea directa o indirectamente. La fluctuación en los valores, de las variables endógenas se explican por el modelo porque todas las variables latentes que las influyen se incluyen en la especificación del modelo. En términos de simplicidad y coherencia en todo el resto del capítulo, los términos independientes y dependientes en lugar de exógenos y endógenos. Las variables, respectivamente, se usan en los diferentes modelos SEM descritos.

CFA. El modelo de análisis factorial

El procedimiento estadístico más antiguo y mejor conocido para investigar las relaciones entre un conjunto de observaciones y las variables latentes son las de análisis factorial. Al usar este enfoque para el análisis de datos, el investigador examina la covariación entre un conjunto de variables observadas para recopilar información sobre sus constructos latentes subyacentes (es decir, factores). Existen dos tipos básicos de análisis factorial: análisis factorial exploratorio y el análisis factorial confirmatorio (para mayor información véase Byrne, 2005a, 2005b).

La EFA está diseñada para la situación en las que las relaciones entre las variables observadas y latentes son desconocidas o inciertas. El análisis, por lo tanto, se desarrolla en un modo exploratorio para determinar cómo y hasta qué punto las variables observadas están vinculadas a sus factores subyacentes. Típicamente, el investigador desea identificar un número mínimo de factores que subyacen (o explican) la covariación entre las variables observadas. Retomando el caso de estudio mencionado al final del capítulo 6, luego de la formulación de los ítems del cuestionario diseñados para medir los constructos latentes, el investigador realiza entonces un EFA para determinar el grado en que las mediciones de los indicadores (o sea, las variables observadas) se relacionaron con los constructos latentes. En el análisis factorial, estas relaciones son representadas por “cargas factoriales”. El investigador esperaría que los elementos diseñados para medir la percepción de los clientes de una empresa que realiza campañas de mercadotecnia digital, tenga diversas cargas factoriales en el resto de sus factores componentes (X_1 . Web tecnología; X_2 . Web precio del servicio; X_3 . Web planeación estratégica; X_4 . Web imagen; X_5 . Web experiencia del usuario; X_6 . Web calidad; X_7 . Web desempeño). Este enfoque analítico del factor se considera exploratorio en el sentido de que el investigador no tiene un conocimiento previo de que los indicadores realmente miden los factores previstos.

A diferencia de la EFA, el CFA se usa apropiadamente cuando el investigador tiene algún conocimiento de la estructura de la variable latente subyacente. Basado en el conocimiento de la teoría, la investigación empírica o ambas, el investigador postula las relaciones entre las medidas observadas y los factores subyacentes a priori y luego prueba esta estructura hipotética estadísticamente. Por ejemplo, en el ejemplo anterior, el investigador argumentaría con base en las cargas factoriales de los indicadores diseñados para medir la percepción de los clientes de una empresa que realiza campañas de mercadotecnia digital, que el indicador X_1 . Web tecnología tiene más impacto que el resto de los indicadores (si fuera el caso estadístico). En consecuencia, la especificación *a priori* del modelo CFA permitiría

que todos los indicadores de X_1 . Web tecnología sean libres de cargarse en dicho factor, y ser libres así también a tener cargas cero en los factores restantes. El modelo final, debería ser entonces evaluado por medios estadísticos para determinar la adecuación de su bondad de ajuste (*goodness-of-fit*) a los datos de la muestra (para discusiones más detalladas del CFA, ver, por ejemplo, Bollen, 1989).

En resumen, el modelo de análisis factorial (EFA o CFA) se centra únicamente en cómo y en qué medida las variables observadas están vinculadas con sus factores latentes subyacentes. Más específicamente, se enfoca a verificar el grado en que las variables observadas son generadas por los constructos latentes subyacentes y, por lo tanto, en la intensidad de la fuerza de las regresiones desde los factores a las variables observadas (es decir, las cargas factoriales) como interés primario. Aunque las relaciones entre los factores también son de interés, la estructura de regresión entre ellas no se considera en el modelo de análisis factorial. Dado que el modelo CFA se enfoca únicamente en el vínculo entre factores y sus variables medidas, dentro del marco del SEM, esto representa lo que se denomina *modelo de medición o medida*.

CFA. Modelo completo de variable latente

A diferencia del modelo de análisis factorial, el modelo completo de variable latente permite la especificación de estructura por regresión entre las variables latentes. Es decir, el investigador puede hipotetizar el impacto de un constructo latente sobre otro en el modelado de una dirección causal. Este modelo se denomina completo (*full o complete*) porque comprende tanto un modelo de medición como un modelo estructural. Así “el modelo de medición representa los enlaces entre las variables latentes y sus mediciones o medidas observadas (es decir, el modelo de CFA) y el modelo estructural representa los enlaces entre las mismas variables latentes”.

Un modelo completo de variable latente que especifica la dirección de la causa desde una única dirección se denomina “recursivo modelo” o “formativo”; uno que permite efectos recíprocos o de retroalimentación se denomina un “*modelo no recursivo*”. Sólo aplicaciones de modelos recursivos o formativos se consideran en este capítulo.

El CFA y el SEM. Modelaje estadístico de propósito general y proceso

Los modelos estadísticos proporcionan una manera eficiente y conveniente de describir la estructura latente subyacente de un conjunto de variables observadas.

Expresados ya sea, de forma esquemática o matemática a través de un conjunto de ecuaciones, tales modelos explican cómo las variables observadas y latentes están relacionadas entre sí. Típicamente, los investigadores postulan un modelo estadístico basado en su conocimiento de la teoría relacionada, en investigación empírica en el área de estudio, o en alguna combinación de ambos. Una vez que se especifica el modelo, el investigador prueba su veracidad basado en datos de muestra que comprenden todas las variables observadas en el modelo. La tarea principal en este procedimiento de prueba de modelo, es determinar la bondad de ajuste (*goodness-of-fit*) entre el modelo hipotético y los datos de muestra. De esta forma, el investigador impone la estructura del modelo hipotético en los datos de la muestra y luego prueba qué tan bien los datos observados se ajustan a esta estructura restringida. Dado que es altamente improbable que exista un ajuste perfecto entre los datos observados y el modelo hipotético, hay necesariamente un diferencial entre los dos; este diferencial se denomina “residual”. El proceso de ajuste del modelo (*model-fitting process*) puede resumirse de la siguiente manera:

$$\text{Datos} = \text{Modelo} + \text{Residual}$$

donde:

Los Datos representan mediciones de puntuación relacionadas tanto con las variables observadas como derivadas de personas que comprenden la muestra.

El Modelo representa la estructura hipotética que vincula las variables observadas con las variables latentes y, en algunos modelos, los que vinculan variables latentes particulares entre sí.

El Residual representa la discrepancia entre el modelo hipotético y los datos observados. Al resumir el marco estratégico general para probar modelos de ecuaciones estructurales, Jöreskog (1993) distinguió entre tres escenarios que denominó estrictamente confirmatorios (*strictly confirmatory*), modelos alternativos (*alternative models*) y generación de modelo (*model-generating*), (tabla 2).

CFA. Notación simbólica

Los modelos de ecuaciones estructurales se retratan esquemáticamente usando configuraciones particulares con cuatro símbolos geométricos (tabla 3).

En la construcción de un modelo de una estructura particular en estudio, los investigadores usan estos símbolos dentro de un marco de configuraciones básicas, cada una de las cuales, representa un componente importante en el proceso analítico (tabla 4).

TABLA 2. Escenarios de Jöreskog.

Escenario	Descripción
Estrictamente confirmatorio (<i>strictly confirmatory</i>)	El investigador postula un modelo único basado en la teoría, recopila los datos apropiados y prueba el ajuste del modelo hipotético a los datos de muestra. De los resultados de esta prueba, el investigador rechaza o no el modelo; no se realizan modificaciones al modelo planteado.
Modelos alternativos (<i>alternative models</i>)	El investigador propone varias alternativas (es decir, que compiten) de modelos, todos los cuales están basados en la teoría. Luego del análisis de un conjunto único de datos empíricos, el investigador selecciona un modelo como el más apropiado para representar los datos de muestra.
Generación de modelo (<i>model-generating</i>)	Representa el caso en el que el investigador, teniendo postulado y rechazado un modelo teórico sobre la base de su pobre ajuste a los datos de la muestra, procede de manera exploratoria (más bien que confirmatoria) a modificar y reestimar el modelo. El enfoque principal en esta instancia es localizar la fuente de "desadaptación" en el modelo y determinar uno que describa mejor los datos de la muestra. Jöreskog (1993) señala que, aunque la especificación pueda ser teórica o basada en datos, el objetivo final, es encontrar un modelo que sea sustantivamente significativo y estadísticamente apropiado. Aún más, postula que a pesar del hecho de que "se prueba un modelo en cada ronda, el enfoque global está en la generación de modelos, más que las pruebas de modelos" (<i>ibid.</i> : 295). Por supuesto, incluso una revisión superficial de la literatura empírica muestra claramente que este escenario es el más común de los tres, por una sencilla razón: los costos asociados con la recopilación de datos, donde es extremadamente raro que un investigador termine su trabajo sobre la base de un modelo hipotético rechazado. En consecuencia, el escenario estrictamente confirmatorio no es practicado en la realidad. Aunque el enfoque de los modelos alternativos tiene escasos exponentes.

Fuente: Jöreskog (1993).

TABLA 3. Notación utilizada en la representación de ecuaciones estructurales.




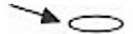
Símbolo	Representan:
Círculo o elipse	Factores latentes, subyacentes no observados
Cuadrados o rectángulos	Variables observables
Flechas en un solo sentido	El impacto de una variable sobre otra
Dobles flechas en dos sentidos	Covarianzas o correlaciones entre pares de variables

Fuente: Jöreskog (1993); Byrne (2006), con adaptación propia.

CFA. Diagrama de trayectoria

Las representaciones esquemáticas de los modelos se denominan diagramas de trayectoria porque proporcionan una descripción visual de las relaciones asumidas entre las variables en estudio. Esencialmente, como se verá más adelante, un diagrama de trayectoria que representa un modelo SEM, en particular, es en realidad el equivalente gráfico de su representación matemática, por lo que un

TABLA 4. Combinación de símbolos en configuración.

Configuración	Significa
	Coefficiente de trayectoria de regresión de una variable observada vs. una variable latente o no observada (o factor)
	Coefficiente de trayectoria de regresión de un factor vs. otro factor
	Medición de error asociado con una variable observada
	Error residual en la predicción de un factor no observable

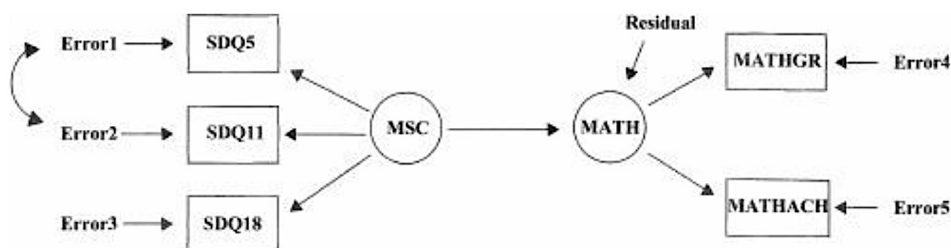
Fuente: Jöreskog (1993); Byrne (2006), con adaptación propia.

conjunto de ecuaciones relaciona las variables dependientes con sus variables explicativas. A fin de ilustrar cómo las cuatro configuraciones de símbolos anteriores pueden representar un proceso causal particular, mostramos la figura 1.

Al revisar este modelo, se observa que hay dos factores latentes no observados: el concepto de matemáticas (*msc. math self-concept*) y el logro en matemáticas (*MATH. math achievement*) con cinco variables observables, donde tres se consideran para medir el MSC (SDQ_5 ; SDQ_{11} ; SDQ_{18}) y dos para medir MATH ($MATHGR$; $MATHACH$). Estas cinco variables observables funcionan como indicadores de sus respectivos factores latentes subyacentes.

Asociado con cada variable observada está un término de error ($Error_1$ - $Error_5$) y, con el factor de predicción (MATH), un término residual (Residual). Hay una distinción importante entre estos dos términos de error. El error asociado con las variables observadas representa *el error de medición*, el cual se refleja en su adecuación con la medición de los factores subyacentes relacionados (MSC, MATH).

FIGURA 1. Modelo a discusión.



Fuente: Byrne (2006).

El error de medición se deriva de dos fuentes: el error de medición aleatorio (*random measurement error* en el sentido psicométrico) y el error de singularidad (*error uniqueness*), un término utilizado para describir el error de varianza que surge de alguna característica considerada específica (o única) para un particular indicador variante. Tal error a menudo representa un error de medición no aleatorio (*nonrandom measurement error*).

Los términos residuales representan el error en la predicción de factores dependientes a partir de factores independientes. Por ejemplo, el término residual que se muestra en la figura 1 representa el error en la predicción de MATH (el factor dependiente) de MSC (el factor independiente). En esencia, tanto los términos de error de medición como los de error residual representan variables no observadas. Por lo tanto, parece perfectamente razonable que, en consonancia con la representación de los factores, pudieran modelarse como elipses (o círculos). De hecho, este es el enfoque de modelado implementado en al menos una SEM programa: AMOS (Arbuckle, 2003). Además de los símbolos que representan las variables, otros se utilizan en los diagramas de trayectoria para indicar procesos hipotéticos que involucran a todo el sistema de variables. En particular, las flechas unidireccionales representan coeficientes de regresión estructural, lo que indica el impacto de una variable sobre otra.

En la figura 1, por ejemplo, la flecha unidireccional que apunta hacia el factor dependiente, MATH, implica que el factor independiente MSC “provoca” el logro matemático (MATH). Del mismo modo, las tres flechas unidireccionales desde MSC a cada una de las tres variables observadas (SDQ_5 , SDQ_{11} , SDQ_{18}) y las que van desde MATH a cada uno de sus indicadores (MATHGR y MATHACH) sugieren que estos valores de puntaje son cada uno influenciado por sus respectivos factores subyacentes. Como tal, estos coeficientes de trayectoria representan la magnitud del cambio esperado en las variables observadas para cada cambio en la variable latente relacionada (o factor). Las flechas unidireccionales que apuntan desde los cinco términos de error indican el impacto del error de medición (aleatorio y único) en las variables observadas y, a partir del residuo, el impacto del error en la predicción de MATH. Finalmente, como antes se señaló, las flechas curvas de dos vías representan covarianzas o correlaciones entre pares de variables. Así, las flechas bidireccionales enlazan $Error_1$ y $Error_2$, como se muestra en la figura 1, implica el error de medida asociado con SDQ_5 en correlación con lo asociado a SDQ_{11} .

CFA y SEM. La ecuación estructural

Es una poderosa técnica estadística multivariable cada vez más utilizada en las ciencias de la administración y en las ciencias sociales en general. Es el resultado de

la evolución y conjunción de varias aproximaciones metodológicas desarrolladas por diversas disciplinas: el análisis de variables latentes o no observables de la psicología, los modelos de ecuaciones simultáneas de la economía o el análisis de trayectorias (*path analysis*) de la sociología, entre otras. El SEM abarca una familia entera de modelos conocidos con muchos nombres, entre ellos, el análisis de la estructura de la covarianza, el análisis de variables latentes, el análisis de factor confirmatorio y a menudo simplemente el análisis LISREL (por cierto, nombre de uno de los programas de *software* más populares). El primer antecedente de un modelo de ecuaciones estructurales se remonta a 1934, año en el que el biómetra Sewall Wright da a conocer el análisis de trayectorias sobre las relaciones de tamaño en mediciones óseas. Esta técnica permitía descomponer la varianza y covarianza de las variables involucradas, en función de los parámetros de un sistema de ecuaciones simultáneas y tenía como fin estudiar el efecto directo e indirecto entre estas variables.

No obstante, su importancia, no es sino hasta 1960 y principios de los setenta que Blalock, Boundon y Duncan reconocen el potencial del análisis de trayectoria y las técnicas relacionadas de “correlación parcial”, como herramientas para analizar datos no experimentales. Un paso decisivo ocurre cuando Jöreskog (1993), Keesling (1972) y Wiley (1973) desarrollan un modelo general de ecuaciones estructurales e incorporan diagramas de trayectoria y otras características del análisis de trayectoria, conocido como modelo LISREL o modelo JKW. Además de facilitar la difusión del análisis de trayectoria, presenta las ecuaciones que se derivan de las covarianzas entre las variables, a través de operaciones matriciales en lugar de que se “lean” del diagrama de trayectoria, y proporciona una descomposición más clara de los efectos directos, indirectos y totales. El desarrollo de modelos con esta combinación de variables (latentes y medidas) se ha incrementado de forma espectacular. Jöreskog (1993) extendió el análisis factorial exploratorio al factorial confirmatorio; desarrolló el modelo factorial de segundo orden, el análisis factorial multigrupo y el ya citado modelo general de ecuación estructural (LISREL).

La influencia de Jöreskog no sólo se limita a los desarrollos propios. Varios de sus estudiantes de doctorado han realizado importantes contribuciones. Por ejemplo, Sörbom (1974) extiende el modelo multigrupo para incluir medias en las variables latentes; Muthén (1997) introduce métodos para incluir variables observadas categóricas; Hägglund (1985) contribuye con el método de mínimos cuadrados por medio de estimación de dos estados (*two-state last-square methods*); Quiroga (1992) realiza estudios de robustez con correlaciones policóricas para desviaciones del supuesto de normalidad, mientras que Yang-Wallentin (1997) desarrolla métodos para estimar relaciones no lineales. Los avances recientes en modelos de ecuaciones estructurales comprenden extensiones para estimaciones

en datos que provienen de muestras complejas, modelos lineales generalizados y series de tiempo.

En muchas disciplinas relacionadas con las ciencias sociales es usual intentar medir la inteligencia, motivación, eficiencia, percepción, habilidad verbal, etcétera (fenómenos de una gran complejidad), a partir de percepciones, opiniones, indicadores y variables relativas o aproximadas. Este tipo de variables recibe un nombre genérico: variables latentes. Las variables manifiestas pueden medirse de manera directa y representan características observables de algún fenómeno subyacente, al contrario de las latentes. Una característica importante de estas variables es que sirven para evidenciar o definir a las variables no observadas o latentes. La forma de modelar un fenómeno que requiere representar relaciones entre variables latentes y variables medidas o manifiestas es a través de los modelos de ecuaciones estructurales.

El modelo de ecuaciones estructurales es una herramienta estadística que busca explicar la relación existente entre diversas variables. Además, examina la relación estructural expresada en una serie de ecuaciones, similar a las series de regresión múltiple. La relación existente entre los constructos (variables dependientes e independientes) se analizan simultáneamente a través de un modelo de ecuaciones estructurales. Los constructos pueden ser factores no observables o variables latentes, quienes comúnmente están representados por múltiples variables o ítems (variables manifiestas).

El modelo de ecuaciones estructurales es la única técnica estadística que combina tanto las técnicas de dependencia como las técnicas de interdependencia: análisis factorial y análisis de regresión lineal.

Como se señaló en otros capítulos las variables exógenas son aquellas variables latentes equivalentes a las variables independientes, las cuales generalmente son medidas a través de una varianza y representan el constructo central de la investigación. Por lo tanto, en un diagrama de trayectoria de un SEM las variables exógenas son fáciles de reconocer porque de ellas salen las flechas que indican una relación existente con otro tipo de variables (endógenas), y no reciben ninguna flecha, por eso se llaman variables independientes.

Los constructos o *variables endógenas* son aquellas variables latentes multi-ítems equivalentes a las *variables dependientes*, las cuales están teóricamente determinadas por factores o dimensiones.

Este tipo de variables son dependientes de otros constructos y la relación de dependencia que tienen es fácil de visualizarse en un path model, porque son los constructos que reciben la flecha de la relación.

Como se señaló al inicio de este capítulo, además de prestarse a ser una presentación esquemática de los procesos causales bajo estudio, los modelos de

ecuaciones estructurales también pueden ser representados por una serie de ecuaciones de regresión (es decir, estructurales) ya que:

- ▶ Las ecuaciones de regresión representan la influencia de una o más variables sobre otra, y
- ▶ Esta influencia, convencionalmente conocida como SEM, es simbolizado por una flecha de una sola cabeza que apunta desde la variable de influencia a la variable de interés, por lo que se puede pensar en cada ecuación como un resumen del impacto de todas las variables relevantes en el modelo (observado y no observado) en una variable específica (observada o no).

Por lo tanto, un enfoque relativamente simple para la formulación de estas ecuaciones, consiste en observar cada variable que tiene una o más flechas apuntando hacia ella y luego registrar la suma de todas esas influencias para cada una de estas variables dependientes. Para ilustrar esta traducción de los procesos de regresión en ecuaciones estructurales, volvamos nuevamente a la figura 1. Podemos identificar seis variables con flechas apuntando hacia ellas; cinco representan las variables observadas (SDQ₅, SDQ₁₁, SDQ₁₈; MATHGR, MATHACH) y uno representa una variable no observada (o factor; MATH). Por lo tanto, sabemos que las funciones de regresión simbolizadas en el modelo mostrado en la figura 1 pueden ser resumido en términos de seis representaciones separadas de tipo ecuación de dependencias lineales, de la siguiente manera: $MATH = MSC + Residual$.

$$SDQ5 = MSC + ERROR1$$

$$SDQ11 = MSC + ERROR2$$

$$SDQ18 = MSC + ERROR3$$

$$MATHGR = MATH + ERROR4$$

$$MATHACH = MATH + ERROR5$$

La teoría estadística define un *coeficiente de regresión*, estar compuesto en realidad de dos elementos: el *verdadero* o coeficiente estructural entre la variable dependiente e independiente y la fiabilidad de la variable predictor. La fiabilidad es el grado en el que la variable independiente está “libre de error”. En todas las técnicas multivariantes expuestas hasta ahora, hemos asumido que no existía error en nuestras variables. Pero sabemos tanto desde la perspectiva práctica como teórica que no podemos medir perfectamente un concepto y que siempre hay algún grado de error de medida. Por ejemplo, cuando preguntamos acerca de algo tan inmediato como la renta de las economías domésticas, sabemos que algunas personas contestarán incorrectamente, tanto por exceso como por defecto o no lo sabrán con precisión. Las respuestas ofrecidas tienen cierto error de medida y por tanto afectan al verdadero coeficiente estructural. El impacto del error de medida

(y la correspondiente disminución de la fiabilidad) puede observarse de una expresión del coeficiente de regresión como lo mostrado en la ecuación 2.

ECUACIÓN 2. Coeficiente de regresión.

$$\beta_{v.x} = \beta_s \times \rho_x$$

donde:

$\beta_{v.x}$: es el coeficiente de regresión observado,

β_s : es el verdadero» coeficiente estructural y

ρ_x : es la fiabilidad de la variable predictor.

A menos que la fiabilidad sea del 100 por cien, la correlación observada siempre infravalora la relación correcta. Dado que todas las relaciones de dependencia se basan en la correlación observada (y resultante del coeficiente de regresión) entre las variables, esperemos se fortalezcan las correlaciones utilizando un modelo de dependencia parahacer estimaciones más precisas de los coeficientes estructurales teniendo en cuenta en primer lugar, la correlación atribuible a cualquiera de los problemas de medida.

El error de medición no sólo está provocado por respuestas inadecuadas sino que se produce también cuando utilizamos conceptos teóricos o más abstractos, tales como la actitud hacia un producto o las motivaciones del comportamiento. Con conceptos tales como éstos, el investigador intenta diseñar las mejores cuestiones para medir el concepto. Los encuestados también pueden no estar seguros de cómo responder o de cómo interpretar las cuestiones de forma diferente a la que pretende el investigador. Ambas situaciones pueden dar lugar al error de medida. Pero si sabemos la magnitud del problema, podemos incorporar la fiabilidad en la estimación estadística y mejorar nuestro modelo de dependencia.

¿Cómo tenemos en cuenta el error de medida? El SEM proporciona el modelo de medida, que especifica las reglas de correspondencia entre las variables latentes y manifiestas. El modelo de medida permite al investigador utilizar una o más variables para un único concepto dependiente o independiente y a continuación estimar (o especificar) la fiabilidad. Por ejemplo, la variable dependiente puede ser un concepto representado por un conjunto de cuestiones. En el modelo de medida el investigador puede evaluar la contribución de cada ítem de la escala, así como incorporar el grado en que la escala mide el concepto (su fiabilidad) en la estimación de las relaciones entre las variables dependientes e independientes. Este procedimiento es similar al desarrollo por el análisis factorial de los ítems de la escala y utilizar las puntuaciones del factor en la regresión.

CFA. Componentes no visibles de un modelo

Aunque en principio, hay una correspondencia uno a uno entre la presentación esquemática de un modelo y su traducción a un conjunto de ecuaciones estructurales, ninguna de estas representaciones nos informa de la historia completa del modelo. Algunos parámetros críticos para la estimación del modelo no se muestran explícitamente y por lo tanto pueden no ser obvios para el modelador de ecuaciones estructurales principiante. Por ejemplo, tanto en el diagrama de trayectoria como en las ecuaciones anteriores, no hay indicación de que las varianzas de las variables independientes sean parámetros en el modelo; de hecho, tales parámetros son esenciales para todos los modelos de ecuaciones estructurales. Del mismo modo, es igualmente importante notar la inexistencia especificada de ciertos parámetros en el modelo. Por ejemplo, en la figura 1, no hay una flecha curva entre Error4 y Error5, lo que sugiere la falta de covarianza entre los términos de error asociados con las variables observadas MATHGR y MATHACH. Del mismo modo, no existe una covarianza hipotética entre MSC y el residual. La ausencia de esta trayectoria provoca la suposición común de que la variable predictora (o independiente) no está de ninguna manera asociada con algún error que surja de la predicción de la variable de criterio (o dependiente).

Modelo SEM. Composición básica

A lo largo de la exposición del SEM, nos referiremos a la necesidad de una justificación teórica para la especificación de las relaciones de dependencia, modificaciones a las relaciones propuestas y muchos otros aspectos de la estimación de un modelo. La teoría ofrece la causa de casi todos los aspectos del SEM. La teoría puede definirse como un conjunto sistemático de relaciones que ofrecen una explicación exhaustiva y consistente de un fenómeno. A partir de esta definición, podemos ver que teoría no es el dominio exclusivo de los académicos sino que puede estar basada en la experiencia y en la práctica obtenida por la observación del comportamiento del mundo real. La teoría es a menudo un objetivo prioritario de la investigación académica, pero la investigación práctica puede desarrollar o proponer un conjunto de relaciones que son tan complejas e interrelacionadas como cualquier teoría de base académica. Por tanto, tanto los investigadores académicos como los de la empresa pueden beneficiarse de unas herramientas analíticas extraordinarias proporcionadas por el SEM. Desde una perspectiva práctica, un enfoque teórico del SEM es una necesidad porque la técnica debe ser casi completamente especificada por el investigador. Mientras que con otras técnicas multivariantes el investigador puede

haber sido capaz de especificar un modelo básico y permitir que valores por defecto de los programas estadísticos completen los resultados de estimación restantes, el SEM no tiene ninguna de estas características, por lo que debe definirse explícitamente cada componente de los modelos y medidas estructurales. Además, cualesquiera modificaciones del modelo deben hacerse a través de acciones específicas dictadas por el investigador. La necesidad de un modelo teórico para guiar el proceso de estimación se hace especialmente crítica cuando se hacen modificaciones del modelo. Dada la flexibilidad del SEM, las oportunidades de sobreajustar el modelo o desarrollar un modelo con escasa generalización son muy elevadas. Por tanto, cuando resaltamos la necesidad de la justificación teórica, nuestro objetivo es que el investigador reconozca que el SEM es un método confirmatorio, guiado más por la teoría que por los resultados empíricos.

Uno de los conceptos más importantes que un investigador en ciencias de la administración debe aprender en relación con las técnicas multivariantes es que no existe un único modo “correcto” de aplicarlas. En lugar de eso, el investigador debe formular los objetivos de la investigación y aplicar la técnica de la forma más apropiada para conseguir los objetivos deseados. En algunos casos, las relaciones están especificadas estrictamente y el objetivo es una confirmación de la relación. Otras veces, las relaciones están vagamente reconocidas y el objetivo es el descubrimiento de las relaciones. En cada caso límite y en los casos intermedios, el investigador debe formular el uso de la técnica de acuerdo con los objetivos de investigación. La aplicación del SEM sigue este mismo principio. Su flexibilidad ofrece al investigador una poderosa herramienta de análisis apropiada para muchos objetivos de investigación. Pero el investigador, debe definir estos objetivos como líneas maestras de la estrategia de modelización. El uso del término *estrategia* se refiere a un plan de acción orientado a la consecución de un objetivo específico. En el caso del SEM, el resultado último es siempre la evaluación de una serie de relaciones. Sin embargo, esto se puede conseguir de muchas maneras. Para nuestros propósitos, definimos tres estrategias distintas en la aplicación del SEM: modelización confirmatoria, modelos rivales y desarrollo del modelo (cuadro 1).

El modelo SEM general se puede descomponer en dos submodelos:

- ▶ El modelo de medición, que define las relaciones entre las variables observadas y las no observadas. En otras palabras, proporciona el vínculo entre las puntuaciones en un instrumento de medición (es decir, las variables indicadoras observadas) y los constructos subyacentes que están diseñados para medir (es decir, las variables latentes no observadas).
- ▶ El modelo de medición, entonces, representa el modelo CFA descrito anteriormente en que especifica el patrón por el que cada medida (o medición) carga a un factor particular. El modelo estructural, por el contrario, define las

TABLA 1. Estrategias de aplicación del SEM.

Estrategia	Descripción
Modelización Confirmatoria	<p>La aplicación más directa de los modelos de ecuaciones estructurales es una estrategia de modelización confirmatoria, donde el investigador especifica un modelo aislado y SEM se utiliza para evaluar su significación estadística. Aquí el investigador está diciendo, <i>tanto si funciona como si no</i>. Aunque pueda parecer que es la aplicación más rigurosa, en realidad no es el test más riguroso del modelo propuesto. La investigación ha mostrado incluso que las técnicas desarrolladas para evaluar los modelos de ecuaciones estructurales tienen un <i>sesgo confirmatorio</i>, que tiende a confirmar que el modelo se ajusta a los datos. Por tanto, si el modelo propuesto tiene un ajuste aceptable por cualquiera de los criterios aplicados, el investigador no ha <i>probado</i> el modelo propuesto, sino que sólo ha confirmado que es uno de los varios modelos posibles aceptables. Varios modelos pueden tener ajustes igualmente aceptables. Por tanto, el test más riguroso se consigue comparando modelos rivales.</p>
Modelos rivales	<p>Obtener un nivel de ajuste aceptable tanto para el modelo conjunto como para los modelos de medida o estructural no asegura al investigador que se haya encontrado el <i>mejor</i> modelo. Varios modelos alternativos pueden ofrecer iguales o incluso mejores ajustes. Como medio de evaluar el modelo estimado con modelos alternativos, las comparaciones de modelos conjuntos pueden llevarse a cabo en una estrategia de modelos rivales. El test más fuerte de un modelo propuesto es identificar y contrastar los modelos rivales que representan las verdaderas y distintas relaciones hipotéticas estructurales. Cuando se comparan estos modelos, el investigador se acerca mucho al contraste de teorías alternativas, que es un contraste mucho más fuerte que una ligera modificación de una única teoría. ¿Cómo genera el investigador este conjunto de modelos alternativos? Un posible origen de modelo rivales son las formulaciones alternativas de la teoría subyacente. Por ejemplo, en una formulación la confianza puede preceder a la lealtad mientras que en la otra la lealtad precede a la confianza. Esto podría constituir la base de dos modelos rivales. Los modelos equivalentes proporcionan una segunda perspectiva sobre el desarrollo de un conjunto de modelos rivales. Se ha mostrado que para cualquier modelo de ecuaciones estructurales, existe al menos otro modelo con el mismo número de parámetros y el mismo nivel de ajuste del modelo que varía en las relaciones representadas. Esto implica que ningún modelo es único en el nivel de ajuste conseguido, y que para cualquier modelo con un ajuste aceptable existe un número indeterminado de modelos alternativos con el mismo nivel de ajuste. Se han definido una serie de <i>nonnas</i> para identificar modelos equivalentes para cualquier modelo estructural. Como <i>nonna</i> general, cuanto más complejo sea el modelo, más modelos equivalentes existen. Una tercera aproximación, el programa TETRAD, es un método empírico que está tomando auge y que examina sistemáticamente un modelo estructural e identifica relaciones adicionales que están apoyadas por los datos. Con la matriz de datos de entrada y la especificación del modelo hecha por el investigador, el programa examina las pautas de las relaciones (tetrads) y aísla aquellas relaciones que pudieran estar empíricamente fundamentadas. El programa TETRAD no estima los parámetros, sino que identifica las relaciones que se van a incluir en el modelo original para formar los modelos rivales. El programa funciona mejor empezando con un modelo simple y añadiendo relaciones. Se suele alegar que este modelo es <i>ateórico</i> y muy <i>mecanicista</i> o de <i>caja negra</i>, pero ofrece a los investigadores intuiciones para sus modelos que no se podrían haber obtenido de otra forma. Un ejemplo habitual de la estrategia de modelos rivales es el proceso de evaluación factorial de invarianza, la igualdad de modelos de factor entre grupos. Existe un procedimiento establecido para evaluar el grado de invarianza, comenzando por los modelos más vagamente</p>

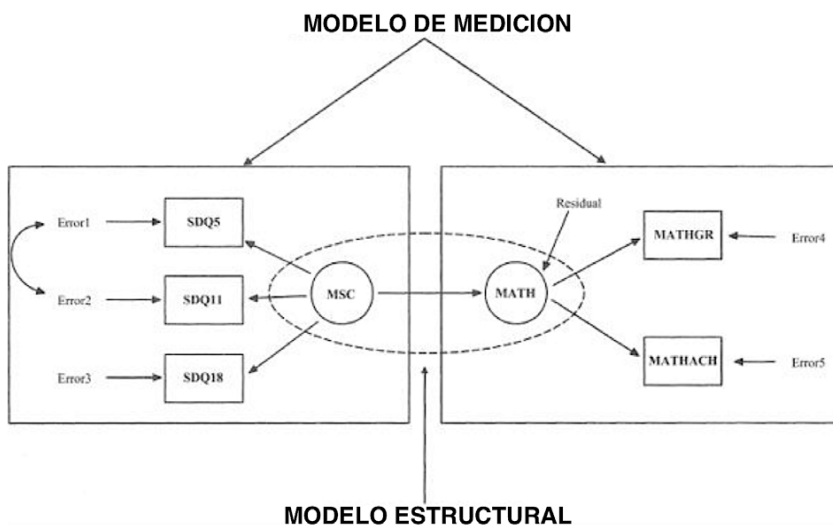
Estrategia	Descripción
	restringidos y a partir de esos modelos añadir paulatinamente restricciones adicionales hasta que se contraste el modelo más restrictivo. Las restricciones se añaden para representar invarianzas a lo largo de grupos, ponderaciones e incluso intercorrelaciones de factores. Esto también es un ejemplo de un enfoque de modelos anidados , en el que el número de constructos e indicadores permanece cons-tante, pero el número de relaciones estimadas cambia. Aunque los modelos rivales son normalmente modelos anidados, también pueden no estar anidados (difieren en el número de constructos o indicadores), con lo que se exigen medidas especializadas de ajuste para comparar entre los modelos.
Desarrollo del modelo	La estrategia de desarrollo del modelo difiere de las dos anteriores estrategias en que aunque se propone un modelo, el propósito del esfuerzo de modelización es mejorarlo a través de modifica- ciones de los modelos de medida y/o estructurales. En muchas aplicaciones, la teoría sólo puede ofrecer un punto de partida para el desarrollo de un modelo con justificación teórica que pueda ser apoyado empíricamente. Por tanto, el investigador ha de emplear el SEM no sólo para contrastar el modelo empíricamente sino también para obtener perspectivas acerca de su reespecificación. Debe tomarse, sin embargo, alguna precaución. El investigador tiene que ser cuidadoso no em pleando esta estrategia en la medida en que el modelo final tenga un ajuste aceptable pero que no pueda ser generalizada a otras muestras o poblaciones. Además, la reespecificación del modelo debe hacerse siempre con apoyo teórico en lugar de justificación empírica.

Fuente: Jöreskog (1993); Hair (*et al.*, 1999); Byrne (2006), con adaptación propia.

relaciones entre las variables no observadas. En consecuencia, especifica la manera particular en que las variables latentes, directa o indirectamente influyen (es decir, “causan”) cambios en los valores de ciertas otras variables latentes en el modelo.

Con fines didácticos para aclarar este importante aspecto de la composición SEM, examinemos ahora la figura 2, el cual, es el mismo modelo presentado en la figura 1 del capítulo 1 pero que ha sido demarcado en medición y componentes estructural. Considerados por separado, Los elementos modelados dentro de cada rectángulo en la figura 2 del capítulo 1 representan dos modelos CFA. Los dos factores dentro de la elipse representan un modelo de variable latente completa y, por lo tanto, no sería de interés en la investigación del CFA. El modelo CFA a la izquierda del diagrama representa un modelo de un solo factor (MSC), medido por tres variables observadas (SDQ_5 - SDQ_{18}), mientras que el modelo CFA a la derecha representa un modelo de un factor (MATH) medido por dos variables observables (MATHGR-MATHACH). En ambos casos, la regresión de las variables observadas en cada factor y las varianzas tanto del factor como los errores de medición son de interés primario; la covarianza de error sería de interés sólo en los análisis del modelo CFA relacionado con MSC.

FIGURA 2 . Ecuación general estructural demarcada en sus componentes: medición y estructural .



Fuente: Jöreskog (1993); Hair (*et al.*,1999); Byrne (2006), con adaptación

El verdadero valor del SEM proviene de los beneficios de utilizar los modelos estructurales y de medida simultáneamente, jugando cada uno de ellos distintos papeles en el análisis conjunto. Para asegurar que ambos modelos están correctamente especificados y que los resultados son válidos, discutiremos ahora un proceso de modelización de ecuaciones estructurales: (1) desarrollar un modelo fundamentado teóricamente, (2) construir un diagrama de secuencias de relaciones causales, (3) convertir el diagrama de secuencias en un conjunto de modelos y relaciones estructurales, (4) elegir el tipo de matriz de entrada y estimar el modelo propuesto, (5) evaluar la identificación del modelo estructural, (6) evaluar los criterios de calidad del ajuste, y (7) interpretar y modificar el modelo si está teóricamente justificado (Hair *et al.*, 1999), (cuadro 2).

CFA y EQS. Formulación de estructuras de covarianza y medias

Los parámetros centrales en los modelos de ecuaciones estructurales que se centran en el análisis de las estructuras de covarianza son los coeficientes de regresión y las varianzas y covarianzas de las variables independientes. Sin embargo, dado que los datos de la muestra comprenden sólo puntajes observados, es necesario que haya una evaluación de mecanismo interno, mediante el cual los

CUADRO 2 . Etapas sugeridas para modelar el CFA y el SEM.**Paso 1. Desarrollo de un modelo basado en la teoría**

La modelización de ecuaciones estructurales se basa en relaciones causales, en las que el cambio en una variable se supone que produce un cambio en otra variable. Encontramos este tipo de enunciado cuando definimos una relación de dependencia, tal y como se encuentra en el análisis de regresión. Las relaciones causales pueden tomar muchas formas y significados, desde la estricta causación encontrada en los procesos físicos, tales como una reacción química, a las relaciones menos definidas encontradas en la investigación del comportamiento, tales como las «causas» del éxito educativo o las razones por las cuales compramos un producto y no otro. La fuerza y convicción con que el investigador puede asumir la causal entre dos variables no descansa en los métodos analíticos escogidos sino en la justificación teórica ofrecida por los análisis. Los «requisitos» para hacer valer la causación tienen profundas raíces en varios enfoques de la filosofía de la ciencia. Hay un acuerdo general con al menos cuatro criterios establecidos para realizar afirmaciones causales: (1) asociaciones suficientes entre dos variables, (2) antecedentes temporales de la causa frente al efecto, (3) falta de alternativas a las variables causales, y (4) una base teórica para la relación. Aunque en muchos casos todos los criterios establecidos para la realización de afirmaciones causales no se cumplen estrictamente, probablemente puedan realizarse afinaciones causales si las relaciones están basadas en una teoría racional. Pero avisamos a cualquier investigador contra la suposición de que las técnicas discutidas en este capítulo ofrezcan por sí mismas un medio de probar la causal sin tener una perspectiva teórica como guía. Utilizar estas técnicas de forma exploratoria es inútil y equivoca al investigador más que ofrecerle resultados apropiados. El error más crítico en el desarrollo de modelos de fundamentación teórica es la omisión de una o más variables predictivas claves, un problema conocido como error de especificación. La consecuencia de omitir una variable significativa consiste en sesgar la evaluación de la importancia de otras variables. Por ejemplo, si suponemos que dos variables (a y b) fueran predictores de c e incluyéramos tanto a como b en nuestro análisis, haríamos la evaluación correcta de su importancia relativa como se muestra por sus coeficientes estimados. Pero si dejamos a la variable b fuera de nuestro análisis, el coeficiente para a sería diferente. Esta diferencia, o sesgo, es el resultado del coeficiente para la variable a que refleja no sólo su efecto sobre c sino el efecto que comparte también con b. Este efecto compartido, sin embargo, está controlado cuando ambas variables están incluidas en el análisis. (ver más sobre el manejo de errores de datos en Mejía Trejo, 2017d). El deseo de incluir todas las variables debe compensarse con las limitaciones prácticas de SEM. Aunque no existe un límite teórico sobre el número de variables del modelo, los problemas prácticos aparecen incluso antes de llegar a los límites de la mayoría de los problemas informáticos. Muy a menudo, las interpretaciones de los resultados, particularmente la significación estadística, se hace bastante difícil a medida que el número de conceptos aumenta (más de 20 conceptos). El investigador nunca debería omitir un concepto solamente porque el número de variables se haga grande, pero también debería reconocer los beneficios de los modelos teóricos concisos y parsimoniosos.

Paso 2. Construcción de un diagrama de trayectorias de relaciones causales

Hasta el momento, hemos expresado las relaciones causales sólo en términos de ecuaciones. Pero existe otro método de representar estas relaciones denominadas diagramas de trayectorias, que son especialmente útiles en el diseño de una serie de relaciones causales. Un diagrama de trayectorias es más que una simple representación visual de las relaciones porque permite al investigador presentar no sólo las relaciones predictivas entre constructos (es decir, las relaciones variable dependiente-independiente), sino también relaciones asociativas (correlaciones) entre los constructos e incluso entre los indicadores.

Elementos de un diagrama de secuencias

Recuerde los dos elementos básicos utilizados en su construcción. Primero, es el concepto de un constructo, que es un concepto teórico que actúa como una piedra angular utilizada para definir relaciones. Un constructo puede representar un concepto tan simple como edad, renta o género o tan complejo como estatus socioeconómico, conocimiento, preferencia o actitud. El investigador en

ciencias de la administración define diagramas de trayectoria en términos de constructos y a continuación encuentra variables para medir cada constructo. Por ejemplo, podemos preguntar la edad a alguien y utilizar esto como medida del constructo edad. De la misma manera, podemos preguntar una serie de cuestiones acerca de las opiniones de una persona y utilizar esto como una medida de actitud. Ambas series de cuestiones ofrecen valores numéricos para los constructos. Podemos evaluar las cuestiones para la cantidad de error medido que poseen e incluir esto en el proceso de estimación. A partir de este momento, utilizamos el término constructo para representar un concepto particular, no importa cómo se mida. Un constructo se representa normalmente en un diagrama de secuencias mediante un óvalo. El segundo elemento básico es la flecha, utilizada para representar relaciones específicas entre constructos. Una flecha directa indica una relación causal directa de un constructo a otro. Una flecha curvada (o una línea sin cabezas de flechas) entre constructos indica simplemente una co-relación entre constructos. Finalmente, una flecha directa con dos cabezas indica una relación recíproca o no recursiva entre constructos. Sólo con estos tres tipos de flechas, se pueden representar todas las relaciones de los modelos estructurales. Todos los constructos de un diagrama de trayectorias pueden clasificarse en dos clases: exógenos y endógenos. Los constructos exógenos, también conocidos como variables origen o variables independientes, no están causados o son explicados por ninguna de las variables del modelo; esto es, no existen flechas apuntando a estos constructos. Los constructos endógenos se predicen mediante uno o más constructos pueden predecir otros constructos endógenos (aquí es donde vemos las interrelaciones que apuntan a la necesidad de los modelos estructurales), pero un constructo exógeno puede estar causalmente relacionado sólo con los constructos endógenos. Por tanto, la distinción entre exógenos y endógenos los determina sólo el investigador, al igual que se hizo al decidir sobre cuáles son variables dependientes e independientes en la regresión.

Dos hipótesis subyacen en los diagramas de trayectorias. En primer lugar, todas las relaciones causales están indicadas. Por tanto, la teoría es la base para la inclusión o la omisión de cualquier relación. Es tan importante justificar por qué no existe una relación causal entre dos constructos como justificar la existencia de otra relación. Pero es importante recordar que el objetivo es modelizar las relaciones entre constructos con el número más reducido de secuencias causales o correlaciones entre constructos que pueden estar teóricamente justificados. El segundo supuesto se refiere a la naturaleza de las relaciones causales que se supone que son lineales. Al igual que se encontró en otras técnicas multivariantes, **las relaciones no lineales no pueden ser estimadas directamente en la modelización de ecuaciones estructurales**, pero modelos estructurales modificados pueden aproximar relaciones no lineales.

Paso 3. Conversión de un diagrama de secuencias en un conjunto de ecuaciones estructurales y especificación del modelo de medida

Después de desarrollar el modelo teórico y de representarlo en un diagrama de trayectorias, el investigador está preparado para especificar el modelo en términos más formales. Esto se hace a través de una serie de ecuaciones que definen: (1) las ecuaciones estructurales que vinculan los constructos, (2) el modelo de medida que especifica qué variables miden qué constructos y (3) una serie de matrices que indican cualquier correlación supuesta entre constructos o variables. El objetivo es vincular definiciones operacionales de los constructos con la teoría para llegar al contraste empírico apropiado. En esta sección se discute cada tipo de ecuación o matriz. En las discusiones, todas las ecuaciones se especifican en terminología general.

Modelo estructural

Trasladar un diagrama de trayectorias a una serie de ecuaciones estructurales es un procedimiento directo. En primer lugar, cada constructo endógeno (cualquier constructo con una o más flechas directas hacia él) es la variable independiente en una ecuación separada. Entonces las variables predictor son todos constructos en los principios o colas, de las flechas que apuntan a la variable endógena. Es así de simple.

Modelo de medida

Hasta ahora nos hemos referido al término de medida en términos generales, pero ahora debemos definirlo en términos específicos. Discutiremos no sólo los procedimientos básicos de especificación de un modelo de medida, sino también los supuestos referentes al número de indicadores por constructo y el proceso de especificación de la fiabilidad del constructo en lugar de en su estimación. Pero antes de hacerlo, revisemos los fundamentos del análisis factorial, que son muy parecidos a los del modelo de medida.

Correspondencia con el análisis factorial

En el análisis factorial, cada variable individual se «explicaba» por su ponderación en cada factor. El objetivo era representar lo mejor posible todas las variables en un número reducido de factores, es decir, los factores referidos dimensiones subyacentes de los datos, que después tendremos que interpretar y clasificar. El análisis factorial (Mejía-Trejo 2017c, capítulo 12), a menudo se clasifica como una técnica exploratoria porque no existen restricciones sobre las cargas de las variables. Cada variable tiene una carga sobre cada factor. El valor de cada factor (puntuación del factor) se calcula mediante las cargas sobre cada variable. También el valor predictor para cada variable se calcula mediante las cargas de la variable para cada factor. Sin embargo, cada variable tiene una carga factorial; por tanto, cada factor es siempre una composición de todas las variables, aunque sus cargas varíen en magnitud. Por consiguiente, un factor es en realidad un constructo latente, definido por las cargas de todas las variables.

Especificación del modelo de medida

Para especificar el modelo de medida, hacemos la transición desde el análisis factorial, en el que el investigador no tiene el control sobre qué variables describen cada factor, a un modo confirmatorio, en el que el investigador especifica qué variables definen cada constructo (factor). Las variables observadas que obtenemos de los encuestados se denominan indicadores en el modelo de medida, porque los utilizamos para medir o indicar, los constructos latentes (factores).

¿Cómo y por qué difiere esta configuración de las cargas del análisis factorial examinadas previamente? La diferencia más evidente es el más reducido número de ponderaciones. En el modelo explicativo del análisis factorial, el investigador no puede controlar las ponderaciones. En el modelo de medida sin embargo, el investigador tiene un control completo sobre las variables descritas por cada constructo. En el ejemplo, cada variable era un indicador de un solo constructo; por tanto, existe un número más reducido de ponderaciones. Aunque una variable puede ser indicador de más de un constructo, este método no es recomendable excepto en situaciones específicas con fuerte carga teórica. El investigador especifica un modelo de medida tanto para los constructos exógenos como para los constructos endógenos exactamente de esta forma.

Determinación del número de indicadores

Ya se ha discutido la lógica y la justificación del uso de indicadores múltiples (variables) para representar un constructo, pero queda una cuestión fundamental: ¿cuántos indicadores deberían utilizarse por constructo? El número mínimo de indicadores por constructo es uno, pero el uso de sólo un único indicador exige al investigador dar estimaciones de fiabilidad. Un constructo puede ser representado por dos indicadores, pero **3 es el número mínimo preferido de indicadores, porque utilizando sólo dos indicadores aumentan las oportunidades de alcanzar una solución no factible**. Además de los fundamentos teóricos que deberían utilizarse para seleccionar las variables como indicadores de un constructo, no existe límite superior en términos de número de indicadores. A efectos prácticos, sin embargo, la mayoría de los constructos deberían ser representados por un número de entre **5 a 7 y indicadores**. La excepción notable es el uso de escalas preexistentes, que pueden contener muchos ítems, cada uno actuando como indicador del constructo. En estos casos, el investigador debería evaluar la unidimensionalidad del constructo y la posibilidad de múltiples subdimensiones que puedan ser representadas en un modelo factorial de segundo orden.

Justificar la fiabilidad del constructo

Una vez que el modelo de medida ha sido especificado, el investigador debe probar después la

fiabilidad de los indicadores. En este paso, el investigador debe determinar el método básico en que se establece la fiabilidad de cada constructo. Existen dos métodos principales para establecer la fiabilidad: (1) estimación empírica o (2) especificación por el investigador.

Fiabilidades estimadas empíricamente

La estimación empírica de la fiabilidad sólo es posible si el constructo tiene dos o más indicadores. Para un constructo con un sólo indicador, el investigador debe especificar la fiabilidad. Para la estimación empírica, el investigador especifica la matriz de ponderaciones tal y como se ha descrito, junto con un término de error para cada variable indicador). Cuando los modelos de medida y estructural están estimados, los coeficientes de ponderación ofrecen estimaciones de las fiabilidades de los indicadores y del constructo conjunto. En este enfoque, la intervención del investigador no tiene impacto sobre el valor de la fiabilidad utilizado en la estimación del modelo excepto en los conjuntos de indicadores incluidos.

Especificar las confiabilidades

En algunos casos resulta apropiado especificar o fijar las confiabilidades. La especificación de las confiabilidades para el indicador(es) de cualquier constructo latente puede parecer que va en contra de los objetivos de la modelización de ecuaciones estructurales; sin embargo, en al menos 3 situaciones está justificado y es altamente recomendable. En un caso, la estimación empírica de la confiabilidad no es posible, aunque el investigador puede saber que el error de medida todavía existe. En otros, los indicadores pueden haber sido utilizados previamente; por tanto, se conocen las confiabilidades antes de utilizarse. Y finalmente, tenemos un enfoque en 2 pasos en el que las confiabilidades se evalúan en primer lugar y a continuación se especifican en el proceso de estimación. Este enfoque de dos pasos separa explícitamente los 2 procesos empíricos y ofrece perspectivas de los dos por separado.

Medidas de ítem único

Con medidas de ítem único, no es posible estimar empíricamente la confiabilidad; por tanto, el investigador se enfrenta con dos posibilidades. La primera, establecer (fijar) la confiabilidad a 1,0, indicando que no existe error de medida en el indicador. Pero como ya hemos discutido, sabemos que esto es erróneo en casi todos los casos, si por ninguna otra razón la confiabilidad está afectada por la calidad de la recogida de datos. Por ejemplo, el género puede ser perfecto o muy cercano (99 por ciento), con errores debidos a errores de codificación. Sin embargo, los ingresos pueden tener un nivel elevado de error (es decir, un 10 por ciento) debido a sesgos de las contestaciones y al nivel de medida. Muy a menudo por tanto, el investigador debería realizar alguna estimación de la confiabilidad y especificar el valor de los indicadores de ítem único.

Uso de medidas o escalas validadas

Muchas veces el investigador emplea una escala o medida que ha sido extensamente contrastada en la investigación previa. Si al usarla el objetivo es obtener una réplica de los efectos encontrado en estudios anteriores, entonces la confiabilidad de la medida o escala debería fijarse en niveles previamente especificados. Este es un ejemplo de las confiabilidades especificadas por el investigador para mantener el control sobre el significado de los constructos. Al fijar la confiabilidad, el investigador fuerza a un indicador para tener la cantidad de varianza apropiada para el constructo y mantener un significado específico para el mismo.

Análisis de dos pasos

Muchos investigadores proponen un proceso de dos pasos de modelización de ecuaciones estructurales en el que el modelo de medida se fija en el segundo paso cuando el modelo estructural está estimado. La razón de este enfoque es que la representación precisa de la fiabilidad de los indicadores se realiza mejor en dos pasos para evitar la interacción de los modelos de medida y estructurales. Aunque no podemos evaluar verdaderamente los modelos de medida y estructural aisladamente, debemos considerar el potencial para den- tro

del constructo frente a los efectos entre constructos en estimación, que pueden ser sustanciales y resultar en lo que se denomina “*confusión interpretacional*”. Un análisis de dos pasos con estimación simultánea tanto de modelos de medida como estructurales es la mejor aproximación cuando el modelo posee tanto fuerte lógica teórica como medidas altamente confiables, resultando en relaciones más precisas y disminuyendo la posibilidad de interacción. Sin embargo, cuando nos enfrentamos con medidas que son menos fiables, o una teoría que es sólo tentativa, el investigador debería considerar un enfoque por pasos para maximizar la interpretabilidad tanto de las medidas como de los modelos estructurales. Se ha producido un debate considerable sobre la conveniencia de este enfoque y sobre aquellos casos en los que está justificado, tanto conceptual como empíricamente.

Métodos para especificar la confiabilidad

Para fijar la confiabilidad de un indicador en una matriz de correlación el investigador especifica el valor de ponderación como la raíz cuadrada de la fiabilidad deseada o estimada, o especifica el término de error de esa variable como 1.0 menos el valor de fiabilidad deseado. Si se utiliza la matriz de covarianza, entonces el término de error o valor de ponderación se multiplica por el valor de la varianza de la variable. Al especificar las confiabilidades, el investigador puede determinar los valores de carga, el término de error o ambos. Dado que especificar bien la ponderación o bien los términos de error determina automáticamente el otro valor, recomendamos que ambos se fijen para el modelo de mayor parsimonia y que un coeficiente no se utilice para estimar un valor que podría ser especificado. Este procedimiento puede ser realizado simplemente en programas estadísticos a través de un tratamiento único o para cada variable. Una variante de la fijación de la fiabilidad de la escala completa es un enfoque propuesto por Hayduk [54, 55] que implica un proceso por pasos. En primer lugar, el investigador selecciona el indicador aislado que se cree es la mejor representación del constructo. Para este indicador, la fiabilidad (ponderación y término de error) es fija. A continuación, pueden añadirse indicadores adicionales y sus ponderaciones y términos de error estimados, permitiendo el cálculo de la fiabilidad del constructo completo. La lógica que hay detrás de este enfoque es especificar el significado deseado del constructo a través de este indicador y a continuación permitir a otros indicadores añadir significados al concepto ya definido. El inconveniente es que cuando todas las ponderaciones y los términos de error se estiman empíricamente de forma simultánea, el investigador no está seguro exactamente de qué constructo representa excepto algún concepto subyacente común al resto de los indicadores.

Correlaciones entre constructos e indicadores

Además de los modelos de medida y estructurales, el investigador también especifica cualquier correlación entre los constructos exógenos o entre los constructos endógenos. Muchas veces los constructos exógenos están correlacionados, representando una influencia compartida sobre las variables endógenas. Las correlaciones entre los constructos endógenos, sin embargo, tienen pocas aplicaciones apropiadas y no están recomendadas para un uso normal debido a que representan correlaciones entre las ecuaciones estructurales que pueden confundir sus interpretaciones. Finalmente, los indicadores en el modelo de medida también pueden estar correlacionadas separadamente de las correlaciones de los constructos. Este método se suele evitar excepto en situaciones específicas, tales como un estudio en que existen efectos conocidos del proceso de recogida de datos o de medida sobre dos o más indicadores, o un estudio longitudinal en el que el mismo indicador se recoge en dos periodos de tiempo.

Paso 4. Selección del tipo de matriz de entrada y estimación del modelo propuesto

Como se muestra en los pasos precedentes, se exige mucho más al investigador de ciencias de la administración que utiliza SEM en términos de especificación del modelo a estimar que con otra técnica multivariante, con la posible excepción del análisis conjunto. Ahora el investigador debe llevar el proceso efectivo de estimación del modelo especificado, incluyendo los supuestos de

entrada de datos en la forma apropiada y selección del procedimiento de estimación. Las decisiones tomadas en estas áreas tienen un impacto profundo en los resultados conseguidos.

Introducción de datos

SEM difiere de otras técnicas multivariantes en que utiliza sólo la matriz de varianza-covarianza o de correlación como sus datos de entrada. Las observaciones individuales pueden ser introducidas en los programas, pero se convierten en uno de estos dos tipos de matrices antes de la estimación. El interés en SEM no está en las observaciones individuales sino el patrón de relaciones entre los encuestados. La entrada del programa es una matriz de correlación o varianza-covarianza de todos los indicadores utilizados en el modelo. El modelo de medida especifica entonces qué indicadores corresponden a cada constructo, y las puntuaciones del constructo latente son las empleadas en el modelo estructural.

Supuestos

El SEM comparte tres supuestos con los otros métodos multivariantes que hemos estudiado: **observaciones independientes, muestra aleatoria de los encuestados y la linealidad de todas las relaciones.** Además, el SEM es más sensible a las características distribucionales de los datos, particularmente el incumplimiento de la normalidad multivariante (crítica en el uso de LISREL) o una fuerte curtosis (apuntamiento) de los datos. Algunos programas informáticos, como EQS, **son menos sensibles a los datos no normales**, pero los datos deberían ser evaluados sin importar qué programa se está utilizando. Los mínimos cuadrados generalizados (*generalized least squares*), es un método de estimación alternativo, pueden ajustarse a estas violaciones, pero este método se hace rápidamente impracticable a medida que el modelo aumenta en tamaño y complejidad; por tanto su uso es limitado. **La ausencia de normalidad** multivariante es particularmente problemática porque infla sustancialmente el estadístico de la Chi-cuadrado y crea un sesgo alcista en valores críticos para determinar la significación de los coeficientes. Dado que los programas sólo aceptan las matrices de varianza-covarianza o correlación, el investigador debe realizar todos los test de diagnóstico sobre los datos antes de utilizarlos en el procedimiento de estimación. Aunque los programas de ecuaciones estructurales no tienen procedimientos de diagnóstico para contrastar estos tres supuestos, pueden ser contrastados con métodos convencionales como PRELIS. El investigador también debería identificar cualquier atípico en los datos antes de convertirlos a forma matricial.

Datos ausentes

Los datos ausentes pueden tener un profundo efecto sobre el cálculo de la matriz de entrada de datos y su capacidad para ser utilizados en el proceso de estimación. Existen 2 formas en las que se pueden incorporar los valores perdidos en el SEM. La primera es el método directo, en el que los parámetros del modelo se estiman tanto con datos completos como incompletos. Esta aproximación se utiliza raramente, sin embargo, dada la complejidad del modelo resultante. Más habitual es el método indirecto, para el cual se estima una matriz de datos de entrada utilizando toda o parte de la información disponible. Existen muchos métodos disponibles para resolver el problema de los datos ausentes, que van desde la supresión según el orden de lista a los métodos de imputación. La investigación reciente ha mostrado que el método SEM introduce menos sesgo en los modelos estimados, pero que las **opciones pareada y en forma de lista** se desarrollan adecuadamente si la proporción de datos ausentes no es muy elevada. Una conclusión del método de listas es que puede reducir seriamente el tamaño muestral. La aproximación pareada puede introducir irregularidades en la matriz de datos de entrada que provocarán después serios problemas en el proceso de estimación. Por tanto, no existe un método único que produzca siempre los mejores resultados, y si es posible que el investigador debiera emplear varias aproximaciones para evaluar la estabilidad de los resultados.

Matriz varianza-covarianza vs. matriz de correlación

Un tema importante en la interpretación de los resultados es el uso de la matriz de varianza -

covarianza frente a la matriz de correlación. El SEM se formuló inicialmente para ser utilizada con la matriz de varianzas covarianzas (de ahí su denominación habitual como análisis estructural de la covarianza). La matriz de covarianza tiene la ventaja de proporcionar comparaciones válidas entre diferentes poblaciones o muestras, característica que no es posible cuando los modelos se estiman con una matriz de correlación. La interpretación de los resultados, sin embargo, es de alguna forma más difícil cuando se utilizan covarianzas dado que los coeficientes deben interpretarse en términos de las unidades de medida de los constructos. La matriz de correlación tiene un uso muy extendido en muchas aplicaciones. Las matrices de correlación tienen un rango común que hace posible las comparaciones directas de los coeficientes dentro de un modelo, dado que simplemente es una matriz de varianzas covarianzas estandarizada en la que la escala de medida de cada variable se modifica dividiendo las varianzas o covarianzas por el producto de las desviaciones estándar. Utilizar las correlaciones es apropiado cuando el objetivo del investigador es sólo entender el patrón de las relaciones entre constructos, pero no explicar la varianza total del constructo. Otro uso apropiado es hacer comparaciones entre las diferentes variables, dado que las escalas de medida afectan a las covarianzas. Los coeficientes obtenidos de la matriz de correlación están siempre en unidades estandarizadas, similares a las ponderaciones beta de la regresión, y con un rango entre -1,0 y +1,0. Además, se ha demostrado que la matriz de correlación ofrece estimaciones más conservadoras de la significación de los coeficientes y no están sesgadas al alza, como se pensaba anteriormente. En resumen, el investigador debería emplear una matriz de varianzas covarianzas siempre que se desarrolle un contraste de la teoría real, en el que las varianzas y las covarianzas satisfacen los supuestos y la metodología y son la forma apropiada de los datos para validar las relaciones causales. Sin embargo, a menudo el investigador se centra sólo en las pautas de las relaciones, no con la explicación total tal y como se necesita en el contraste de la teoría, siendo aceptable la matriz de correlación. Si se utiliza la matriz de correlación, el investigador debería interpretar con cautela los resultados y su generalización a diferentes situaciones. *utilizadas* Uno de los modos más extendidos de cálculo de las correlaciones o covarianzas entre las variables observadas es la correlación de producto momento de Pearson. También es la forma más habitual de correlación utilizada en el análisis multivariante, simplificando al investigador el cálculo de las matrices de varianzas y covarianzas. El supuesto de la correlación del momento producto es que ambas variables se miden métricamente. Esto hace la correlación momento producto inadecuada para el uso con medidas no métricas (ordinales o binarias). Para permitir la incorporación de las medidas no métricas en modelos de ecuaciones estructurales, el investigador debe emplear diferentes tipos de correlación. Si ambas variables son ordinales con tres o más categorías (poli-cótomicas), entonces es apropiada la correlación poli-córica. Si las medidas no métricas son binarias, entonces se utiliza la correlación tetracórica. Para casos en los que la medida métrica está relacionada con una medida ordinal poli-cótica, la correlación poli-serial es la que representa la relación. Si una medida binaria está relacionada con una medida métrica, se utiliza una correlación biserial.

Tamaño muestral

Incluso aunque las observaciones individuales no sean necesarias, como sí ocurría con el resto de los métodos multivariantes, el tamaño muestral juega un papel importante en la estimación e interpretación de los resultados del SEM. El tamaño muestral, como con cualquier otro método estadístico, ofrece bases para la estimación del error muestral. La cuestión crítica en SEM es el tamaño muestral necesario. Aunque no existe ningún criterio que dicte el tamaño de muestra necesario, existen al menos 4 factores que afectan a los requisitos del tamaño muestral: (1) mala especificación del modelo, (2) tamaño del modelo, (3) no cumplimiento de la normalidad, y (4) procedimiento de estimación.

Mala especificación del modelo

La mala especificación del modelo se refiere a la medida en que el modelo sufre por los errores de especificación. El error de especificación es la omisión de variables relevantes del modelo especificado. Todos los modelos de ecuaciones estructurales sufren de errores de especificación en la medida en que no puede ser incluido cada indicador o constructo

potencial. El impacto de los constructos e indicadores omitidos, sin embargo, debería de ser insignificante si el investigador ha incluido todos aquellos relevantes según la teoría. El tamaño muestral afecta a la capacidad del modelo para ser estimado correctamente e identificar el error de especificación si se desea. Por tanto, si el investigador está preocupado por el impacto del error de especificación, las exigencias sobre el tamaño muestral deberían aumentarse por encima de aquellos que se hubiesen requerido en otro modelo.

Tamaño del modelo

El tamaño de muestra mínimo absoluto debe ser al menos tan grande como el número de covarianzas y correlaciones de la matriz de datos de entrada. Sin embargo, más habitual es un mínimo de al menos cinco encuestados para cada parámetro estimado, **con un ratio de 10 encuestados por parámetro considerado más apropiado**. Por tanto, a medida que la complejidad del modelo aumenta, también aumentan los requisitos. Nótese que estos requisitos difieren del concepto de grados de libertad y hacen referencia al número de encuestados originales utilizado para calcular la matriz de correlación o covarianza.

Incumplimientos de la normalidad

En la medida en que los datos no cumplan los supuestos de normalidad multivariante, el ratio de los encuestados respecto de los parámetros necesita aumentar a una tasa generalmente aceptada de **15 encuestados para cada parámetro**. Aunque algunos procedimientos de estimación se diseñan específicamente para tratar con datos no normales, siempre se anima al investigador a ofrecer un tamaño de muestra suficiente para que el impacto de error muestral sea mínimo, especialmente para datos no normales.

Procedimiento de estimación

Se ha demostrado que el método de estimación de máxima verosimilitud (MLE), es el procedimiento de estimación más habitual, **ofrece resultados válidos con muestras de tan sólo 50 observaciones**, pero una muestra de este tamaño *no es recomendable*. Se acepta generalmente que el tamaño de muestra mínimo que asegura el uso apropiado de **MLE es de 100 a 150**. A medida que aumentamos el tamaño muestra por encima de ese valor, el método MLE aumenta en su sensibilidad al tamaño muestra por encima de ese valor. A medida que el tamaño muestra) aumenta (**pasando de 400 a 500**), el método se hace más sensible y casi no se detecta ninguna diferencia, haciendo que todas las medidas de calidad del ajuste indiquen un ajuste muy pobre. Aunque no existe un tamaño muestra) correcto, **se recomiendan tamaños que vayan entre 100 y 200**. Un enfoque es contrastar siempre un modelo con un tamaño de muestra de **200**, sin importar cuál fuera el tamaño muestra original, debido a que algunos investigadores han propuesto **200 como un tamaño de muestra crítico**. Si se selecciona un procedimiento de estimación asintóticamente libre de distribución, los requisitos del tamaño muestra aumenta sustancialmente, ya que exige mayores tamaños de muestra para compensar la dependencia sobre los supuestos de distribución mediante otros métodos.

Existen muchos factores que afectan al tamaño muestra exigido. Normalmente, recomendamos un **tamaño muestral de 200**, aumentando si se sospecha que existe mala especificación, el modelo es muy grande o complejo, los datos exhiben características no normales o se utiliza un procedimiento de estimación alternativo. Una evaluación aislada es el **diagnóstico de la N crítica**, que es el tamaño de muestra que haría significativo el nivel de ajuste del modelo (medido por Chi-Cuadrada), llegando al nivel especificado de significación. La medida Chi-Cuadrado está disponible en la mayoría de los programas y es una base para la comparación del tamaño muestral.

Estimación del modelo

Una vez que están especificados los modelos estructurales y de medida y que se ha seleccionado el tipo de datos de entrada, el investigador debería elegir cómo se estimará el

modelo. En el caso del SEM, tiene varias opciones tanto para el procedimiento de estimación como respecto al programa informático a utilizar.

Técnicas de estimación

Los intentos iniciales de estimación del modelo de ecuaciones estructurales se realizaron con la regresión de los mínimos cuadrados ordinarios (OLS. *ordinary least squares*). Pero estos esfuerzos fueron superados rápidamente por la estimación máxima verosimilitud (ML. *maximum likelihood*), que es eficiente y no sesgada cuando se cumplen los supuestos de normalidad multivariante. Como tal y se ha convertido en una técnica ampliamente empleada en la mayoría de los programas informáticos como EQS. La sensibilidad de MLE a la no normalidad, sin embargo, creó una necesidad de técnicas de estimación alternativas y pronto aparecieron métodos tales como los mínimos cuadrados ponderados (WLS. *weighted least squares*), mínimos cuadrados generalizados (GLS) y asintóticamente libre de distribución (AGL). La técnica AGL ha recibido recientemente atención particular debido a su insensibilidad a la no normalidad de los datos. Su principal conclusión es la exigencia de un aumento del tamaño de muestra. Todas las técnicas de estimación alternativas han empezado a difundirse ampliamente a medida que ha aumentado el uso de computadoras personales, haciéndoles factibles resolver cualquier problema habitual.

Procesos de estimación

Además de la técnica de estimación empleada, el investigador puede también escoger entre varios procesos de estimación. Estos procesos van desde la estimación directa del modelo, que es similar a lo que hemos visto en el resto de las técnicas multivariantes, a métodos que generan miles de estimaciones del modelo para las cuales se obtienen los resultados finales del modelo. Existen cuatro procesos: estimación directa, *bootstrapping*, simulación y análisis *jackknife*.

Estimación directa

El proceso de estimación más habitual es el de estimación directa, en el cual se estima directamente un modelo con un procedimiento de estimación elegido. En este proceso, se estima primero, el parámetro; a continuación el intervalo de confianza (y error estándar) de cada parámetro estimado que se basa en el error muestral. Tanto los parámetros estimados como su intervalo de confianza provienen del modelo estimado de una muestra aislada.

Bootstrapping

Sin embargo, existen varias alternativas que no descansan en un único modelo de estimación sino en estimaciones del parámetro calculado y sus intervalos de confianza basados en estimaciones múltiples. La primera opción es *bootstrapping*, que se realiza en cuatro pasos: En el primero, se diseña la muestra original para que actúe como la población a efectos muestrales. En el segundo paso, se vuelve a muestrear la muestra original un número especificado de veces (quizá hasta varios miles) para generar un gran número de nuevas muestras, siendo cada una un subconjunto aleatorio de la muestra original. En el tercer paso, se estima el modelo y para cada muestra nueva se guardan los parámetros estimados. En el último paso, las estimaciones de los parámetros finales se calculan como la media de las estimaciones de los parámetros de todas las muestras. El intervalo de confianza no se estima por un error muestral, sino que se observa directamente examinando la distribución efectiva de los parámetros estimados alrededor de la media. De esta forma, las estimaciones del parámetro final y sus intervalos de confianza se obtienen directamente de las estimaciones de modelos múltiples para varias muestras aisladas y no descansan en supuestos como la distribución estadística de los parámetros.

Simulación

El investigador puede emplear técnicas de simulación, que también descansan en muestras múltiples y modelos estimados. Los procesos de simulación difieren de "*bootstrapping*" en que durante el proceso de generación de nuevas muestras, el programa de simulación puede cambiar

ciertas características de la muestra para cumplir con lo que se propone el investigador. Por ejemplo, el grado de correlación entre variables puede ser variable a lo largo de las muestras en cierta forma sistemática. De esta forma, el investigador no sólo tiene una variación asintótica de la muestra entre las muestras sino también una pauta sistemática especificada en el procedimiento de simulación. Con estas muestras, se estiman otra vez los modelos para cada muestra y los resultados compilados, como en el proceso de *bootstrapping*.

Análisis *Jackknife*. En este proceso de estimación, se crean de nuevo muestras repetidas a partir de la muestra original. Este método difiere de los procedimientos de simulación y "*bootstrapping*", sin embargo, en el método de creación de nuevas muestras. En lugar de crear un gran número de nuevas muestras aleatorias, el proceso *Jackknife* crea N muestras nuevas, donde N es el tamaño muestral original. Cada vez que se crea una nueva muestra, se omite una observación diferente. Por tanto, cada nueva muestra tiene un tamaño de muestra de N-1 con una observación diferente omitida en cada muestra. La ventaja de este proceso es la facilidad de identificar observaciones influyentes mediante el examen de los cambios en los parámetros estimados. Si se desea, la estimación del último parámetro puede calcularse como el parámetro medio, pero en los casos de tamaños de muestras reducidas no existen nuevas muestras para calcular adecuadamente el intervalo de confianza.

Programas informáticos

Una vez que se ha elegido el procedimiento de estimación, el siguiente paso es escoger el programa informático utilizado para estimar de forma efectiva el modelo. El programa más utilizado es LISREL (*Linear Structural RELations*) un modelo verdaderamente flexible para varias situaciones de investigación (sección-cruzada, experimental, cuasi-experimental y estudios longitudinales). LISREL ha encontrado aplicaciones a lo largo de todos los campos de estudio y se ha convertido casi en sinónimo de la modelización de ecuaciones estructurales. EQS tiene supuestos menos exigentes sobre la normalidad multivariante de los datos y LVPLS se ajusta mejor para la predicción aunque es limitado a efectos de interpretación de los resultados. AMOS se ha ganado una creciente popularidad en los últimos años debido a la sencillez del interfaz para el usuario y se ha comparado recientemente con LISREL y EQS [59]. Sin embargo, existen varios programas informáticos alternativos, entre ellos, PROC CALIS de SAS, COSAN. Todos estos programas se encuentran disponibles en versiones que pueden ser utilizadas por computadoras personales.

Resolución del problema de *definida no positiva*

Un problema habitual para todos los usuarios del SEM es el mensaje de error del computador: "la matriz... es definida no positiva". Lo que ha ocurrido es que tanto la matriz de datos de entrada como la matriz de datos estimados es singular, lo que significa que existe una dependencia lineal o inconsistencia entre algún grupo de variables. Existen muchas causas de este problema, pero se han encontrado algunas generales. Si el error se produce en la matriz de entrada de datos, las causas más probables son: (1) el enfoque para tratar con los datos ausentes utilizado, especialmente la eliminación pareada; o (2) una dependencia lineal entre las variables, incluyendo todos los ítems de la escala y el total de la escala en la matriz de entrada. En estos dos casos, el investigador debería generar una nueva matriz de datos, empleando un proceso de datos ausentes alternativo o eliminando las variables infractoras. Si el problema se produce en la matriz de datos estimados, entonces el investigador debe corregir cualquier varianza de error negativa (conocida como casos Heywood y que se describen en el paso 6) o intentar valores de entrada distintos.

PASO 6. Valoración de la identificación del modelo estructural

Durante el proceso de estimación, la causa más probable de la interrupción súbita del programa de la computadora o de producir resultados sin sentido o ilógicos es un problema de identificación del modelo estructural. Un problema de identificación, en términos sencillos

es la incapacidad del modelo propuesto para generar estimaciones aisladas. Se basa en el principio de que debemos tener una ecuación aislada y diferenciada para estimar cada coeficiente, basado en la premisa de que "se deben tener más ecuaciones que incógnitas" como principio básico del álgebra. Sin embargo, a medida que el modelo estructural se hace más complejo, no existe un enfoque garantizado para asegurar que el modelo está identificado (Ver más en el apartado El concepto de identificación de modelo).

Grados de libertad

A efectos de identificación, el investigador se centra en el tamaño de las matrices de correlación o covarianzas relativas respecto al número de coeficientes estimados. La diferencia entre el número de correlaciones o covarianzas y el número efectivo de coeficientes en el modelo propuesto se denomina grados de libertad. Al igual que los grados de libertad que encontramos en la regresión múltiple o MANOVA (Mejía-Trejo, 2017d), un grado de libertad es un elemento no restringido de la matriz de datos. El número de grados de libertad para un modelo propuesto se calcula como

$$gl = 1/2[(p+q)(p+q+1)] - t$$

donde:

p = número de indicadores endógenos

q = el número de indicadores exógenos

t = número de coeficientes estimados en el modelo propuesto

La primera parte de la ecuación calcula el tamaño no redundante de la matriz de covarianza o correlación (es decir, la mitad inferior o superior de la matriz más la diagonal). Entonces cada coeficiente estimado gasta un grado de libertad. La principal diferencia entre los grados de libertad utilizados en el SEM comparado con otras técnicas multivariantes es que el número de parámetros estimados se compara con el número de elementos de la matriz de datos, NO con el tamaño muestral! En el SEM el tamaño muestral se utiliza para estimar el error de muestreo, pero no afecta a los grados de libertad.

Normas para la identificación

Aunque no existe una regla aislada que establezca la identificación de un modelo, el investigador dispone de varias «normas» o heurística. Las dos normas más básicas son las condiciones de orden y rango. Las condiciones de orden afirman que los grados de libertad del modelo deben ser mayores o iguales a cero. Esto corresponde a lo que hemos denominado como modelo identificado o modelo sobreidentificado. Un modelo identificado tiene exactamente cero grados de libertad. Aunque esto ofrece un ajuste perfecto del modelo, la solución no tiene interés puesto que no se puede generalizar. Un modelo sobreidentificado es el objetivo de todos los modelos de ecuaciones estructurales. Tiene más información en la matriz de datos que el número de parámetros a estimar, lo que significa que tiene un número positivo de grados de libertad. Al igual que en otras técnicas multivariantes, el investigador se esfuerza por conseguir un ajuste aceptable con el mayor grado de libertad posible. Esto asegura que el modelo es tan generalizable como sea posible. Un modelo que no llega a cumplir la condición de orden se conoce como un modelo infraestimado. Este modelo tiene grados de libertad negativos, lo que significa que se intentan estimar más parámetros de lo que permite la información disponible. El modelo no puede ser estimado hasta que algunos parámetros sean fijos o restringidos. La condición de orden es una condición de identificación necesaria, pero no suficiente. El modelo también debe cumplir la condición de rango, lo que exige que el investigador determine algebraicamente si cada parámetro se identifica (estima) especialmente. Pero incluso, para los modelos más sencillos, se trata de un ejercicio muy complejo para ser considerado directamente por el investigador. En su lugar, hay heurística disponible. Primero está la norma de las 3 medidas, que evalúa que cualquier constructo con tres o más indicadores siempre estará identificado. También existe la norma del modelo recursivo, que dice que los modelos recursivos con constructos identificados

(norma de las tres medidas) siempre estarán identificados . Un modelo recursivo NO tiene relaciones no recursivas o recíprocas en el modelo estructural .

Diagnóstico de los problemas de identificación

Los programas de ecuaciones estructurales también realizan contrastes para diagnosticar problemas de identificación. LISREL tiene un test sencillo de identificación durante el proceso de estimación para examinar la matriz de información, mientras que EQS tiene el **test de rango de Wald (1943)**. Aunque estos contrastes identifican la mayoría de los problemas de identificación, pueden no evaluar la unicidad de cada parámetro estimado, como exige la condición de rango.

El investigador puede llevar a cabo contrastes cuando la ecuación se identifica para ver si los resultados son inestables debido al nivel de identificación. En primer lugar, el modelo puede ser reestimado varias veces, cada una de ellas con un valor de partida diferente. El investigador puede especificar un valor inicial para cualquier parámetro estimado, un punto de partida del proceso de estimación. Si el valor de partida no está disponible, el programa lo calcula automáticamente por diversos métodos. Si los resultados no convergen en el mismo punto para diversos valores de partida, la identificación debería examinarse más a fondo. El segundo test evalúa el efecto de la identificación sobre un único coeficiente, que consiste en estimar en primer lugar el modelo y a continuación obtener la estimación de coeficiente. Posteriormente, se fija el coeficiente a su valor estimado y se reestima la ecuación. Si el ajuste conjunto del modelo varía significativamente, esto indica problemas de identificación. Otro enfoque es observar los posibles síntomas de un problema de identificación. Estos incluyen: (1) errores estándar muy elevados para uno o más coeficientes, (2) la incapacidad del programa para invertir la matriz de información, (3) estimaciones muy poco razonables o estimaciones imposibles tales como varianzas de error negativas, o (4) elevadas correlaciones ($\pm 0,90$ o superiores) entre los coeficientes estimados.

Causas y soluciones para los problemas de identificación

Si se localiza un problema de identificación, el investigador debería buscar tres posibles causas: (1) un mayor número de coeficientes estimados relativos al número de correlaciones o covarianzas, indicada por el reducido número de grados de libertad, similar al problema de sobreajuste de los datos encontrado en otras técnicas multivariantes; (2) el uso de efectos recíprocos (flechas causales de dos sentidos entre dos constructos); o (3) fallo en la fijación de la escala de un constructo. Discutiremos este procedimiento más adelante en nuestro análisis de los datos del ejemplo.

La única solución para un problema de identificación es definir más restricciones para el modelo, esto es, eliminar algunos de los coeficientes estimados. El investigador debería seguir un proceso estructurado, añadiendo gradualmente más restricciones (eliminando secuencias del diagrama de trayectorias) hasta que el problema esté solucionado. Al hacerlo, el investigador está intentando conseguir un modelo sobreidentificado que tiene grados de libertad suficientes con las cuales evaluar, si es posible, la cantidad de error de medida y de muestreo y ofrecer mejores estimaciones de las verdaderas relaciones casuales. Para llegar a este fin, se recordará el siguiente proceso: (1) construir un modelo teórico con el mínimo número de coeficientes (incógnitas) que puedan ser justificados. Si se encuentran problemas de identificación, procederemos con las soluciones por este orden: (2) fijar las varianzas de error de medida de los constructos si es posible, (3) fijar algunos coeficientes estructurales que sean conocidos con confiabilidad, y (4) eliminar las variables problemáticas. Si los problemas de identificación todavía existen, el investigador debe reformular el modelo teórico para ofrecer más constructos relativos al número de relaciones causales examinadas.

Paso 6. Evaluación de los criterios de calidad de ajuste

El primer paso de la evaluación de los resultados es una inspección inicial de las "estimaciones infractoras". Una vez que el modelo está establecido como para ofrecer estimaciones aceptables, a continuación debe evaluarse la calidad del ajuste a diversos niveles: en primer lugar para el modelo conjunto y a continuación para los modelos estructurales y de medida por separado.

Estimaciones infractoras

En primer lugar se examinan los resultados buscando estimaciones infractoras. Se trata de coeficientes estimados tanto en los modelos de medida como los estructurales que exceden los límites aceptables. Los ejemplos más normales de estimaciones infractoras son: (1) varianzas de error negativas o varianzas de error no significativas para cualquier constructo, (2) coeficientes estandarizados que sobrepasan o están muy cerca de 1.0 o (3) errores estándar muy elevados asociados con cualquier coeficiente estimado. Si se encuentran estimaciones infractoras, el investigador debería resolver en primer lugar cada caso antes de evaluar cualquier resultado específico del modelo, en la medida en que cambios en una parte del modelo puedan tener efectos significativos sobre otros resultados. Se han utilizado varios enfoques para la resolución de estos problemas en la discusión de los problemas de identificación. Si se corrigen los problemas de identificación y aún así, los problemas todavía se mantienen, existen otras soluciones. En el caso de varianzas de error negativas (también conocidas como casos de Heywood), una posibilidad es fijar las varianzas de los errores infractores a un valor positivo muy pequeño (.005). Aunque este remedio cumple los requisitos prácticos del proceso de estimación, sólo enmascara el problema subyacente y debe ser considerado cuando se interpreten los resultados. Si las correlaciones de la solución estandarizada exceden de 1.0, o dos estimaciones están altamente correlacionadas, entonces el investigador deberá considerar la eliminación de uno de los constructos o deber asegurarse que se ha establecido entre los constructos una verdadera validez discriminante. En muchos casos, tales situaciones son el resultado de **modelos ateóricos**, establecidos sin la suficiente justificación teórica o modificados solamente a partir de consideraciones empíricas.

Ajuste global del modelo

Una vez que el investigador ha establecido que no existen estimaciones infractoras, el siguiente paso es evaluar el ajuste global del modelo con una o más medidas de calidad del ajuste. La calidad del ajuste mide la correspondencia entre la matriz de entrada real u observada (covarianza o correlación) con la que se predice mediante el modelo propuesto.

Al desarrollar cualquier modelo estadístico, el investigador debe ser precavido con el sobreajuste del modelo a los datos. Al discutir la regresión, mostramos que deberían mantenerse ciertas relaciones (**quizá 5 a 1**, ver Mejía-Trejo, 2017c) entre el número de coeficientes estimados y el número de encuestados. Esta relación debería mantenerse también en el SEM. El investigador debe esforzarse en tener un gran número de grados de libertad, si el resto no cambia. Al hacerlo así, el modelo consigue parsimonia, el objetivo de un mejor ajuste del modelo para cada coeficiente estimado. Cuanto mejor ajuste pueda conseguirse con pocos coeficientes, mejor será el contraste del modelo y más confianza podemos tener en que los resultados no sean producto del sobreajuste de los datos.

Las medidas de calidad del ajuste son de tres tipos: (1) medidas absolutas del ajuste, (2) medidas del ajuste incremental, o (3) medidas de ajuste de parsimonia. Las medidas absolutas del ajuste evalúan sólo el ajuste global del modelo (tanto los modelos de medida como los estructurales colectivamente), sin ajuste para el grado de sobreajuste que pudiera ocurrir. Las medidas del ajuste incremental comparan el modelo propuesto con otro modelo especificado por el investigador. Finalmente, las medidas de ajuste de parsimonia ajustan las medidas de ajuste para ofrecer una comparación entre modelos con diferentes números de coeficientes estimados, siendo el propósito determinar la cantidad del ajuste conseguido por cada coeficiente estimado. El investigador se enfrenta con la cuestión de qué medidas elegir. No surge ninguna medida aislada o conjunto de medidas como las únicas medidas necesarias. Como el SEM ha evolucionado en los últimos años, las medidas de calidad del ajuste se han estado desarrollando continuamente y se han propuesto medidas adicionales. Se anima al investigador a buscar y emplear una o más medidas de cada tipo. La evaluación de la calidad del ajuste de un modelo es más un proceso relativo que un criterio absoluto. La aplicación de diversas medidas del ajuste múltiple permitirá al investigador ganar en consenso tanto para los tipos de medidas como para la aceptación del modelo propuesto. Un nivel aceptable de calidad del ajuste conjunto NO garantiza que todos los constructos cumplan los requisitos del ajuste del modelo, ni que el modelo estructural puede ser mantenido con certeza absoluta. El investigador deberá evaluar cada una de estas áreas

separadamente para confirmar su cumplimiento de los requisitos o como medio de identificar problemas potenciales que afectan a la calidad del ajuste conjunto del modelo.

Ajuste del modelo de medida

Una vez que se ha evaluado el ajuste del modelo conjunto, podemos evaluar la unidimensionalidad y la confiabilidad de la medida de cada constructo. La unidimensionalidad es un supuesto que subyace al cálculo de la confiabilidad y se demuestra cuando los indicadores de un constructo tienen un ajuste aceptable sobre un modelo de un único factor (unidimensional). El uso de medidas de confiabilidad, tales como la alfa de Cronbach NO asegura la unidimensionalidad sino que en su lugar supone que existe. Se anima al investigador a que re-alice contrastes de unidimensionalidad sobre todos los constructos de indicadores múltiples antes de evaluar su confiabilidad. El siguiente paso es examinar las ponderaciones estimadas y evaluar la significación estadística de cada una. Si la significación estadística no se consigue, el investigador puede desear eliminar el indicador o intentar transformarlo para un mejor ajuste del constructo.

Confiabilidad compuesta

Antes de realizar el examen de las ponderaciones de cada indicador, se utiliza una medida fundamental para la evaluación del modelo de medida, la confiabilidad compuesta de cada constructo. La confiabilidad es una medida de la consistencia interna de los indicadores del constructo, que representa el grado en que éstos “*indican*” el constructo común latente (no observado). Medidas más confiables ofrecen al investigador una mayor confianza de que todos los indicadores individuales son consistentes en sus medidas. Un valor umbral comúnmente aceptado para aceptar la hipótesis de fiabilidad es *.7*, aunque no se trate de un estándar absoluto, y los valores por debajo de *.7* se han demostrado aceptables si la investigación tiene naturaleza exploratoria. Debemos tener en cuenta, sin embargo, que **la confiabilidad no asegura validez**. La validez es la medida en que los indicadores miden “*con precisión*” lo que se supone que están midiendo. Por ejemplo, pueden ser confiables diversas medidas de cómo y por qué los consumidores compran productos, pero el investigador puede asumir erróneamente que miden la lealtad a la marca cuando de hecho son indicadores de intenciones de compra. En este caso, los indicadores son un conjunto confiable de medidas pero una medida inválida de lealtad a la marca. El supuesto de validez descansa en la especificación del investigador de indicadores para un constructo latente.

La confiabilidad y la varianza extraída (véase la sección siguiente) de un constructo latente deben calcularse por separado para cada constructo con indicadores múltiples en el modelo. Aunque LISREL y EQS no los calculan directamente, se proporciona toda la información necesaria. (Ver más en Capítulo 3).

Varianza extraída

Otra medida de confiabilidad es la medida de varianza extraída. Esta medida refleja la cantidad total de la varianza de los indicadores tenida en cuenta por el constructo latente. Los mayores valores de la varianza extraída se producen cuando los indicadores son verdaderamente representativos del constructo latente. La varianza extraída es una medida complementaria del valor de la confiabilidad del constructo. Esta medida es muy parecida a la medida de la fiabilidad pero difiere en que las ponderaciones estandarizadas se elevan al cuadrado antes de sumarlas. En general, se sugiere que el valor de la varianza extraída debería exceder de *.5* para un constructo. Los ejemplos reales de los cálculos tanto para la medida de la fiabilidad como la varianza extraída, se ofrecen en el ejemplo del análisis factorial confirmatorio de este mismo capítulo (ver más en capítulo 3 y 4).

Ajuste del modelo estructural

El examen más obvio del modelo estructural consiste en la significación de los coeficientes estimados. Los métodos de modelización de ecuaciones estructurales ofrecen no sólo coeficientes estimados sino también errores estándar y valores *t* calculados para cada coeficiente. Si

podemos especificar el nivel de significación que juzgamos apropiado (es decir .05), entonces cada coeficiente estimado puede ser contrastado por la significación estadística (es decir, que sea diferente de cero) para las relaciones causales supuestas. Sin embargo, dadas las propiedades estadísticas del MLE y sus características (tamaños de muestra reducidos), se anima al investigador a ser conservador en la especificación del nivel de significación, eligiendo niveles pequeños (.025 o .01) en lugar del tradicional nivel de .05. La selección de un valor crítico también depende de la justificación teórica de las relaciones propuestas. Si se supone una relación positiva o negativa, se puede emplear un test de significación con una sola cola. Sin embargo, si el investigador no puede preespecificar la dirección de la relación, entonces se necesita utilizar un test de dos colas. La diferencia está en los valores críticos de la t utilizados para evaluar la significación. Por ejemplo, para el nivel de significación de .05, el valor crítico para un test de una sola cola es de 1.645, pero aumenta a 1.96 para un test de dos colas. Por tanto, el investigador puede detectar con mayor precisión las diferencias causales. Aunque no se ha fijado ningún límite que defina lo que se consideran como correlaciones elevadas, los valores que exceden .9 siempre deberían ser examinados, y muchas veces las correlaciones que exceden .8 pueden ser indicadoras de problemas.

Comparación de modelos anidados o rivales

Las estrategias de modelización más habituales son: estrategia de desarrollo de modelo y modelos rivales consisten en la comparación de los resultados del modelo para determinar el modelo mejor ajustado de un conjunto de modelos. En una estrategia de desarrollo del modelo, el investigador comienza con un modelo inicial y sigue con una serie de re-especificaciones del modelo, con las que cada vez se espera mejorar el ajuste del modelo mientras se mantenga la concordancia con la teoría subyacente. Para ayudar a la comparación entre modelos, se han desarrollado una gran cantidad de medidas para evaluar el ajuste del modelo. Una clase de medidas evalúa el ajuste del modelo conjunto en términos absolutos, ofreciendo una medida específica del ajuste. Una conclusión sobre estas medidas es que no tienen en cuenta el número de relaciones utilizadas en la obtención del ajuste del modelo. Para medir la parsimonia del modelo, se han propuesto una serie de medidas de ajuste de parsimonia. Su objetivo es determinado por el ajuste por coeficiente, dado que el ajuste absoluto siempre mejorará a medida que se añaden coeficientes. Anderson y Gerbing (1988) propusieron un procedimiento exhaustivo para este propósito, en el que se especifican una serie de modelos rivales. Puede mostrarse que las diferencias entre modelos son simplemente la diferencia entre los valores de la Chi-Cuadrado para los diferentes modelos. A continuación se puede contrastar la significación de la diferencia de la chi-cuadrado con los grados de libertad apropiados, esto es, la diferencia en el número de coeficientes estimados para los dos modelos. El único requisito es que el número de constructos o indicadores sea el mismo, de tal forma que el modelo nulo sea el mismo para ambos modelos (es decir, son modelos anidados). El efecto de añadir o destruir una o más relaciones causales se puede contrastar también de esta forma haciendo comparaciones entre modelos con y sin las relaciones. Si los modelos no se hacen anidados (tener un número diferente de indicadores o constructos), el investigador debe basarse en las medidas de ajuste de parsimonia descritas previamente, en la medida en que el test de la chi-cuadrado no es apropiado para este caso.

Paso 7. Interpretación y modificación del modelo

Una vez que el modelo se considera aceptable, el investigador debe examinar en primer lugar los resultados y su correspondencia con la teoría propuesta. ¿Están corroboradas y son estadísticamente significativas las principales relaciones de la teoría? ¿Añaden los modelos rivales mayor perspectiva sobre las formulaciones alternativas de la teoría como para que puedan ser tenidas en cuenta? ¿Están todas las relaciones en la dirección supuesta (positiva o negativa)? Todas estas cuestiones y muchas más pueden ser contestadas a partir de los resultados empíricos. En el proceso de respuesta a estas tres cuestiones, el investigador puede tener necesidad de considerar dos supuestos de interpretación: el uso de las soluciones estandarizadas frente a las no estandarizadas y la re-especificación del modelo.

Soluciones estandarizadas frente a las no estandarizadas

Un aspecto de la evaluación de una relación estimada es la valoración del tamaño del parámetro. Al igual que en otras técnicas multivariantes, tales como la regresión múltiple, existe una diferencia notable en las soluciones estandarizadas y no estandarizadas en términos de su interpretación y uso. En los modelos de ecuaciones estructurales, los coeficientes estandarizados tienen todos igual varianza y un valor máximo de 1.0, por tanto, efectos tamaño muy aproximados, como se mostró con las ponderaciones beta en la regresión. Los coeficientes cercanos a cero tienen poco efecto sustantivo, si es que tienen alguno, mientras que un aumento en el valor corresponde a un aumento de la importancia en las relaciones causales. Los coeficientes estandarizados son útiles en la determinación de la importancia relativa, pero son específicos de una muestra y no son comparables entre las muestras. Los coeficientes sin estandarizar corresponden a las ponderaciones de la regresión en una regresión múltiple en la que se expresan en términos de la escala del constructo, en este caso su varianza. Esto hace comparables a estos coeficientes para las muestras y retiene sus efectos escala. Dado que la escala varía para cada constructo, sin embargo, la comparación entre coeficientes es más difícil que con los coeficientes estandarizados.

Re-especificación del modelo

Una vez que la interpretación del modelo se ha completado, el investigador normalmente busca métodos para mejorar el ajuste del modelo y/o su correspondencia con la teoría subyacente. En tales casos, puede iniciar la re-especificación del modelo, el proceso de añadir o eliminar los parámetros estimados del modelo original. Antes de tratar algunos enfoques para identificar la modificación del modelo, aconsejamos al investigador hacer tales modificaciones con cuidado y sólo después de obtener justificación teórica para lo que se considera empíricamente deseable. Las modificaciones del modelo original deberían hacerse sólo después de una consideración deliberada. Si se hacen las modificaciones, el modelo debería tener validación cruzada (es decir, estimado sobre un conjunto distinto de datos) antes de que el modelo modificado sea aceptado.

Un proceso de re-especificación del modelo

Antes de identificar cualquier posible re-especificación del modelo, el investigador deberá clasificar todas las relaciones (estimadas o no) en una de 2 categorías: teóricas o empíricas. Las relaciones teóricas son esenciales a la teoría subyacente y no pueden ser modificadas. Están fuera de los límites de la reespecificación. La categoría empírica contiene relaciones que se añaden para mejorar el ajuste del modelo. Estas pueden ser re-especificadas. El objetivo es determinar un conjunto de modelos teóricos anidados, donde el conjunto de modelos se contempla como un conjunto de diferentes niveles de parsimonia para la misma teoría subyacente. De esta forma, el modelo se convierte en una serie de modelos rivales que llevan a varios niveles de corroboración para la teoría.

Indicadores empíricos de posibles reespecificaciones

¿Dónde puede buscar el investigador la mejora de los modelos? La primera indicación viene del examen de los residuos de la matriz de las predicciones de la covarianza y correlación. Los residuos estandarizados (también denominados residuos normalizados) representan las diferencias entre la matriz de covarianza o co-relación observada y la matriz de covarianza o correlación estimada. Con las últimas versiones de LISREL (versión 7 y superiores), se ha mejorado el cálculo de los residuos

y el patrón de evaluación de los residuos «significativos» ha cambiado (el umbral anterior estaba en $\pm 2,0$). Los valores residuales mayores que $\pm 2,58$ se consideran ahora estadísticamente significativos al nivel de $0,05$. Los residuos significativos indican un error de predicción sustancial para un par de indicadores (es decir, una de las covarianzas o correlaciones de los datos de entrada originales).

Un residuo estandarizado indica sólo que existe diferencia pero no arroja luz acerca de cómo puede ser reducida. El investigador debe identificar el remedio mediante la suma o modificación de las relaciones causales.

Otro apoyo en la evaluación del ajuste de un modelo especificado consiste en los índices de modificación, que se calculan para cada relación no estimada. El valor del índice de modificación corresponde aproximadamente a la reducción en la Chi-Cuadrado que se produciría si el coeficiente fuera estimado. Un valor de $3,84$ o superior sugiere que se obtiene una reducción estadísticamente significativa en la chi-cuadrado cuando se estima el coeficiente. Aunque los índices de modificación pueden ser útiles en la evaluación del impacto de modificaciones basadas teóricamente, el investigador nunca hace cambios de modelo basándose sólo en índices de modificación. Este enfoque ateoórico es completamente contrario al “*espíritu*” de la técnica y debería ser evitado en todos los casos. La modificación del modelo debería tener una justificación teórica antes de ser considerada e incluso entonces el investigador debería ser escéptico acerca de los cambios. Además del índice de modificación, LISREL y EQS también ofrecen un parámetro de cambio esperado, que hace referencia a la magnitud y la dirección de cada parámetro fijo (no estimado). Este parámetro difiere del índice de modificación en que no indica el cambio en el ajuste del modelo conjunto (χ^2) y en su lugar representa el cambio en el valor del parámetro efectivo. EQS ofrece el **multiplicador de Lagrange y el estadístico de Wald**, que evalúa el efecto de liberar un conjunto de parámetros simultáneamente. Una vez hechas las modificaciones en el modelo, el investigador debe volver al paso 4 del proceso de 7 pasos y reevaluar los modelos modificados. Si se anticipan modificaciones amplias del modelo, los datos deberían dividirse en 2 muestras, una de ellas ofreciendo la base para la estimación y modificación del modelo, y la otra validación para el modelo final.

Una recapitulación del proceso de siete pasos

La modelización de ecuaciones estructurales ofrece al investigador más flexibilidad que cualquier otro método multivariante discutido. Pero junto con esta flexibilidad viene el potencial uso inapropiado del modelo. Un asunto primordial en cualquier aplicación del SEM, es que deber tener una fundamentación teórica incondicional, apoyada por un test confirmatorio robusto para una serie de relaciones casuales. Sin embargo, cuando se aplica el método de forma exploratoria, el investigador se enfrenta con la muy elevada probabilidad de no conseguir resultados mediante la búsqueda o de datos y las relaciones identificadas tendrán poca generalización mediante la simple extrapolación de las relaciones específicas de los datos de la muestra que se están estudiando.

Fuente: Jöreskog (1993); Hair (et al., 1999); Byrne (2006), con adaptación

datos se transponen a los parámetros del modelo. Esta tarea se realiza a través de un modelo matemático que representa todo el sistema de variables. Dichos sistemas de representación pueden y deben variar con cada programa de computadora SEM; el mecanismo utilizado por el programa EQS se discute más adelante. Al igual que con cualquier forma de comunicación, uno debe primero entender el lenguaje antes de poder entender el mensaje transmitido; así es en la comprensión de la especificación de los modelos SEM.

EQS. Notación

El programa EQS considera que todas las variables caen en una de dos categorías: variables medidas (observadas) o variables no medidas (no observadas). Todas las variables medidas se designan como V y constituyen los datos “reales” de un estudio. Todas las demás variables son hipotéticas y representan la red estructural del fenómeno bajo investigación. Aunque conceptualmente es innecesario, tiene sentido en la práctica diferenciar las variables no medidas, debido a:

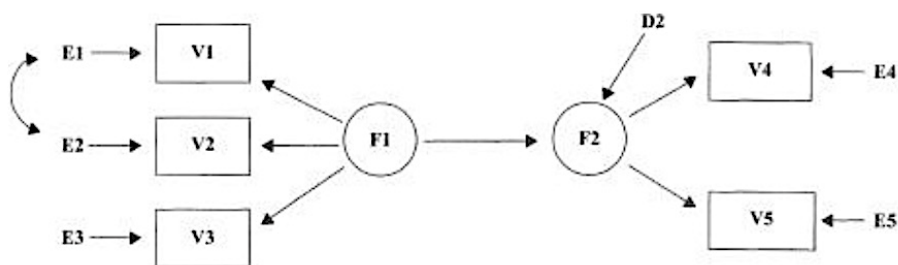
- ▶ el propio constructo latente (considerado generalmente como un factor en EQS),
- ▶ designada como F;
- ▶ un residual asociado con la medición de cada variable observada (V),
- ▶ designada como E; y
- ▶ un residuo asociado con la predicción de cada factor, designado como D.

Los términos residuales a menudo se conocen como disturbios (*disturbance*), que es la terminología utilizada en el programa EQS. El término disturbio se usa en lugar de residual a lo largo del resto de este libro. Finalmente, para simplificar todas las letras E y D están numeradas para corresponderse con las V y F con las que están asociados, respectivamente.

EQS. Diagrama de trayectoria

Para comprender completamente cómo funciona el sistema de símbolos y etiquetado, veamos la figura 3.

FIGURA 3. EQS. Modelo general de ecuación estructural .



Fuente: Jöreskog (1993); Hair et al., (1999); Byrne (2006), con adaptación.

Como se observa, es muy similar a lo planteado desde la figura 1, agregando sólo los comentarios que llevan en el número que acompaña a cada variable

etiquetada en el modelo, ya que:

- ▶ EQS automáticamente numera cada variable observada de acuerdo con su ubicación de entrada de datos. Como tal, la primera variable en el conjunto de datos se designaría V1, el segundo V2, y así sucesivamente.
- ▶ El valor numérico asociado con cada error (E) y el término de perturbación (D) son consistentes con su relación observada (V) y no observada (F) variables, respectivamente. Por lo tanto, aunque sólo hay un término de perturbación en este modelo, se denomina D2 en lugar de D1.

EQS. Representación del sistema Bentler-Weeks

Como se discutió anteriormente, dado que los datos de muestra constituyen puntajes observados solamente, cada programa SEM requiere algunos medios para transponer estos puntajes en parámetros del modelo. Esto se logra a través de un modelo matemático que represente a todo el sistema de variables. En EQS, el modelo matemático deriva del trabajo de Bentler y Weeks (1979, 1980). Lo valioso del modelo de Bentler-Weeks (*idem*), es que todas las variables del modelo, pueden clasificarse ya sea como variables dependientes o independientes. Cualquier variable que tenga una flecha unidireccional dirigida a ella representa una variable dependiente; si no hay una flecha unidireccional dirigida a él, una variable se considera independiente. Como es habitual, las variables dependientes se explican en términos de otras variables en el modelo, mientras que las variables independientes sirven como variables explicativas. No es tan habitual, sin embargo, la conceptualización de Bentler-Weeks (*idem*) sobre lo que es una variable dependiente o independiente. De hecho, la interpretación de este concepto es mucho más amplia de lo común. Según Bentler y Weeks (*idem*), cualquier variable que no sea una variable dependiente se considera automáticamente una variable independiente, sin importar de si se trata de una puntuación observada, un factor no observado o un término de perturbación. Por ejemplo, en la figura 3, las variables dependientes, se clasifican en: V1, V2, V3, V4, V5 y F2; las variables independientes, son: E1, E2, E3, E4, E5, F1 y D2. Una variable dependiente, entonces, es cualquier variable que se puede expresar como una función de regresión estructural de otras variables. Por lo tanto, para cada variable dependiente, esta función de regresión puede resumirse en la forma de una ecuación. Al igual que con las ecuaciones basadas en las variables que se muestran en la figura 1, cada función de regresión modelado en la figura 3 se puede traducir a ecuaciones específicas de EQS. Como se hará evidente en las aplicaciones ilustradas más adelante, estas ecuaciones sirven esencialmente para definir el modelo del programa. Las

ecuaciones relacionadas son las siguientes:

$$F2 = F1 + D2$$

$$V1 = F1 + E1$$

$$V2 = F1 + E2$$

$$V3 = F1 + E3$$

$$V4 = F2 + E4$$

$$V5 = F2 + E5$$

Está claro que las flechas unidireccionales vinculan a los factores con las variables observadas y el factor 1 con el factor 2 representan coeficientes de regresión, sin embargo la explicación con respecto a la vinculación en términos de la perturbación o disturbios asociados a sus variables a través de las flechas de acceso unidireccional, pueden ser algo menos obvias. Aunque estas flechas también simbolizan los coeficientes de regresión, sus trayectorias están implícitas en la predicción de una variable a otra; así se consideran conocidos y, por lo tanto, se fijan a 1.0. Por ejemplo, en el lenguaje simple regresión, la predicción de V1 a partir de F1 puede escribirse como:

$$V1 = b_n F1 + E1$$

donde b_n representa el peso beta desconocido asociado con el predictor F1 y E1 significa un error en esta predicción. Tome en cuenta que no hay un peso beta asociado con el término de error, lo que indica que no se debe estimar. Por implicación entonces, el peso beta para E1 se considera conocido y se fija arbitrariamente a 1 (la varianza de E1 necesita ser fijado a 1 porque tanto el coeficiente de regresión como la varianza no pueden ser estimados simultáneamente).

De manera similar, la predicción de F2 a partir de F1 se puede escribir como

$$F2 = b_{12} F1 + D2,$$

donde D_2 representa un error en la predicción, aunque esto implica la predicción de un factor de otro (mientras que en la predicción anterior, la ecuación implicaba la predicción de una variable observada a partir de un factor (por lo tanto, la distinción entre la términos E y D). Finalmente, un corolario importante del modelo de Bentler-Weeks es que las varianzas de las variables dependientes o sus covarianzas con otras variables nunca son parámetros del modelo; más bien, sigue siendo explicado por esos parámetros. Por el contrario, las varianzas y covarianzas de las variables independientes son parámetros importantes que necesitan ser estimados.

EQS. Uso del programa

Ahora que se comprende la notación básica de EQS, es hora de ver:

- ▶ Cómo pueden vincularse estos símbolos juntos para formar declaraciones y párrafos del programa admisibles utilizados en la construcción de un archivo de entrada SEM, y
- ▶ Cómo ejecutar estos archivos de entrada para que el modelo especificado pueda ser probado.

El objetivo principal aquí es presentar una visión multifacética y exhaustiva de las capacidades del programa; en consecuencia, este capítulo es necesariamente extenso. Sin embargo, después de trabajar, a través de este material, se espera que tenga una mejor apreciación para el programa banco de construcción de modelos y los recursos de prueba de modelo que le esperan al usar EQS. Comenzamos esta visita guiada de la programación de EQS inspeccionando primero todos los requisitos y la mayoría de los opcionales componentes del archivo de entrada. Para una visión más completa de cómo se combinan estos componentes, se revisará su aplicación con respecto a tres modelos diferentes (es decir, el CFA de primer orden, segundo orden CFA y SEM completo). Se presenta el importante concepto de identificación de modelo (o estadístico). Posteriormente, se aborda la cuestión de cómo crear un archivo de entrada EQS. Aquí, se ofrece tres opciones:

- ▶ Manual.
- ▶ Interactivo, usando la opción de comando `BUILD_EQS`.
- ▶ Gráfico, usando la opción `DIAGRAMMER`.

Continuando con la siguiente etapa, preparamos la salida y análisis de datos. Finalmente, se presenta un resumen de posibles mensajes de error que pueden ocurrir en el uso de EQS.

EQS. Componentes del archivo de entrada

Un archivo de entrada EQS describe el modelo bajo estudio. Se compone de varias declaraciones agrupadas dentro de párrafos según ciertas reglas. En esta subsección, trataremos sobre instrucciones y datos básicos que rigen la construcción de archivos de entrada EQS, luego se examinarán los componentes básicos de un archivo de entrada EQS y, finalmente, se revisará el método manual para formular y ejecutar archivos de entrada EQS relacionados con tres modelos SEM diferentes.

EQS. Reglas básicas en la creación de archivos de entrada

Existen una serie de instrucciones básicas, las cuales las resumimos a continuación:

Keywords (palabras clave)

Como en la mayoría de los programas de computadora, EQS usa un sistema de palabras clave que el programa interpreta como comandos básicos. Las palabras clave principales en EQS son los nombres utilizados para identificar párrafos particulares que comprende el archivo de entrada (por ejemplo, el párrafo /SPECIFICATIONS). Las palabras clave secundarias representan cada declaración dentro de un párrafo; funcionan como subcomandos. Como tales, se refieren a un aspecto particular del párrafo, lo que permite al usuario elegir una de varias opciones relacionadas con este componente. Por ejemplo, un aspecto del párrafo /SPECIFICATIONS involucra el método de estimación. De las opciones disponibles, el usuario puede desear emplear una estimación generalizada de mínimos cuadrados (generalized least squares estimation), que se establecería como ME = GLS, donde ME es la forma abreviada de la palabra clave METHOD, un subcomando que indica el método de estimación que se utilizará, y GLS es la palabra clave de tres letras que indica que la estimación generalizada de mínimos cuadrados es el método de estimación que se utilizará.

Dos reglas principales rigen el uso de los párrafos en las palabras clave:

- a. Siempre deben ir precedidas de una barra oblicua (por ejemplo, /SPECIFICATIONS), y
- b. La línea en la que aparece la palabra clave no debe contener ninguna otra entrada información. Un uso opcional del párrafo en las palabras clave es que pueden abreviarse a las tres primeras letras (por ejemplo, SPE). Sin embargo, el usuario puede preferir más de tres letras porque da como resultado una palabra clave significativa, y esta elección es bastante aceptable. Por ejemplo, abreviatura /SPECIFICATIONS a /SPEC parece ser una mejor opción.

Descriptive statements

Toda la información que describe el modelo y los datos debe expresarse con declaraciones específicas escritas dentro del párrafo apropiado. Un punto y coma (;) debe separar cada instrucción. Aunque no es necesario, a menudo es útil para construir archivos de entrada, de tal manera que los haga fáciles de leer. Una forma

de hacerlo es comenzar párrafos de palabras clave en la columna 1, con todas las demás declaraciones dentro del párrafo con sangría de algunos caracteres. Finalmente, todas las declaraciones de entrada deben especificarse en no más de 80 columnas por línea. En ocasiones, es posible que desee insertar un comentario recordatorio en el archivo de entrada con respecto a algún aspecto de los datos o análisis. EQS le permite incluir tales comentarios en cualquier línea del archivo de entrada por medio de un signo de exclamación (!), que debe preceder al comentario. El programa ignora todo el material a la derecha del símbolo de exclamación.

File editors

Para que se ejecute EQS, el archivo de entrada debe ser un archivo simple que no contenga caracteres de control ocultos. Dado que la mayoría de los procesadores de texto (por ejemplo, *Word*™, *Wordperfect*™) usan dichos símbolos en la formulación del texto, los caracteres invisibles primero deben eliminarse antes de que EQS pueda leerlos. Sin embargo, esta restricción es fácil direccionado al guardar el archivo de entrada en formato ASCII, que EQS puede leer.

Data input

Los datos en forma de una matriz de covarianza o correlación se pueden incrustar dentro del archivo de entrada o residir en un archivo externo. Los datos de insumo siempre deben residir en un archivo separado. Detalles específicos sobre la integración de los archivos de datos con el archivo de entrada EQS se abordan en la siguiente subsección.

Basic components of the EQS input file

Como se discutió anteriormente, la estructura básica de un archivo de entrada EQS comprende una serie de declaraciones agrupadas dentro de párrafos, cada uno de los cuales se introduce por medio de una palabra clave precedida por una barra oblicua (/). Sin embargo, no todos los párrafos son necesarios. Así, serán revisadas las características más importantes relacionadas con seis párrafos básicos aunque existen más que se refieren más bien, a aplicaciones particulares del modelo en diseño. Ver el Manual EQS 6.1 (Bentler, 2005) o la guía del usuario (Bentler Wu, 2002), ambas disponibles a través de la ventana de Ayuda (Help) del programa.

/TITLE (optional)

Aunque este párrafo es opcional (es decir, no se requiere para ejecutar un trabajo de EQS), no sólo es altamente recomendable usarlo ya que incluye una generosa cantidad de información en el título. Es muy común que la documentación al inicio sencilla y fácil de interpretar, con el tiempo “pierda su sentido”. El uso libre de la información del título, es particularmente útil cuando se implementan varias ejecuciones EQS para un conjunto de datos determinado. EQS permite tantas líneas como desee en este título sección. Un ejemplo de un párrafo /TITLE es el siguiente:

<p>/TITLE CFA of BDI-French Version Initial Model</p>
--

Fuente: EQS 6.2, con datos propios.

/SPECIFICATIONS (SPEC; Obligatorio)

Este párrafo define los datos a analizar, así como el método de análisis que se utilizará. En particular, detalla:

- ▶ Información relacionada con los datos (ubicación, forma de matriz);
- ▶ La cantidad de casos y aportes variables;
- ▶ El método de estimación deseado; y
- ▶ Otra información que pueda ser necesaria para asegurar que EQS, corra.

Hay tres factores importantes con respecto al párrafo/SPEC:

1. Aunque la información de entrada se puede colocar en cualquier orden, un punto y coma (;) debe delinear cada operando;
2. Información relacionada con el tamaño de muestra y el número de variables debe ser siempre provisto; y
3. Las palabras clave del subcomando se pueden abreviar para formato de tres y dos letras.

Ahora damos un vistazo general a algunos aspectos de este párrafo.

Data (DA)

El uso de esta palabra clave se rige por dos condiciones:

1. Los datos que se analizarán existen en algún archivo externo, y

2. El entorno informático implica un sistema interactivo. Como tal, la instrucción se leería: DA = 'C:\EQS61\Files\FRBDI.dat,
donde

C:\EQS61\Files representa la ubicación de los datos y
FRBDI.dat representa el nombre del archivo de datos, los cuales están encerrados dentro de citas. Como se explicó anteriormente, estos datos pueden estar en forma de matriz en bruto, de correlación o de covarianza.

Variables (VAR)

El número indicado aquí debe representar el número total de variables en el conjunto de datos; no representa la cantidad de variables que se analizarán en una ejecución específica de EQS. El programa ordena automáticamente qué variables incluir en los análisis en su lectura posterior de los párrafos: /EQUATIONS y /VARIANCES.

Cases (CAS)

Este término define el número de sujetos que comprenden los datos de muestra. El número debe representar todos los casos, independientemente que cualquiera de ellos pueda eliminarse más tarde. Tales eliminaciones, ocurren según lo especifica el usuario, así como el procedimiento analítico de EQS, nunca altera realmente el conjunto de datos original, por lo tanto, el número de casos siempre permanece igual.

Si los datos de entrada están en forma de una matriz o correlación o covarianza, es fundamental que el número de casos sea correcto porque EQS no tiene forma de verificar su precisión; un número incorrecto dará lugar a estadísticas incorrectas. Sin embargo, si los datos de entrada comprenden puntajes brutos (*raw scores*), este número no necesita ser específico; EQS registra automáticamente el número de casos en el archivo, que se usa en todos los cálculos. Cualquier discrepancia entre el número especificado y el número real de casos se observa en la salida.

Method (ME)

El usuario EQS, puede elegir alguno de los nueve métodos de estimación:

- ▶ Maximum Likelihood (ML. Máxima verosimilitud)
- ▶ Least Squares (LS. Mínimos cuadrados).
- ▶ Generalized Least Squares (GLS. Mínimos cuadrados generalizados).
- ▶ Elliptical LS (ELS. Elíptica LS).

- ▶ Elliptical GLS (EGLS. Elíptica GLS).
- ▶ Elliptical Reweighted (ERLS. Elíptica ponderada LS).
- ▶ Heterogeneous Kurtosis GLS (HGKLS. Curtosis heterogénea GLS).
- ▶ Heterogeneous Kurtosis RLS (HKRLS. Curtosis heterogénea RLS).
- ▶ Arbitrary distribution GLS (AGLS).

Si no se especifica algún método, el programa automáticamente usa la estimación de ML; en otras palabras, ML representa el método predeterminado (para detalles relacionados con estos métodos de estimación, ver el Manual de Bentler, 2005).

Una característica especial del programa EQS es que también permite la especificación de varias estimaciones métodos en un único envío de trabajo (por ejemplo, ME = ML, AGLS). Se puede especificar un máximo de dos métodos cuando AGLS está incluido y un máximo de tres métodos de lo contrario.

Los métodos, elíptico y de distribución arbitraria, siempre van precedidos automáticamente por sus contrapartidas teóricas normales. Más específicamente, los métodos ELS, EGLS y ERLS están precedidos por los métodos teóricos normales de LS, GLS y ML, respectivamente; AGLS está precedido por LS. No obstante, el usuario puede anular estos métodos predeterminados simplemente especificando otro método válido (para más detalles con respecto a la especificación y uso apropiado de estos métodos, véase Bentler, 2005).

Finalmente, una característica extremadamente valiosa exclusiva del programa EQS es la disponibilidad de estadística robusta (*robust statistics*), asociada con cualquier método de estimación seleccionado, excepto AGLS. Por ejemplo, al especificar ME = ML, ROBUST, la salida proporciona una estadística chi-cuadrado robusta (χ^2), llamada estadística escalada de Satorra-Bentler (S-B χ^2 ; Satorra Bentler, 1988, 1994) y errores estándar robustos (Bentler Dijkstra, 1985), los cuales han sido corregidos para falta de normalidad en muestras grandes. El S-B χ^2 se ha mostrado como el cálculo más aproximado a χ^2 que la prueba usual estadística y los errores estándar robustos para ser corregidos en muestras grandes a pesar del hecho de que los supuestos de distribución con respecto a las variables sean incorrectos (Bentler, 2005). Aunque estas robustas estadísticas son computacionalmente exigentes, se ha demostrado que tienen un mejor rendimiento que las estadísticas no corregidas donde el supuesto de la normalidad no se sostiene y es mejor que la asintótica GLS sin distribución (AGLS. *asymptotic distribution-free* GLS) en todas las muestras excepto en las más grandes (Chou, Bentler y Satorra, 1991; Hu, Bentler y Kano, 1992). Sin embargo, una advertencia importante sobre el uso de estadísticas robusta es que pueden ser calculadas sólo a partir de datos sin procesar (raw data).

Analysis (ANAL)

Esta palabra clave describe el tipo de matriz que se analizará si se trata de algo más que una matriz de covarianza, que es la predeterminada. Si los análisis se basan en la matriz de covarianza, pero los datos se ingresaron como una matriz de correlación, las desviaciones estándar se deben agregar a la entrada. Esto se logra proporcionando un párrafo separado introducido con la palabra clave /STANDARD DEVIATIONS (o /STA) y luego listar las desviaciones estándar en la línea siguiente, dejando un espacio en blanco entre cada una; más de una línea puede usarse si es necesario. Se deben recordar tres instrucciones al usar el párrafo /STA:

1. Debe haber exactamente el mismo número de desviaciones estándar que variables en el que se lee la matriz de correlación,
2. Las desviaciones estándar deben estar en el mismo orden que las variables en la matriz de entrada, y,
3. No se usa punto y coma (;) para finalizar esta sección de entrada.

Matrix (MA)

Esta palabra clave describe la forma de los datos de entrada, es decir, si es una correlación, covarianza o matriz de datos sin procesar (*raw data matrix*). Como se señaló anteriormente, siempre se supone que los datos brutos (*raw data*) residen en un archivo externo; las matrices de datos de covarianza o correlación se pueden incrustar en el archivo de entrada (es decir, como un archivo interno) o residen en un archivo separado (externo). Cuando las matrices de covarianza (o correlación) residen como un archivo interno, los datos se especifican en un párrafo separado, etiquetado /MATRIX (MAT). Si, por otro lado, residen en un archivo externo, su ubicación se anota a través de la palabra clave DATA (DA). Una vez que EQS ha localizado el archivo de datos, necesita saber cómo leer los datos. Si los datos están en formato libre, que es el predeterminado, EQS puede leer el archivo siempre que cada elemento en la matriz esté separado por al menos un espacio en blanco. Si los datos no están en formato libre, esta información se proporciona por medio de una declaración Fortran, que debe ser precedida inmediatamente por la palabra clave FORMAT (FO).

Finalmente, si el usuario desea que el análisis no sea con base en la matriz de covarianza (que es el valor predeterminado), la matriz que se analizará se debe anotar utilizando la palabra clave ANALYSIS (ANAL). Ahora, intentemos todo esto y observemos dos párrafos de especificación que describen una matriz de datos de entrada. Veamos los siguientes casos:

CASO 1. La matriz de datos como un archivo interno.

```
/SPECIFICATIONS  
CASE=250; VAR=4; ME=ML; MA=COR;  
/MATRIX  
1.00  
.34 1.00  
.55 .27 1.00  
.48 .33 .63 1.00  
/STANDARD DEVIATIONS  
1.09 .59 .98 1.10
```

Fuente: EQS 6.2, con datos propios.

Nota:

- ▶ La entrada de datos nunca requiere un punto y coma (;). Por lo tanto, no hay “;” en los párrafos la `/MATRIX` y `/STANDARD DEVIATIONS`.
- ▶ Aunque los datos de entrada estén en forma de matriz de correlación, la especificación de las desviaciones estándar permite que los análisis se basen en la matriz de covarianza.

CASO 2. La matriz de datos como un archivo externo.

```
/SPECIFICATIONS  
CASE=250;VAR=4; ME=ML, ROBUST; MA=COV;  
DA='C:\EQS61\Files\FRBDI.dat'; FO=(4F2.0);
```

Fuente: EQS 6.2, con datos propios.

Además de los subcomandos del párrafo `/SPEC` que se muestran aquí, otros son opcionales o usados con aplicaciones particulares.

/LABELS (opcional)

Este párrafo se puede usar para identificar los nombres de las variables observadas (Vs) y/o latentes (Fs) en el modelo. Las etiquetas pueden tener de 1 a 8 caracteres de longitud y están asignadas sólo a las variables de tipo V y F. Las variables observadas deben numerarse de acuerdo con su posición en el conjunto de datos. Por lo tanto, la especificación que `V5 = MATH` indica que la quinta variable en la matriz de datos debe etiquetarse como “MATH”. Aunque la numeración de variables latentes es arbitraria, debe ser secuenciada lógicamente dentro del contexto del modelo. Porque EQS asigna automáticamente V1 a la primera variable observada, V2 a la segunda variable, y así en adelante, estas designaciones se utilizan como etiquetas si el usuario no proporciona nombres para las variables; del mismo modo a las variables latentes se les asignan automáticamente las etiquetas F1, F2, etc. Si el

usuario proporciona etiquetas sólo para algunas de las variables, estos nombres anulan las etiquetas predeterminadas (por ejemplo, V1).

/EQUATIONS (EQU; Obligatorio)

Esta palabra clave señala información de especificación con respecto al modelo en estudio. Específicamente, el párrafo define cada trayectoria de regresión en el modelo. Por medio de una serie de declaraciones de ecuación, el párrafo */EQU* especifica todos los vínculos entre las variables independientes y dependientes (y viceversa) e identifica aquellos parámetros a restringir (es decir, a cero, 1.0 o algún otro valor) y aquellos que se estiman libremente. Antes de completar esta parte del archivo de entrada, es fundamental construir un diagrama de trayectoria de su modelo en que:

- a. Todas las variables observadas y latentes están claramente etiquetadas;
- b. Todas las trayectorias de regresión estructural son especificadas;
- c. Se especifican todos los términos de error y perturbación, junto con sus trayectorias de regresión relacionadas;
- d. Todas las covarianzas hipotéticas (entre las variables independientes); y
- e. Todos los parámetros a ser estimados (incluidas las varianzas de factores) se identifican por medio de un asterisco (*). Una vez que obtenga esta visual representación de su modelo, es fácil completar todas las ecuaciones en el párrafo */EQU* simplemente leyendo fuera del diagrama de trayectoria.

Nota: Con EQS ahora hay un párrafo recientemente establecido: */MODEL* que permite al usuario expresar las especificaciones del modelo de una manera más simplificada y manera concisa. En adición, las variables E y D son generadas automáticamente por el programa. Cuando se utiliza este párrafo, reemplaza a la palabra clave */EQU*, también como las palabras clave */VARY* */COV*.

EQS. Cómo escribir ecuaciones

Como se señaló, la finalización de estas ecuaciones se lleva a cabo de la siguiente manera:

Para cada variable dependiente (es decir, cualquier variable que tenga una flecha apuntando hacia ella), escriba una ecuación que resuma el impacto directo sobre ella de otras variables en el modelo. Por lo tanto, siempre habrá tantas ecuaciones en el párrafo */EQUATIONS* como variables dependientes en el modelo probado. La variable dependiente siempre estará en el lado izquierdo de la ecuación, con todas las variables independientes (es decir, explicativas) que aparecen a

la derecha (recordemos que en EQS, los términos de variables dependientes e independientes, se definen en el contexto del modelo de Bentler-Weeks). Finalmente, una vez las ecuaciones se han formado, asegúrese de insertar asteriscos (*) junto a todos los parámetros que se estimarán. EQS proporciona estimaciones sólo para los parámetros con asterisco (*); todos los demás se consideran fijos.

Se debe tomar en cuenta el uso de los valores de inicio, que se refieren al punto en el cual un programa comienza un proceso iterativo para establecer estimaciones de parámetros. Los usuarios pueden permitir que el programa suministre estos valores (EQS usa el valor predeterminados tales como 0.0 para covarianzas) o proporcionar sus propios valores. Los valores de inicio proporcionados por el usuario representan una mejor estimación de cuál será el valor esperado de una estimación de parámetro en particular. Estos valores de mejor estimación son incluidos en la ecuación, actuando como modificadores de los parámetros a estimar; siempre precediendo al asterisco (*). Aunque no es necesario especificar los valores de inicio para la mayoría de los trabajos de EQS, a menudo facilitan el proceso iterativo. Cuando los modelos complejos están bajo estudio; tales modelos pueden generar problemas de no convergencia si el inicio de los valores proporcionados por EQS son inadecuados. Sin embargo, para el usuario de EQS, el problema de tratar de determinar los valores de inicio apropiados se han ido para siempre como resultado de una característica ingeniosa llamada la opción RETEST, que genera automáticamente estos valores de inicio y luego le permite al usuario editar el archivo de entrada en consecuencia. (Para una discusión más extensa de los valores de inicio, véase Bollen, 1989). Algunos párrafos para examinar, son los casos:

CASO 1. Ecuaciones sin valores de inicio.

/EQUATIONS

VI = FI + EI;

F3 = *F1 + *F2 + D3;

Fuente: EQS 6.2, con datos propios..

CASO 2. Ecuaciones con valores de inicio.

/EQUATIONS

VI = .9 FI + EI;

F3 = .6 *F1 + -.2*F2 + D3;

Fuente: EQS 6.2, con datos propios..

Nota:

El valor de inicio asociado con F1 se fija en .9 por elección.

Los valores de inicio asociados con E1 y D3 no tienen ningún valor especificado; estos parámetros son considerados conocidos y fijados a 1.00 por defecto del programa.

EQS/VARIANCES (VAR; Obligatorio)

Este párrafo especifica el estado de las varianzas relacionadas con las variables independientes en el modelo. Como tal, cada varianza debe identificarse ya sea como un parámetro fijo en el modelo, o como uno que se estima libremente. (Recuerde que las varianzas para las variables dependientes nunca se especifican, independientemente de si son fijas o libres). Como en el párrafo /EQUATIONS, las varianzas a ser estimadas se identifican con un asterisco (*). Veamos un par de casos del párrafo /VAR.

CASO 1. Varianzas sin valores de inicio.

/VARIANCES

F1, F2 = *;

F1 to F3 = *;

F4-F6 = *;

Fuente: EQS 6.2, con datos propios..

Nota:

- ▶ Las listas consecutivas de variables se pueden especificar usando “to” o un guión (-). Como tal, F1 to F3 indica que se deben estimar las varianzas de F1, F2 y F3. Del mismo modo, F4, a través de F6 especifica que las varianzas de F4, F5 y F6 deben ser estimadas.
- ▶ Las etiquetas de variables (como se define en el párrafo /LABEL) se pueden usar en lugar de las etiquetas V o F. Sin embargo, dado que las etiquetas con menos de ocho caracteres pueden causar un error, se recomienda que, en este caso, se use el guión (-) en lugar de la convención “to”.

CASO 1. Varianzas sin valores de inicio.

/VARIANCES

F1, F2 = .3*;

F1 to F3 = .4*;

F4-F6 = .6*;

Fuente: EQS 6.2, con datos propios..

EQS/CCOVARIANCES (OV; Opcional)

Por razones obvias, este párrafo es necesario sólo cuando se especifican covarianzas en el modelo. Así, se usa para especificar tanto covarianzas fijas distintas de cero como libres entre las variables independientes. Sin embargo, cualquier variable involucrada en una covarianza también debe tener su varianza especificada en el párrafo /VAR; en consecuencia, las covarianzas tampoco se pueden especificar para las variables dependientes. (Si desea obtener dos variables dependientes covariantes, una opción es especificar su covarianza entre sus términos de perturbación porque éstos, los residuos son siempre variables independientes en un modelo). Al especificar una covarianza, el par de variables se establece y se separa por una coma, como se ejemplifica a continuación:

/COVARIANCES

E1, E3 = *;

F1 to F3 = *;

F4-F6 = *;

Fuente: EQS 6.2, con datos propios..

Nota:

El uso de “to” y “-” indica que las covarianzas relacionadas con todos los posibles pares de variables, deben ser estimadas. Por ejemplo, la especificación de “F1 to F3 = *” indica que los parámetros estimados son F1, F2, F1, F3, F2, F3.

El último párrafo de todos los archivos de entrada de EQS debe ser /END. Esta palabra clave marca la terminación del programa entrada.

EQS. Casos de ejemplos de archivos de entrada

Para proporcionar una vista más completa de cómo las diversas secciones del archivo de entrada EQS se relacionan con el diagrama de trayectoria de un modelo en particular, ahora revisamos tres modelos simples aunque diferentes:

- ▶ modelo CFA de primer orden (figura 4),
- ▶ modelo CFA de segundo orden (figura 5) y
- ▶ modelo SEM completo (figura 6).

Sin embargo, debido a limitaciones de espacio, solo se muestra el archivo de entrada completo para la figura 4; los párrafos de /SPECIFICATIONS y /LABELS están ausentes para los dos modelos restantes. Aunque parte de los mismos archivos de entrada aparecen más adelante en el capítulo, se presentan en su forma original

según lo estructurado por la opción BUILD_EQS. Por lo tanto, los archivos presentados en esta subsección proporcionan una idea de cómo hacer que el archivo sea más conciso. Para obtener el máximo beneficio de esta subsección, estudie cada figura mientras se refiere a su declaración de entrada respectiva.

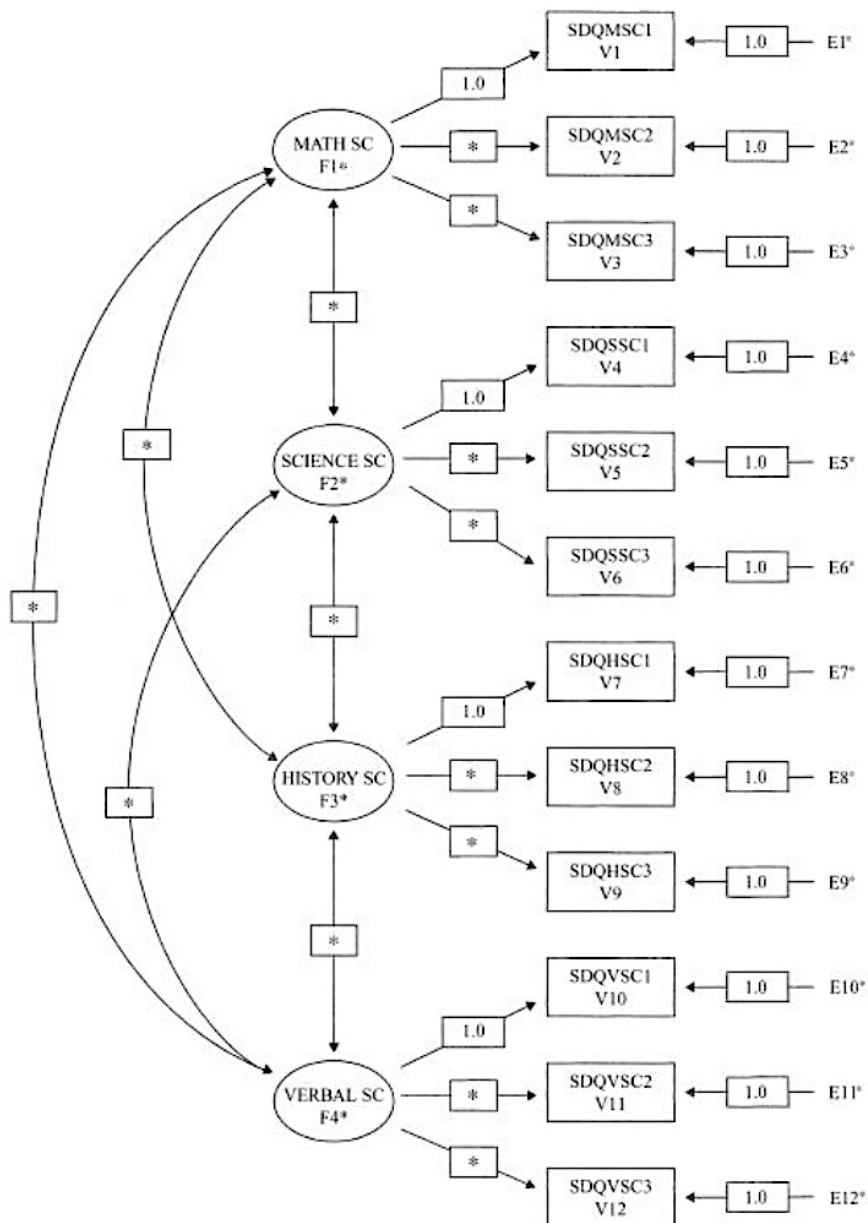
CFA. Modelo de primer orden

Aunque este archivo de entrada parece ser bastante claro, tres características son dignas de hacer notar (Partiremos de observar y analizar la figura 4):

1. La convención “to” se ha utilizado en los párrafos /VARIANCES y /COVARIANCES, indicando que las varianzas de todas las variables independientes (F1-F4) deben estimarse libremente, así como sus covarianzas.
2. Recuerde que una variable independiente puede tener, ya sea su trayectoria o su varianza estimada pero NO ambas. Por lo tanto, debido a que solo estamos interesados en las varianzas de los términos de error, tenga en cuenta que en el párrafo /EQUATIONS, sus ponderaciones beta relacionadas sean asignadas automáticamente por el programa a valor de 1.0. Aunque estos valores no se muestran en el archivo de entrada (vea E1-E12), son visibles en la figura 4.
3. Finalmente, tenga en cuenta que el primer indicador de medición para cada factor (V1, V4, V7, V10) se ha especificado como fijo (es decir, no hay asterisco junto a estos parámetros). Como se muestra en la figura 4, estos parámetros son limitados a un valor igual a 1.0; EQS asigna automáticamente este valor a estos parámetros fijos, por lo tanto, el valor de 1.0 no aparece en el archivo de entrada. Dos puntos deben hacerse a este respecto: (a) los valores asignados a estos parámetros no necesitan ser 1.0, -aunque se puede asignar cualquier número a estos parámetros, el valor de 1.0 ha sido típicamente la asignación de elección; y (b) el parámetro restringido no necesita ser limitado a la primera variable indicadora; cualquiera de un conjunto de parámetros congénéricos puede ser elegido (por ejemplo, como se muestra en la figura 4. SDQMSC1, SDQMSC2 y SDQMSC3 sirven todos como medidas de la variable MATHSC, por lo tanto, representan un conjunto congénérico de variables indicadoras). Más allá estas notaciones técnicas, sin embargo, el punto más importante con respecto a las cargas de factoriales fijas es que abordan el problema de la identificación del modelo o *model identification* (también denominado *identificación estadística-statistical identification*).

Lo fundamental a saber, es si un modelo al estar estadísticamente identificado comprende el número de estimaciones de parámetros en el modelo. Por lo tanto, una advertencia muy importante al trabajar con modelos de ecuaciones estructurales es contar siempre con el número de parámetros

FIGURA 4. Modelo CFA de primer orden con notación EQS y valores asignados de inicio.



Fuente: Byrne (2006).

```

/TITLE
Example 1st-order CFA Input File
Initial Model
/SPECIFICATIONS
DA=6C:\EQS61\FILES\CFASC.DAT; FO=(12F1 .0); CASE=361; VAR=12; ME=ML; MA=RAW;
/LABELS
V1=SDQMSC1;      V2=SDQMSC2;      V3=SDQMSC3;
V4=SDQSSC1;      V5=SDQSSC2;      V6=SDQSSC3;
V7=SDQHSC 1;     V8=SDQHSC3;      V9=SDQHSC3;
V10=SDQVSC1;     V11=SDQVSC2;     V12=SDQVSC3;
F1=MSC; F2=SSC; F3=HSC; F4=VSC;

/EQUATIONS
V1 = F1 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
      V4 = F2 + E4;
      V5 = *F2 + E5;
      V6 = *F2 + E6;
            V7 = F3 + E7;
            V8 = *F3 + E8;
            V9 = *F3 + E9;
                  V10 = F4 + E10;
                  V11 = *F4 + E11;
                  V12 = *F4 + E12;

/VARIANCES
F1 to F4 = *;
E1 to E12 = *;
/COVARIANCES
F1 to F4 = *;
/END

```

Fuente: Byrne (2006).

libremente estimados antes de ejecutar los análisis. Como un prerrequisito a la discusión de la identificación del modelo, entonces, vamos a contar la cantidad de parámetros a estimar para el modelo de la figura 4 en la que encontramos que hay 12 coeficientes de regresión (cargas factoriales) y 16 varianzas (12 varianzas de error, 4 varianzas de factor) y 6 covarianzas de factor. Los 1s asignados a uno de cada conjunto de parámetros de trayectoria de regresión representan un valor fijo de 1.00; como tal, estos los parámetros NO deben ser estimados. En total, entonces, hay 30 parámetros para ser estimados para el modelo CFA representado en la figura 4. A continuación, se presenta una breve discusión del importante concepto de identificación de modelos.

El concepto de identificación de modelo

La identificación de modelo es un tema complejo, difícil de explicar en términos no técnicos. Aunque una explicación completa del principio de identificación excede el alcance de este libro, no es crítico para la comprensión y el uso por parte del lector, lo que sigue es una breve explicación “no matemática” de la idea básica que subyace a este concepto. Esencialmente, la llamada “t- rule” se aborda, una de las varias pruebas asociadas con la identificación. En términos generales, el problema de la identificación se centra en si existe un conjunto único de parámetros consistente con los datos. Esta pregunta influye directamente en la transposición de la matriz de varianza-covarianza de variables observadas (los datos) en los parámetros estructurales del modelo en estudio. Si es determinada una solución única para los valores de los parámetros estructurales, el modelo se considera identificado. Como consecuencia, los parámetros se consideran estimables y el modelo, por lo tanto, comprobable. Por otra parte, si un modelo no puede identificarse, indica que los parámetros están sujetos a arbitrariedad, lo que implica que diferentes valores de parámetros definen el mismo modelo. Si es el caso, el logro de la coherencia de las estimaciones para todos los parámetros no son posibles y, por lo tanto, el modelo no puede evaluarse empíricamente. Por medio de un ejemplo simple, el proceso sería conceptualmente similar a tratar de determinar valores únicos para X e Y cuando la única información disponible es que $X + Y = 15$. Generalizando este ejemplo a la estructura del análisis de covarianza, el problema de identificación del modelo se centra en la medida en que se puede inferir un conjunto único de valores para los parámetros desconocidos de una matriz de covarianza dada de las variables analizadas reproducidas por el modelo.

Los modelos estructurales pueden ser simplemente identificados, sobreidentificados o no identificados. Así:

- a. Un modelo identificado (just-identified), es uno en el que hay una correspondencia uno a uno entre los datos y los parámetros estructurales. Es decir, el número de varianzas y covarianzas de datos es igual a la cantidad de parámetros a estimar. Sin embargo, a pesar de la capacidad del modelo de ofrecer una solución única para todos los parámetros, este modelo no es científicamente interesante porque no tiene grados de libertad y, por lo tanto, nunca puede ser rechazado.
- b. Un modelo sobreidentificado (overidentified) es aquel en el que el número de parámetros a estimar es menor que el número de puntos de datos (es decir, varianzas y covarianzas de las variables observadas). Esta situación resulta en grados positivos de libertad que permiten el rechazo del modelo, por lo que

se hace de uso científico. El objetivo en SEM es el especificar un modelo tal que cumpla con el criterio de sobreidentificación.

- c. Finalmente, un modelo no identificado (underidentified), es uno en el que el número de parámetros a estimar excede el número de varianzas y covarianzas (es decir, puntos de datos). Como tal, el modelo contiene información insuficiente (datos de entrada) para lograr una solución determinada de estimación de parámetros; es decir, es posible un número infinito de soluciones para un modelo no identificado.

Al revisar el modelo CFA en la figura 4, determinemos con cuántos puntos de datos hay que trabajar (es decir, ¿cuánta información tenemos con respecto a nuestros datos?). Como se señaló anteriormente, estos puntos de datos constituyen las varianzas y covarianzas de las variables observadas; con las variables p , hay $p(p + 1) / 2$ de tales elementos. Dado que hay doce variables observadas, esto significa que hay $12(12 + 1) / 2 = 78$ puntos de datos. Antes de esta discusión de identificación, determinamos un total de 30 parámetros desconocidos. Así, con 78 puntos de datos y 30 parámetros para estimar, tenemos un modelo sobreidentificado con 48 grados de libertad. Sin embargo, es importante observar que la especificación de un modelo sobreidentificado es necesaria, pero no condición suficiente para resolver el problema de identificación. De hecho, la imposición de restricciones en particular, los parámetros, a veces pueden ser beneficiosos para ayudar al investigador a alcanzar un modelo sobreidentificado.

El problema de la escala adecuada

Vinculado con el problema de la identificación está el requerimiento de que cada variable latente debe tener su escala determinada. Este requisito surge porque las variables latentes no se observan y, por lo tanto, no tienen escala métrica definida. Este requisito se puede lograr de una de dos maneras:

- a. Mediante el enfoque con que se inicia la presente obra de determinar: dimensionalidad, confiabilidad y validez al constructo para su prueba de EFA y CFA.
- b. Ligar la especificación del modelo de medición, mediante el cual la variable latente no medida se mapea en su variable del indicador observado relacionado. Este requisito de escalamiento se cumple restringiendo a un valor distinto de cero (típicamente 1.0), un parámetro de carga de factorial en cada conjunto congénico de cargas. Esta restricción vale para variables latentes tanto independientes como dependientes. Al revisar la figura 4, esto significa que para una de las tres trayectorias de regresión que conducen

desde cada factor (SC) a un conjunto de indicadores observados, algunos valores fijos deben ser especificados; este parámetro fijo se denomina variable de referencia (*reference variable*). Con respecto al modelo en figura 4, por ejemplo, la escala se estableció limitando a un valor de 1.0, el primer parámetro en cada conjunto de variables observadas. Con una mejor idea de los aspectos importantes de la especificación de un modelo CFA en general, con la especificación que se utiliza del programa EQS en particular, y las nociones básicas asociadas con la identificación del modelo, es posible analizar los siguientes modelos de entrada.

CFA. Modelo de segundo orden

Este modelo apoya la noción de que las variables MATH SC (matemáticas de auto-concepto *math self-concept*), SCIENCE SC (ciencia de autoconcepto *science self-concept*), HISTORY SC (historia de auto-concepto *history self-coincept*) y los VERBAL SC (autoconceptos verbales *verbal self-concept*) son causados por el constructo de ACADEMIC SC (autoconcepto académico *academic self-concept*) global de orden superior. Partiremos de observar y analizar la figura 5.

Por lo tanto, aunque la estructura de primer orden de éste el modelo sigue siendo básicamente el mismo que en la figura 4, se debe tomar en cuenta cuatro características importantes:

1. Primero, en contraste con el ejemplo CFA de primer orden en el que todas las ecuaciones en el párrafo /EQUATIONS representaban las trayectorias de regresión entre todas las variables observadas y sus factores subyacentes relacionados, las ecuaciones especificadas para éste el modelo de segundo orden también incluyen trayectorias de regresión entre factores. Específicamente, estos parámetros representan el impacto de F5 (una variable independiente) en F1, F2, F3 y F4.
2. Dado que la estimación de todas las cargas factoriales de orden más alto son típicamente de interés en modelos de segundo orden, la varianza del único el factor de orden superior, F5, se ha limitado a 1,0, como se especifica en el párrafo /VARIANCES. (Alternativamente, si la varianza de F5 hubiera sido de interés, una de las cargas de orden superior tendría que estar limitada a 1.0 (u otra distinta de cero). Ésta la restricción, por supuesto, aborda el problema de la identificación del modelo discutido anteriormente.
3. En tercer lugar, dado que se supone que la variable ACADEMIC SC causa cada uno de los cuatro factores de primer orden, F1 a F4 ahora representan variables dependientes en el sentido de Bentler-Weeks. Como tal, MATH SC,

SCIENCE SC, HISTORY SC son las hipótesis predichas por el académico, pero con algún grado de error, que es capturado por el término de perturbación asociado con cada uno de estos factores. Por lo tanto, en el párrafo /VARIANCES, tenga en cuenta también que las varianzas de las perturbaciones (las D's) se designan como estimadas libremente. Relacionado, sus caminos son restringidos automáticamente a 1.0 por el programa, como se muestra en la figura 5.

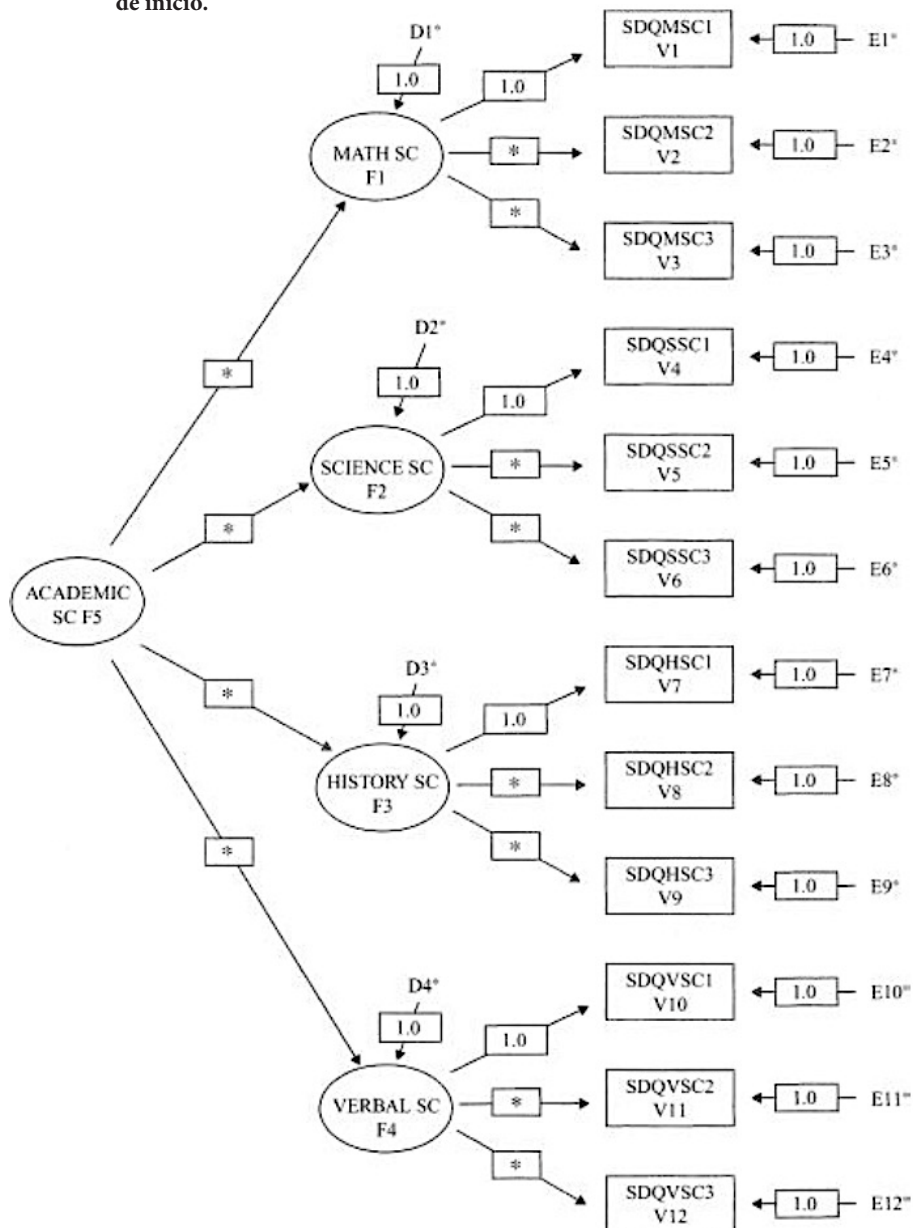
4. Finalmente, en los modelos de segundo orden, cualquier covarianza supuesta entre los factores de primer orden se explica por el (los) factor (es) de orden superior. En consecuencia, observe la ausencia de flechas de doble punta que unen los cuatro factores de primer orden en el diagrama de trayectoria así como la ausencia de un párrafo /COVARIANCES en el archivo de entrada.

SEM. Modelo completo

Al revisar tanto el diagrama de trayectoria, así como el archivo de entrada para este modelo completo de SEM, tenga en cuenta tres particularidades (partiremos de observar y analizar la figura 6):

1. De los cuatro factores que componen este modelo, sólo F1 y F2 son variables independientes (en el sentido de Bentler-Weeks); todos los demás son variables dependientes en el modelo. En consecuencia, sólo las varianzas para F1 y F2, así como sus covarianzas, se pueden estimar como se especifica en el archivo de entrada.
2. En segundo lugar, como el modelo de segundo orden que se muestra en la figura 5, una vez más, las ecuaciones de regresión involucran conjuntos de factores. En el presente caso, estas ecuaciones se especifican sólo para F3 y F4 porque cada una se explica por otros factores en el modelo.
3. En tercer lugar, dado el estado de la variable dependiente de F3 y F4 en el modelo, tenga en cuenta que la varianza de ninguno de ellos está especificado en el párrafo/VARIANCES; relacionado con esto, tenga en cuenta la especificación de las varianzas asociadas con sus respectivas perturbaciones, D3 y D4, respectivamente. Finalmente, aunque F3 parece funcionar como una variable independiente, así como una dependiente, esto no es así. Una vez que una variable se define como variable dependiente en un modelo, mantiene esa designación en todos los análisis relacionados con la hipótesis modelo.

FIGURA 7.5. Modelo hipotético CFA de segundo orden con notación EQS y valores asignados de inicio.



Fuente: Byrne (2006).

```

/TITLE
Example 2nd-order CFA Input File
Initial Model
↓
/EQUATIONS
V1 = FI +E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
    V4 = F2 + E4;
    *F2 + E5;
    *F2 + E6;
        V7 = F3 + E7;
        V8 = *F3 + E8;
        V9 = *F3 + E9;
            V10 = F4 + E10;
            V11 = *F4 + E11;
            V12 = *F4 + E12 ;

FI = *F5 + D1;
    F2 = *F5 + D2;
        F3 = *F5 + D3;
            F4 = *F5 + D4;

/VARIANCES
F5 = 1.0;
D1 to D4 = *;
E1 to E12 = *;

/END

```

Fuente: Byrne (2006).

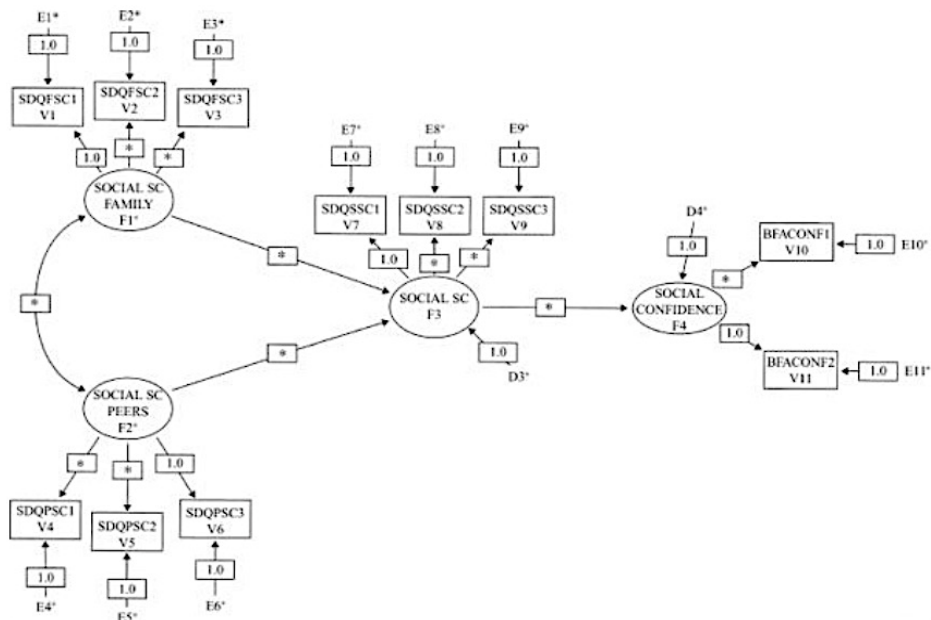
CFA. Modelo de primer orden. Análisis de caso de estudio 1

Recordar que el CFA es un método estadístico multivariable cuyo propósito principal es definir la estructura subyacente en una matriz de datos.

Su objetivo principal es seleccionar factores que expliquen las interrelaciones entre las variables. Las variables originales juegan el papel de variables dependientes que se explican por factores comunes y únicos, que no son observables.

En resumen: sirve para medir modelos teóricos. Véase el modelo del diagrama de trayectoria (figura 7).

FIGURA 6. Modelo hipotético SEM con notación EQS y valores asignados de inicio.



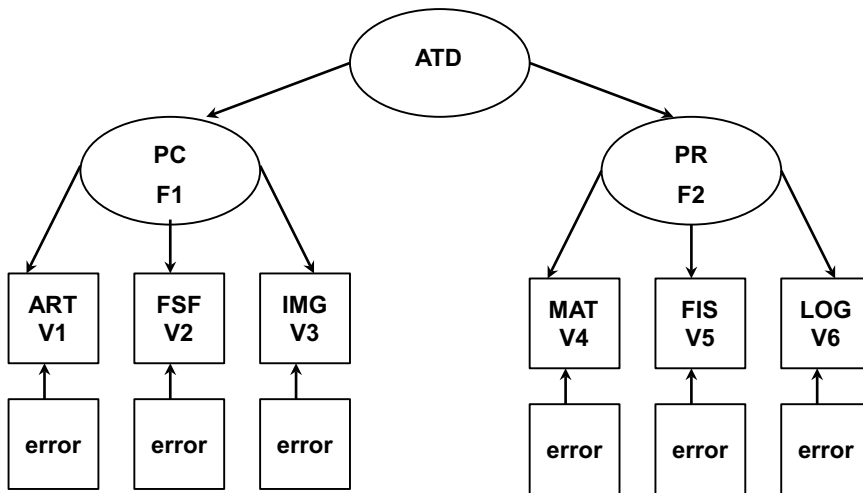
```

/TITLE
Example Full SEM Input File
Initial Model
↓
/EQUATIONS
V1 = F1 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
    V4 = F2 + E4;
    *F2 + E5;
    *F2 + E6;
        V7 = F3 + E7;
        V8 = *F3 + E8;
        V9 = *F3 + E9;
            V10 = F4 + E10;
            V11 = *F4 + E11;

F3 = *F1 + *F2 + D3;
F4 = *F3 + D4;
/VARIANCES
F1 = *; F2 = *;
D3 = *; D4 = *;
E1 to E11 = *;
/COVARIANCES
F1, F2 = *;
/END
    
```

Fuente: Byrne (2006).

FIGURA 7. Diagrama de trayectoria de la asertividad en la toma de decisiones (ATD).



Fuente: Elaboración propia.

CFA. Problema

Suponga que en una investigación, se han recogido las notas de 275 ejecutivos de la industria del software, calificados en seis áreas: Arte (ART); Filosofía (FSF); Imaginación (IMG); Matemáticas (MAT); Física (FIS); y Lógica (LOG). El investigador no posee ninguna hipótesis al respecto de si la asertividad en la toma de decisiones (ATD) es unidimensional o bidimensional (pensamiento creativo y pensamiento racional). Suponiendo que las notas miden la (ATD), tiene la siguiente hipótesis: “La ATD de los ejecutivos de la industria del software es explicada mediante las notas del ART; FSF e IMG que componen al PC, así como las notas de MAT; FIS y LOG que componen al PR”.

Dentro de la encuesta de captura de datos, se tienen entre otros: ventas, número de empleados (ambos como indicadores de crecimiento), años de funcionamiento en la empresa y en el puesto de dirección (se analizan empresas jóvenes y/o maduras de 10 a más años, así como sus ejecutivos), si es o no familiar (para distinguir las pymes de las que no), género (estudios de género pertinentes), edad de los ejecutivos (*gerentes maduros menos emprendedores a los jóvenes?*), grado de estudios del directivo, entre otros relevantes. Aunque el modelo es general, dichas preguntas de segmentación ya aportan datos importantes, incluso por cada uno de los factores

por separado. Con esto, es factible utilizar diversos factores parciales “parcelar” en modelos lineales, es decir, sólo con dos factores. Siguiendo la fórmula:

$$N = \frac{n(n - 1)}{2}$$

donde:

n = número de factores.

N = cantidad de relaciones lineales.

Eventualmente, sería posible hacer artículos que describan de constructos lineales esto es, “parcelar”, del modelo para explicar cualquier fenómeno relacionado, hasta lograr la integración de todo el modelo completo sugerido. Lo importante es el modelo, no tanto los datos (es posible conservar bases de datos con más de cuatro años de antigüedad, aún útiles para la investigación). Cabe señalar, que en CFA NO se tienen hipótesis (modelo de medición), puesto que se requiere ajustar (confiabilidad y validez) el modelo. Hecho lo anterior, ya es posible determinar las hipótesis y probarlas con la parte de modelo estructural. Al momento, es más práctico trabajar el CFA con matrices de covarianza que de correlación.

CFA. Formulación de matriz

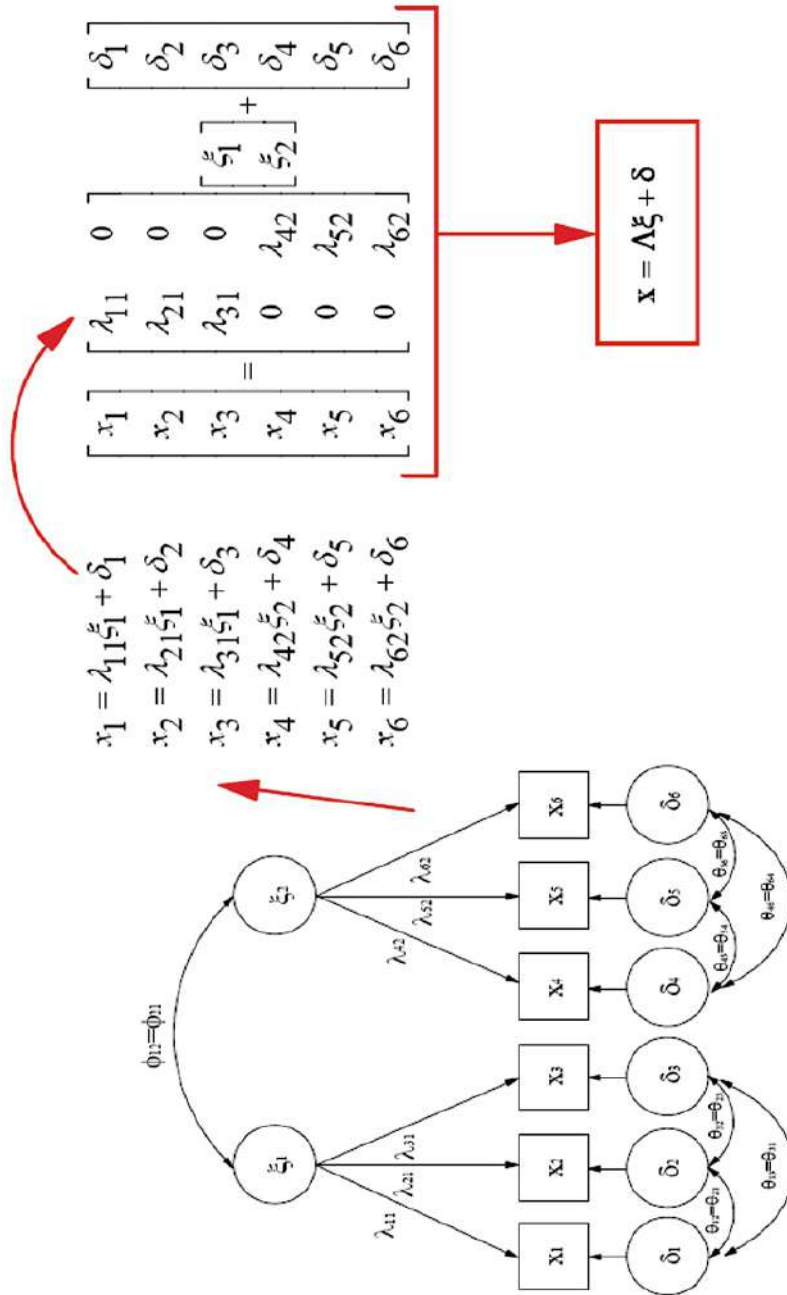
Aunque está más allá de los alcances de esta obra la explicación detallada, una formulación matemática del CFA se muestra en la figura 8.

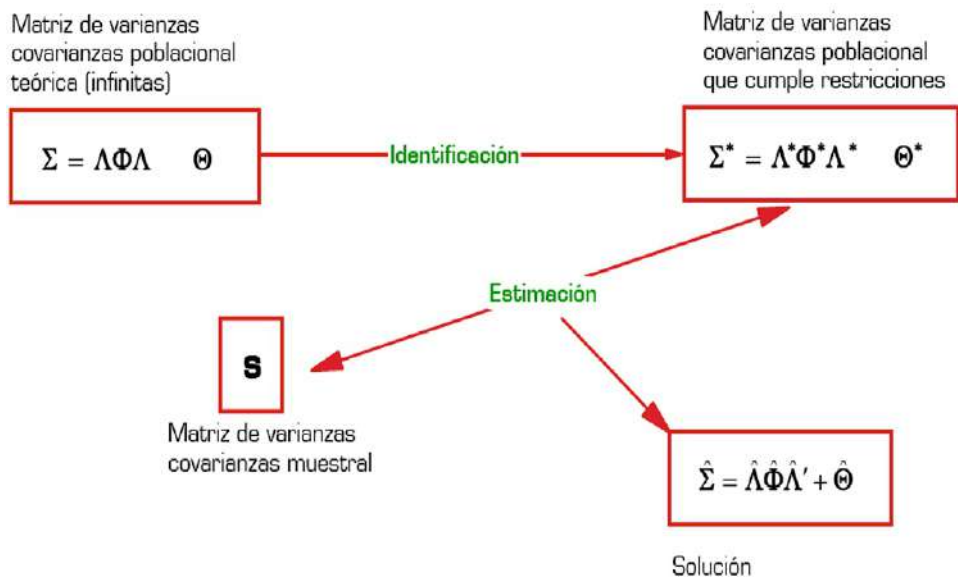
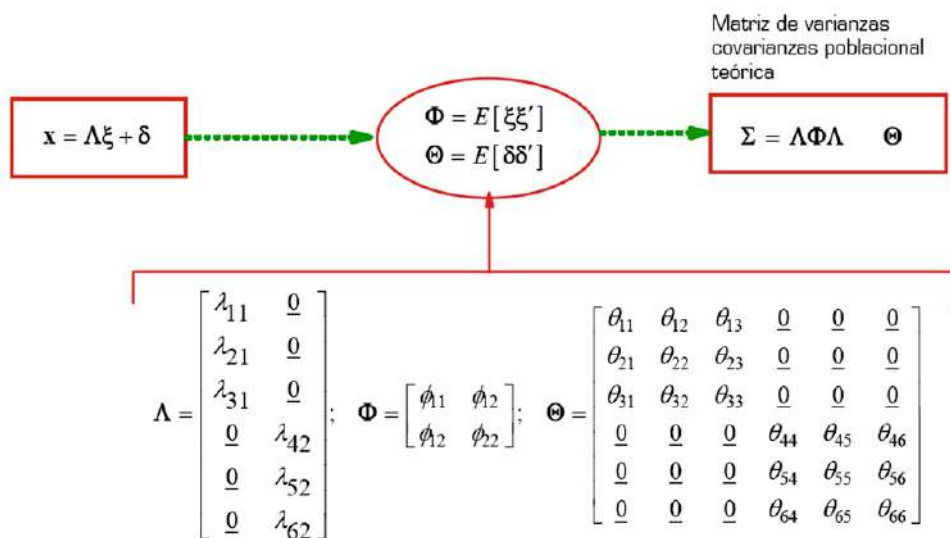
CFA. Condiciones de aplicabilidad

A este momento se deben confirmar las condiciones de aplicabilidad, tales como:

- ▶ Las variables manifiestas (los ítems de las distintas escalas) deben ser variables métricas.
- ▶ Debe existir la normalidad, homoscedasticidad y linealidad.
- ▶ Las relaciones entre las variables deben ser lineales y aditivas
- ▶ No debe existir problema de multicolinealidad (asegurar variabilidad, evitar resultados sesgados).
- ▶ El modelo debe ser sobreidentificado.
- ▶ Debe existir un número mínimo de observaciones (Hair *et al.*, 1999 = 100 y Bagozzi y Yi, 1988 =200), o bien cinco observaciones por cada parámetro. Deben existir al menos tres indicadores (variables manifiestas) por cada variable latente (factor o dimensión).

FIGURA8. CFA. Formulación de matriz.



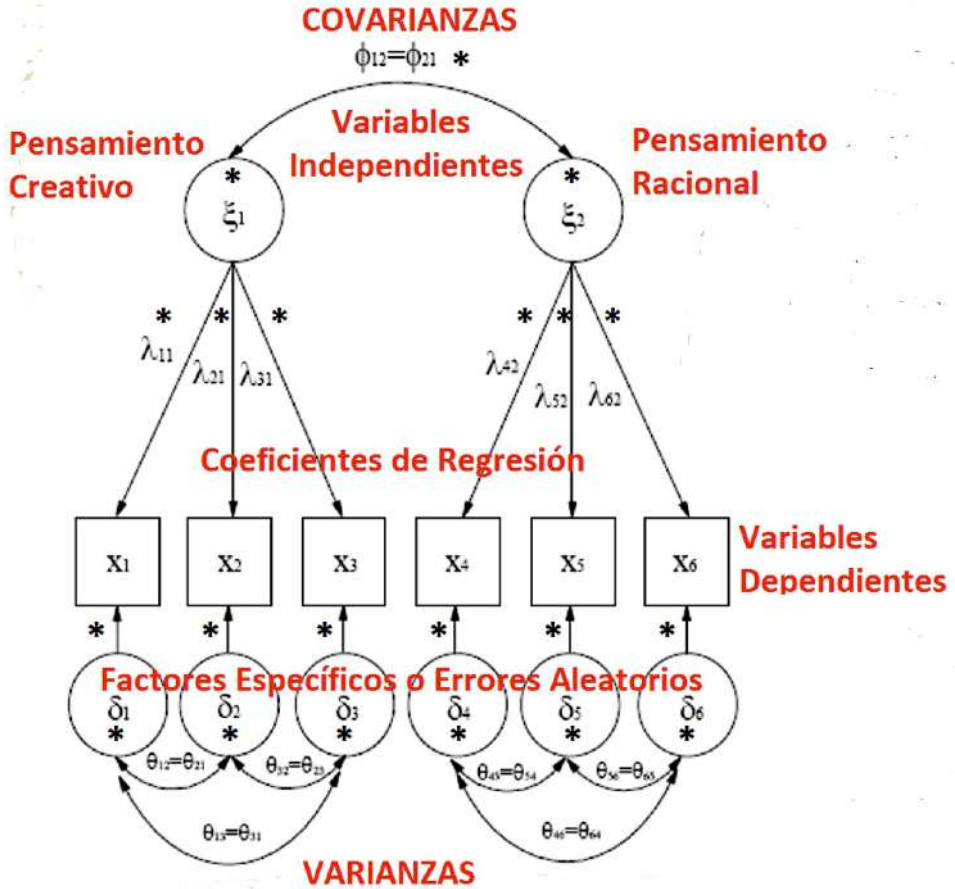


Fuente: Jöreskog (1993); Hair (et al., 1999); Byrne (2006), con adaptación.

- Pueden existir un máximo de 20 a 30 indicadores como ideal.

Por lo que una representación matemática del CFA bajo análisis, se muestra en la figura 9.

FIGURA 9. CFA aproximación matemática del diagrama de trayectoria del estudio de caso.



Fuente: Jöreskog (1993); Hair et al., (1999); Byrne (2006), con adaptación propia.

donde:

Variables observadas o manifiestas o indicadores, son V1 a V6: ART, FSE, IMG, MAT, FIS, LOG (representadas dentro de un cuadrado).

Variables latentes o no observables o subyacentes, como:

- ▶ Factores comunes: (compartidos por más de una variable observada): pensamiento creativo (PC) y pensamiento racional (PR).
- ▶ Factores específicos o errores: aleatorios que pueden haberse producido en la medida de una variable observada (representadas con un cuadrado).
- ▶ Covarianzas entre las variables latentes: representadas con flecha de doble punta.
- ▶ Coeficientes de regresión, trayectoria, carga factorial: representadas con flecha de una sola punta.

Como se observará, la realización de un CFA, sugiere realizar los siguientes pasos (figura 9):

1. Especifique un término de error por cada variable dependiente.
2. Sólo las variables independientes, pueden tener covarianza (y/o varianza. punto 4).
3. Las variables independientes NO tienen término de error.
4. Se deben estimar las varianzas para cada variable independiente del modelo.
5. Se deben estimar las covarianzas para cada par de variables independientes.
Nota: valores $>.8$ de covarianza NO son deseables ya que indicaría que los factores tienden a medir lo mismo; sin embargo $<.1$ NO son deseables ya que indicaría que los factores serían interdependientes, porque juntas deben medir un mismo constructo.
6. No se estiman las covarianzas de los términos de error (éstos son también variables independientes), dado que lo interesante es la covarianza entre factores, no entre errores.
7. Hay que estimar los coeficientes de regresión.
8. El símbolo de asterisco (*) es para indicar que el software realice la estimación del valor de la relación.

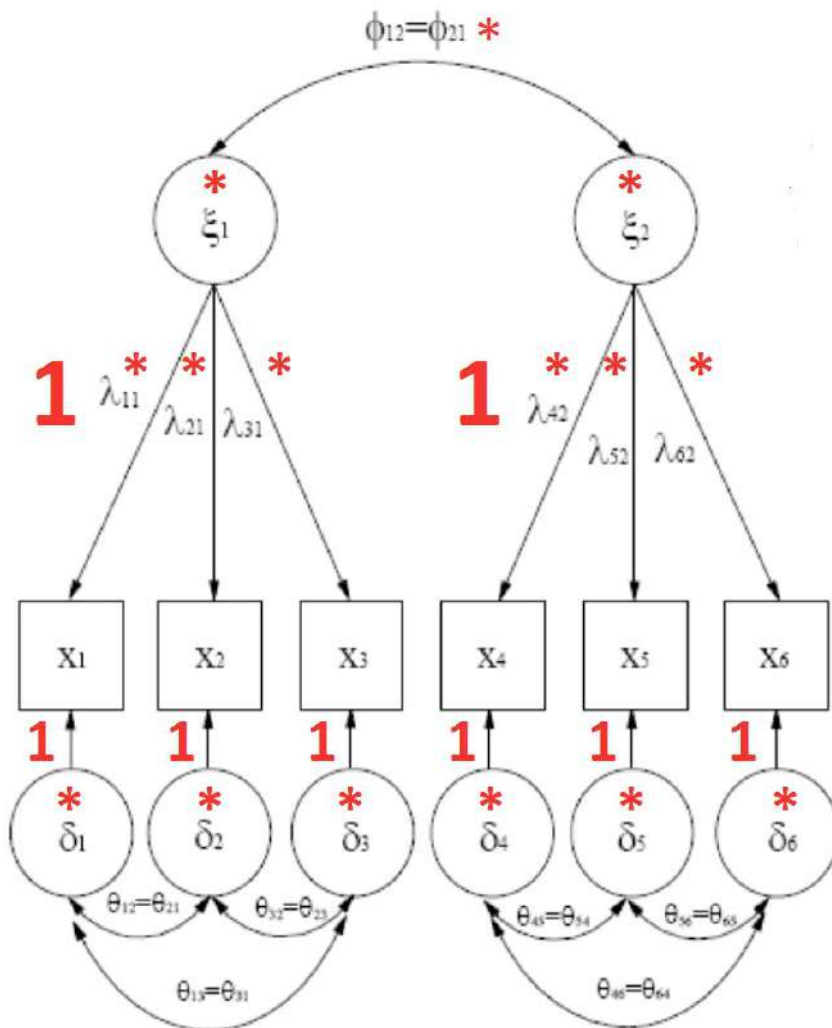
CFA. Identificación

Se plantea la pregunta: ¿pueden los parámetros del modelo ser determinados de manera única? Se debe establecer una escala para los factores comunes (esto es, evitar indeterminación entre la varianza y las cargas factoriales), mediante:

- ▶ Fijar la carga factorial a 1 (al menos 1 variable de cada factor debe tener varianza 1).
- ▶ Fijar la varianza a 1 (1 a todos los errores...NO ambas).
- ▶ Fijar arbitrariamente el coeficiente de regresión del término de error a 1.
- ▶ Comparar el número de datos con el número de parámetros que han de estimarse.

- ▶ Datos varianzas-covarianzas muestrales: $q(q+1)/2$.
- ▶ Parámetros a estimar:
 - Varianzas de los factores comunes.
 - Coefficientes de regresión entre valores observados y factores comunes.
 - Varianzas de los errores (figura 10).

FIGURA 10. Identificación.



Fuente: Jöreskog (1993); Hair (et al., 1999); Byrne (2006), con adaptación propia.

CFA. Estimación

Se puede realizar mediante la consideración de los métodos discutidos en el apartado Method (ME):

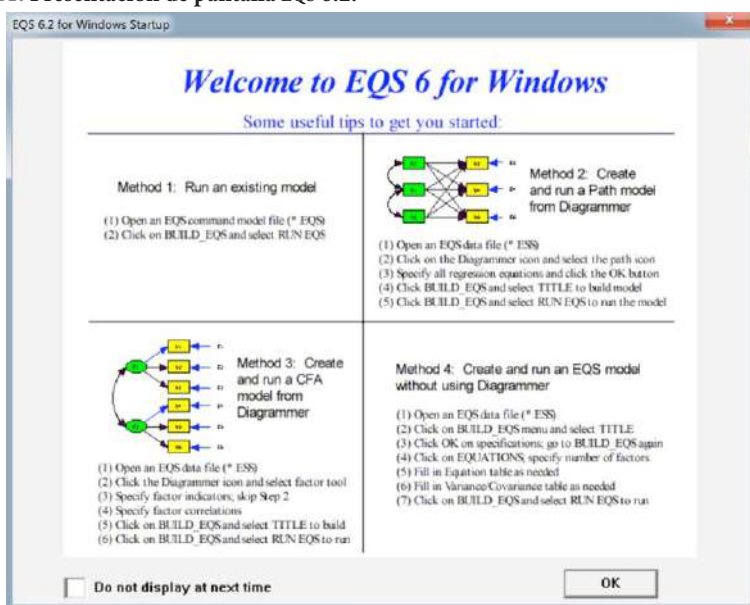
- ▶ Mínimos cuadrados no ponderados.
- ▶ Mínimos cuadrados generalizados.
- ▶ Teoría de la distribución asintótica
- ▶ Con libre distribución asintótica.
- ▶ Máxima verosimilitud (el más utilizado).

De este método, de acuerdo con Ullman (1996), se recomienda:

- ▶ ML y GLS siempre que sea lograble la normalidad e independencia.
- ▶ Si no se cumple la normalidad, ML y GLS sólo deberá realizarse para tamaños muestrales muy grandes (mayores a 2500).
- ▶ Si no se cumple la normalidad USL, aunque no existan contrastes estadísticos y los estimadores dependen de la escala de medida de las variables observadas.

Nota: Bentler en su versión de EQS 6.2 estandariza los datos para conservar la normalidad (figura 11).

FIGURA 11. Presentación de pantalla EQS 6.2.



Fuente: EQS 6.2.

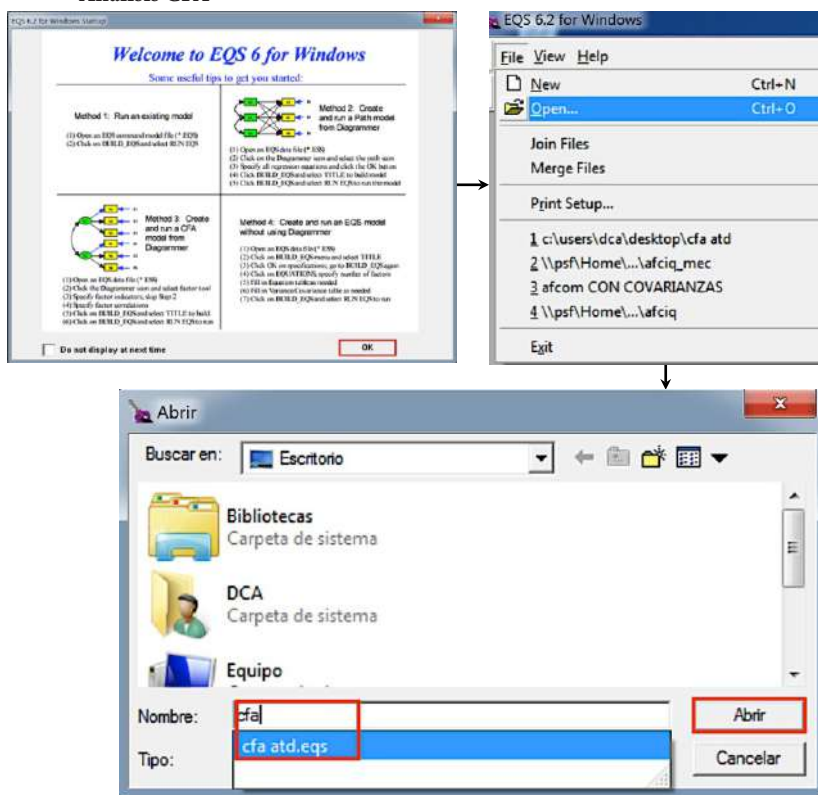
CFA y EQS. Entrada de datos al programa

A fin de entender el CFA aplicando EQS, realizaremos un análisis al programa nombrado como CFA ATD.eqs, y se documenta en el título: *CFA ATD. CFA en la asertividad en la toma de decisiones mediante la siguiente serie de comandos:*

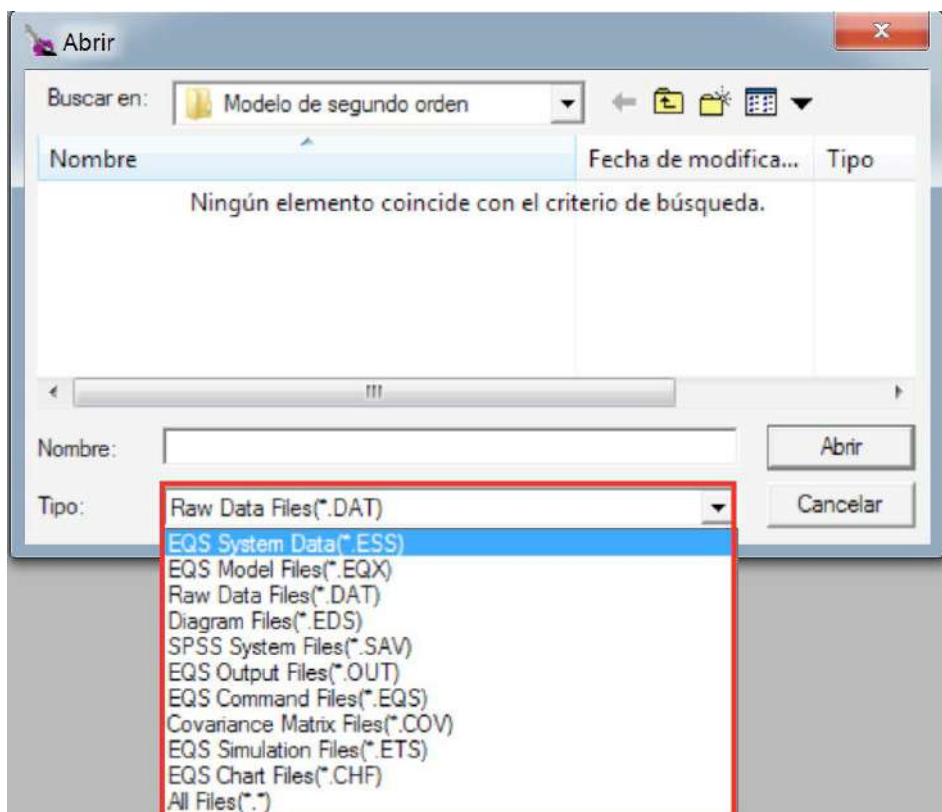
Icono EQS 6.2 for windows->OK->File->Open->Nombre: cfa atd.eqs->Tipo: EQS System Data (*.ESS)->Abrir.

Observe que el programa EQS, ofrece una serie de formatos de archivos con los cuales opera, siendo los más utilizados los EQS Command Files (*.EQS), con los que se configura el archivo de inicio de programa; EQS Output Files (*.OUT), que genera el archivo de datos de salida para análisis; SPSS System files (*.SAV), que permite visualizar el contenido de las bases de datos de SPSS; EQS System Data (*.ESS), utilizado para ingresar datos externos, entre otros.

FIGURA 12. EQS. Secuencia de comandos para apertura de programa de entrada de datos. Análisis CFA



Fuente: EQS 6.2. con datos propios.



Fuente: EQS 6.2, con datos propios..

En el que se tiene codificada la matriz de covarianzas de datos del estudio de caso, donde se aprecian las instrucciones a nivel de título del proyecto, las especificaciones de la base de datos, el método de análisis, la matriz de covarianzas y las desviaciones estándar consideradas, entre otros datos relevantes. Observe que si tiene seis factores observables, en 275 casos, analizados con matriz de correlación, método de máxima verosimilitud (ML. *máximum likelihood*) y análisis por covarianza. La cantidad de elementos en la matriz aplicando la fórmula 1 es de 15. Este dato está reportado en las magnitudes de la matriz de covarianzas de datos (figura 13).

De acuerdo con la codificación mostrada en la figura 12, se tiene de manera relevante las siguientes condiciones de aplicabilidad cumplidas:

- ▶ Observe el párrafo /TITLE, donde se dan detalles del nombre del archivo y lo que hace. En nuestro caso: CFA ATD. Asertividad en la toma de decisiones.

FIGURA 13. EQS. Codificación de entrada de datos. Caso de estudio 1

```

EQS 6.2 for Windows - [cfa atd]
File Edit View Insert Format Build_EQS Window Help
. . . . . 1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . .
/TITLE
CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES
/SPECIFICATIONS
VARIABLES= 6; CASES= 275;
MATRIX=COR; METHOD=ML; ANAL=COV;
/MATRIX
1.000
0.493 1.000
0.401 0.314 1.000
0.278 0.347 0.147 1.000
0.317 0.318 0.183 0.587 1.000
0.284 0.327 0.179 0.463 0.453 1.000

/STANDARD DEVIATIONS
1.0900 0.5900 0.9800 1.1000 0.4100 1.1100

/LABELS
V1=ART; V2=FSF; V3=IMG; V4=MAT; V5=FIS; V6=LOG;

/EQUATIONS
V1=F1+E1;
V2=*F1+E2;
V3=*F1+E3;
V4=F2+E4;
V5=*F2+E5;
V6=*F2+E6;

/VARIANCES
E1=*;
E2=*;
E3=*;
E4=*;
E5=*;
E6=*;
F1 TO F2=*;

/COVARIANCES
F1 TO F2=*;

/PRINT
FIT=ALL;

/LMTEST

/WTEST

/END

```

Fuente: EQS 6.2, con datos propios..

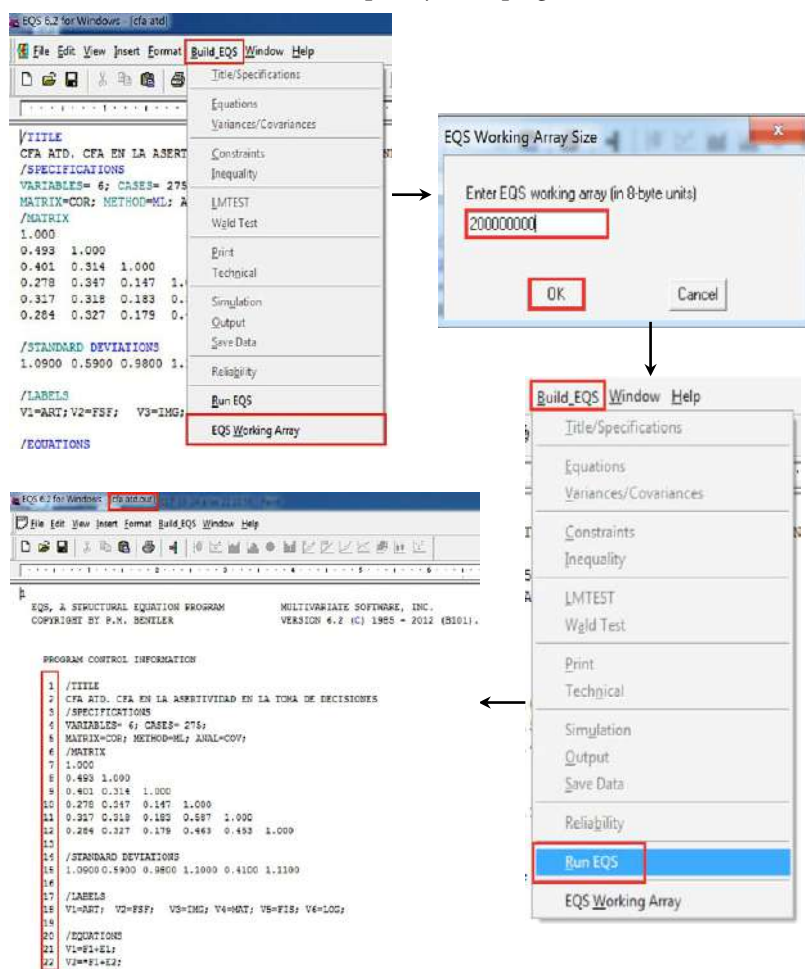
- ▶ Observe el párrafo /SPECIFICATIONS, donde se muestran las instrucciones que declaran las seis variables observables bajo estudio; los 275 casos de los datos de ejecutivos a analizar; con una matriz de correlación (COR), bajo el método de máxima verosimilitud (ML); con análisis por covarianza (COV). Se muestra la matriz de los 15 casos de enlaces que arrojan las seis variables (fórmula 1), integrada en el programa de entrada de datos EQS.
- ▶ Observe el párrafo /STANDARD DEVIATIONS con las desviaciones estándar propuestas.
- ▶ Observe el párrafo /LABELS con las etiquetas de las seis variables que describen al modelo y su equivalente en EQS (variables VX, que describen el modelo teórico).
- ▶ Observe el párrafo /EQUATIONS, en la parte de ecuaciones se eligió $V1=F1+E1$, SIN ASTERISCO (*), con el fin de afirmarlo con varianza=1 y cumplir los requerimientos (así también, se replicó en el caso de V4); V2 ya tiene * para que se calcule la relación de la carga factorial de $V2=*F1+E2$ hacia F1 (es decir, que tanto explica V2 a F1); como los errores no tienen asterisco, se afirman por tanto con varianza= 1; idem con V3, y así consecutivamente.
- ▶ Observe el párrafo /VARIANCES, éstas cumplen la condición en errores y en factores. Es factible la instrucción E1 to E6=*, pero no se sugiere en errores ya que limita el análisis por eliminación y, por lo tanto, el ajuste del modelo. En factores no hay problema ya que éstos NO se deben eliminar (es por ello que se justifica F1 to FX).
- ▶ Observe el párrafo /COVARIANCES, donde se afirma que sólo las variables independientes covarían (F1 to F2=* o F1=*; F2=*; es lo mismo).
- ▶ Observe el párrafo /PRINT FIT ALL, donde se indica que se requiere imprimir todos los índices de ajuste para su análisis.
- ▶ Observe el párrafo /LMTEST, que significa se genere el test de Lagrange, el cual es, univariado o multivariado. El primero indica que tanto aporta cada una de las variables a su correspondiente factor. El segundo, indica qué variable tiene que medir en qué factor para mejorar el ajuste. Es decir, qué cambios se deben realizar en las variables para que midan factores distintos para que se mejore el modelo. No significa finalmente, que la V1 sea parte o mida a F2, sino que su contribución disminuye el error y mejora el ajuste del modelo.
- ▶ Observe el párrafo /WTEST, o TEST DE WALD, que implica determinar qué relación entre factores mejora el ajuste modelo.
- ▶ Observe el párrafo /END, término de la programación.

CFA y EQS. Cómo ejecutar el programa

Una vez definidas las ecuaciones del modelo teórico, y cumplidas las condiciones de aplicabilidad, se corre el programa mediante la siguiente secuencia de comandos:

Build EQS->EQS Working Array->Enter EQS working array (in 8-bits units): cambiar de 2000000 a 200000000 (se asegura capacidad de memoria)->OK->Build EQS->Run->Generación de archivo de salida cfa atd.out (figura 14).

FIGURA 14. EQS. Secuencia de comandos para ejecutar programa de entrada de datos.



Fuente: EQS 6.2, con datos propios.

CFA y EQS. Análisis de resultados

Como se observa, el archivo de salida generado cfa atd.out, muestra el mismo contenido de instrucciones que el archivo de entrada cfa atd.eqs, sólo que con las líneas de comando numeradas y adicionalmente, con los reportes de resultados como producto del cálculo EQS. Por ejemplo, las secciones de reporte de las figuras 15 y 16, que muestran los arreglos de datos que cumplen con la distribución esperada.

FIGURA 15. EQS. Reporte de matriz de covarianza analizada con el CFA.

```

DETERMINANT OF INPUT MATRIX IS  0.23735D-01

PARAMETER ESTIMATES APPEAR IN ORDER,
NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

RESIDUAL COVARIANCE MATRIX  (S-SIGMA)  :

```

	ART V1	FSF V2	IMG V3	MAT V4	FIS V5	
ART V1	0.000					
FSF V2	-0.007	0.000				
IMG V3	0.044	-0.014	0.000			
MAT V4	-0.055	0.028	-0.076	0.000		
FIS V5	-0.003	0.003	-0.014	0.004	0.000	
LOG V6	0.027	0.053	0.003	-0.004	-0.006	
LOG V6						0.000

```

AVERAGE ABSOLUTE RESIDUAL = 0.0163
AVERAGE OFF-DIAGONAL ABSOLUTE RESIDUAL = 0.0228

```

Fuente: EQS 6.2, con datos propios..

En la figura 17 se aprecia que los datos están adecuadamente distribuidos, ya que no existen desviaciones estándar ni a la izquierda ni a la derecha del intervalo.

Si existiera el caso, es indicativo de que algunas variables no están aportando al ajuste del modelo. Se debe tener mucho cuidado a la eliminación de variables para justificar el modelo de estudio (eliminar 10% al 20% de variables).

FIGURA 16. EQS. Reporte residuos estandarizados con el CFA.

24-OCT-17 PAGE: 3 EQS Licensee:
 TITLE: CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

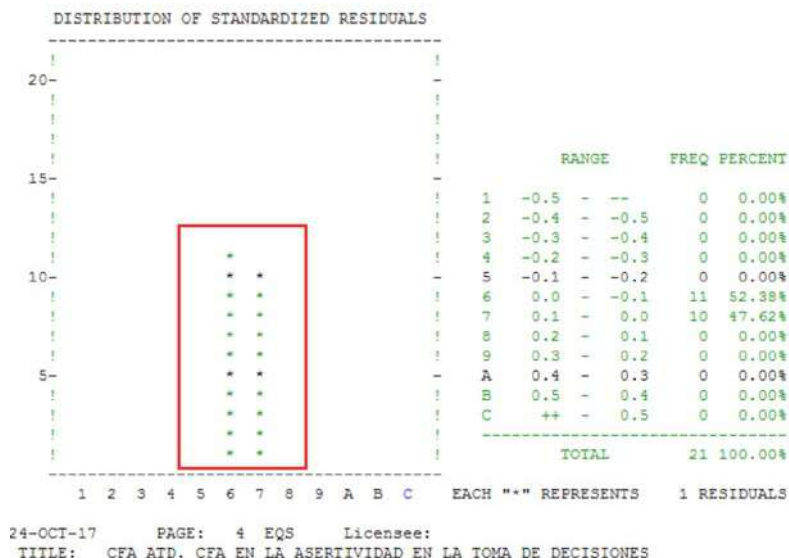
MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

LARGEST STANDARDIZED RESIDUALS:

NO.	PARAMETER	ESTIMATE	NO.	PARAMETER	ESTIMATE
1	V6, V2	0.081	11	V2, V1	-0.011
2	V4, V3	-0.070	12	V5, V4	0.010
3	V4, V1	-0.046	13	V5, V1	-0.007
4	V4, V2	0.043	14	V6, V4	-0.004
5	V3, V1	0.041	15	V6, V3	0.003
6	V5, V3	-0.035	16	V2, V2	0.000
7	V3, V2	-0.024	17	V3, V3	0.000
8	V6, V1	0.022	18	V6, V6	0.000
9	V6, V5	-0.014	19	V5, V5	0.000
10	V5, V2	0.013	20	V4, V4	0.000

Fuente: EQS 6.2, con datos propios..

FIGURA 17. EQS. Gráfico distribución estandarizada de los residuos con el CFA.



Fuente: EQS 6.2, con datos propios..

Por otro lado, es importante analizar la sección de reporte de índices de ajuste (*fit indices*), los cuales en reporte muestran entre otros resultados relevantes (figura 18), donde lo encuadrado en rojo, es la referencia a analizar):

- ▶ El metodo de máxima verosimilitud ML.
- ▶ Se observa que $p = .35579 > p = .05$, o sea, aprox. 36% de error, es decir, la probabilidad de que el modelo se equivoque, es muy alta; esto es importante para ya no considerarlo.
- ▶ Los índices de ajuste (*fit indices*) más utilizados son los primeros tres:
- ▶ Índice de ajuste normalizado o Bentler-Bonnet Normed Fit Index, NFI.
- ▶ Índice de ajuste no normalizado o Bentler-Bonnet Non Normed Fit Index, NNFI.
- ▶ Índice de ajuste comaprativo o Comparative Fit Index, CFI.
- ▶ Error medio cuadrático de aproximación del error o Root Mean-Square Error of Approximation, RMSEA.

Si el promedio de los tres primeros se encuentre entre .8-.89 el modelo se encuentra ajustado estadístico (Hair et al., 1999; Bagozzi y Yi, 1988); si es $>.9$, tiene ajuste excelente; $RMSEA = <.08$.

En conclusión, al momento el modelo está ajustado pero con un alto error.

FIGURA 18. EQS. Reporte índices de ajuste con el CFA.

```

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE          =      392.818 ON      15 DEGREES OF FREEDOM

INDEPENDENCE AIC =      362.818   INDEPENDENCE CAIC =      293.566
MODEL AIC =      -7.158           MODEL CAIC =      -44.092

CHI-SQUARE =      8.842 BASED ON      8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      0.35579

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS      9.156.

FIT INDICES
-----
BENTLER-BONETT NORMED FIT INDEX =      0.977
BENTLER-BONETT NON-NORMED FIT INDEX =      0.996
COMPARATIVE FIT INDEX (CFI) =      0.998
BOLLEN'S (IFI) FIT INDEX =      0.998
MCDONALD'S (MFI) FIT INDEX =      0.998
JORESOG-SORBOM'S GFI FIT INDEX =      0.989
JORESOG-SORBOM'S AGFI FIT INDEX =      0.971
ROOT MEAN-SQUARE RESIDUAL (RMR) =      0.027
STANDARDIZED RMR =      0.031
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =      0.020
90% CONFIDENCE INTERVAL OF RMSEA (      0.000,      0.075)

```

Fuente: EQS 6.2, con datos propios..

Cabe preguntarse, ¿en qué medida la matriz de varianzas-covarianzas predicha se ajusta a la muestra? Esto se puede resolver si se aplican los siguientes criterios:

Matriz residual de covarianzas. Todos los residuos deberían estar cercanos a 0 y simétricamente distribuidos

Chi-cuadrado no significativa. Contrasta la hipótesis nula de que la matriz de varianzas-covarianzas estimada no es cualquier matriz positiva definida, sino que está condicionada al modelo.

No se debe rechazar esta hipótesis.

La prueba es muy sensible al tamaño de la muestra (tendencia a rechazo en muestras muy grandes).

Considerar los estadísticos: NFI; NNFI; CFI; IFI; MFI; GFI; AGFI cercanos c/u a .9 (Nunca es ≥ 1 . Esto representaría multicolinealidad y no es posible que midan lo mismo), y Error de la media cuadrática residual Standarizada (SRMR. *standardized root-mean square residual*) cercano a .05 (SRMR, se utiliza cuando en primera instancia el RMSEA=0 que implicaría alta multicolinealidad entre las variables del modelo y tres índices de ajuste al menos sean $>.9$. Esto genera la necesidad de aclarar el ajuste del modelo).

Que el proceso de estimación haga convergencia.

Continuando el análisis, se prosigue a revisar la sección de reporte del alfa de Cronbach (*reliability coefficients*) que reporta la confiabilidad del modelo. Se destaca que lo importante a revisar es la confiabilidad y validez de la escala NO del modelo (figura 19).

FIGURA 19. EQS. Reporte confiabilidad alfa de Cronbach del modelo con el CFA.

RELIABILITY COEFFICIENTS	
CRONBACH'S ALPHA	= 0.712
RELIABILITY COEFFICIENT RHO	= 0.777

Fuente: EQS 6.2, con datos propios..

Otro reporte a observar es la sección de iteraciones (*iterative summary*) que registra la cantidad de iteraciones que realiza el proceso hasta que hace convergencia en cero, en este caso se señalan cinco iteraciones (figura 20).

Continuando el análisis de resultados que reporta EQS, se observa la sección de reporte de medición de las ecuaciones (*measurement equations with standard errors and tests statistics significant at the 5% are marked with @*). Aquí se reportan el valor de las ecuaciones NO estandarizadas, con la varianza de cada una de las variables NO estandarizadas. Así, donde hay @ la p es significativa (figura 21).

FIGURA 20. EQS. Iteraciones del sistema con el CFA.

ITERATIVE SUMMARY			
ITERATION	PARAMETER ABS CHANGE	ALPHA	FUNCTION
1	0.147746	1.00000	0.04518
2	0.025186	1.00000	0.03297
3	0.006181	1.00000	0.03233
4	0.001730	1.00000	0.03228
5	0.000515	1.00000	0.03227

Fuente: EQS 6.2, con datos propios..

FIGURA 21. EQS. Reporte medición de las ecuaciones con el CFA.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

VARIANZA 1 DADO QUE NO SE LES ASIGNO ASTERISCO (*) COMO DISEÑO ORIGINAL

ART	=V1	=	1.000	F1	+	1.000	E1
-----	-----	---	-------	----	---	-------	----

VALOR DE LA VARIANZA DE V2 RESPECTO A F1 TODAS LAS VARIANZAS DE LOS ERRORES TIENEN VALOR 1

FSF	=V2	=	.509*	F1	+	1.000	E2
			.068				
ERROR DE LA VARIANZA DE V2							
VALOR t DE STUDENT ES SIGNIFICATIVA AL .05			7.467@				
IMG	=V3	=	.604*	F1	+	1.000	E3
			.096				
			6.320@				
MAT	=V4	=	1.000	F2	+	1.000	E4
FIS	=V5	=	.373*	F2	+	1.000	E5
			.039				
			9.467@				
LOG	=V6	=	.817*	F2	+	1.000	E6
			.096				
			8.552@				

Fuente: EQS 6.2, con datos propios..

De este reporte se debe analizar que todas las variables que presentan varianzas a sus respectivos factores presenten valores t de Student con una significancia (@) al .05 para continuar el análisis, situación que es el caso.

Interpretando el modelo, se requiere preguntarse en este punto, sobre los parámetros estimados:

- ¿Toman valores adecuados teóricamente?
- ¿Existen correlaciones superiores a la unidad?
- ¿Existen cargas factoriales estandarizadas fuera del intervalo (-1,+1)?
- ¿Son los residuos estandarizados normalmente grandes o pequeños?
- ¿Hay estimaciones negativas de las varianzas?
- ¿Son significativos?

$$FSF = V2 = .509 * F1 + 1.000 E2$$

$$.068$$

$$7.4676$$

donde:

Valor t (coeficiente/error estándar)

t > 1.96 valor significativo a .05 o 5%

t > 2.56 valor significativo a .01 o 1%

Así, la asignación de asteriscos (*) de error/confiabilidad/ valor de t, es:

*** p < .01 = 1% error o 99% confiabilidad, cuando valor de t > 2.56

** p < .05 = 5% error o 95% confiabilidad, cuando valor de t = 1.96 <= 2.55

* p < .10 = 10% error o 90% confiabilidad, cuando valor de t = 1.65 <= 1.95

Revisando los reportes emitidos por EQS, encontramos la sección de reporte de varianzas de variables independientes (*variances of independent variables*) entre factores y, entre las variables y sus errores.

¿Varianza negativa? (figuras 22 y 23).

Ambos reportes validan lo analizado hasta el momento.

La sección de reporte muy importante de analizar sus resultados, es la de la covarianza entre las variables independientes (*covariances among independent variables*), (figura 24).

Es muy importante considerar el análisis de este reporte, ya que a partir del mismo se determina la validez del modelo. Se recuerda que:

- ▶ El valor de la covarianza entre factores nunca debe ser 1, ya que entonces los factores miden lo mismo;
- ▶ El valor de la covarianza NO debe ser > .8 ya que puede tener problemas debido a que la covarianza es sumada con 2 veces el error, lo que tiende a ser 1 y, por lo

FIGURA 22. EQS. Reporte de varianzas entre factores con el CFA.

```

24-OCT-17      PAGE:   6  EQS      Licensee:
TITLE:   CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

VARIANCES OF INDEPENDENT VARIABLES
-----
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

          V                      F
          ---                      ---
          I  F1  -   F1                      .636*I
          I                      .117 I
          I                      5.444@I
          I                      I
          I  F2  -   F2                      .698*I
          I                      .112 I
          I                      6.244@I
          I                      I
    
```

Fuente: EQS 6.2, con datos propios..

FIGURA 23. EQS. Reporte de varianzas entre las variables y sus errores con el CFA.

```

24-OCT-17      PAGE:   7  EQS      Licensee:
TITLE:   CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

VARIANCES OF INDEPENDENT VARIABLES
-----
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

          E                      D
          ---                      ---
E1 - ART          .552*I                      I
          .088 I                      I
          6.255@I                      I
          I                      I
E2 - FSF          .183*I                      I
          .025 I                      I
          7.296@I                      I
          I                      I
E3 - IMG          .728*I                      I
          .071 I                      I
          10.281@I                      I
          I                      I
E4 - MAT          .512*I                      I
          .075 I                      I
          6.828@I                      I
          I                      I
E5 - FIS          .071*I                      I
          .010 I                      I
          6.807@I                      I
          I                      I
E6 - LOG          .767*I                      I
          .079 I                      I
          9.655@I                      I
          I                      I
    
```

Fuente: EQS 6.2, con datos propios..

FIGURA 24. EQS. Reporte covarianzas entre las variables independientes con el CFA.

```

24-OCT-17      PAGE:   8  EQS      Licensee:
TITLE:   CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

COVARIANCES AMONG INDEPENDENT VARIABLES|
-----|
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

      V
      ---
      I F2 - F2
      I F1 - F1
      I
      I

SE LEE: LA COVARIANZA ENTRE F1 Y F2 ES DE .388 CON UN
ERRO DE .068 Y UN VALOR t DE STUDENT SIGNIFICATIVO
AL .05
      .388*I
      .068 I
      5.712@I
      I
    
```

Fuente: EQS 6.2, con datos propios..

tanto, generar problemas de validez (*colinealidad*). Ver el apartado Cálculo de la validez discriminante. Test del intervalo de confianza.

- ▶ El valor de la covarianza NO debe ser $<.1$ ya que los factores serían interdependientes (lo que significa que nada habría que ver entre ellas; sin relación alguna en el modelo).

En nuestro caso (.388), NO existe el 1 y NO es $>.8$ cumple el requerimiento hasta el momento para verificar su validez.

Al seguir analizando los reportes emitidos por EQS, encontramos el apartado Solución estandarizada (*standardized solution*), es decir, son resultados que ya han eliminado el error a cada una de las variables que intervienen en el modelo. Lo visto hasta el momento no se encontraba estandarizado.

¿Estimaciones estandarizadas >1 como valor absoluto? (figura 25).

FIGURA 25. EQS. Reporte solución estandarizada con el CFA.

```

24-OCT-17      PAGE:   9  EQS      Licensee:
TITLE:   CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEO

STANDARDIZED SOLUTION:
CARGAS FACTORIALES ESTANDARIZADOS DE LAS 6 VARIABLES RESPECTO A
LOS FACTORES Y LOS ERRORES CON R CUADRADA DE C/U
R-SQUARED

ART  =V1  =  .732 F1  +  .682 E1  .535
FSF  =V2  =  .688*F1  +  .725 E2  .474
IMG  =V3  =  .492*F1  +  .871 E3  .242
MAT  =V4  =  .759 F2  +  .651 E4  .577
FIS  =V5  =  .760*F2  +  .650 E5  .578
LOG  =V6  =  .615*F2  +  .789 E6  .378
    
```

Fuente: EQS 6.2, con datos propios..

Antes de finalizar el método, se reporta la apartado de correlaciones entre variables independientes (correlations among independent variables).

¿Correlaciones >1? No puede ser, porque sería el mismo factor. Deseable cargas factoriales de factores >.6 (figura 26).

FIGURA 26. EQS. Reporte correlaciones entre variables independientes con el CFA.

```

24-OCT-17      PAGE: 10  EQS      Licensee:
TITLE:  CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

CORRELATIONS AMONG INDEPENDENT VARIABLES
-----
          V          F          VERIFICAR QUE EL VALOR SEA <1
          ---          ---          -----
          I F2 - F2          .582*I
          I F1 - F1          I
          I                I

-----
          END OF METHOD
-----
    
```

Fuente: EQS 6.2, con datos propios..

FIGURA 27. EQS. Reporte test de Wald con el CFA.

```

24-OCT-17      PAGE: 11  EQS      Licensee:
TITLE:  CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

WALD TEST (FOR DROPPING PARAMETERS)
MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS

          CUMULATIVE MULTIVARIATE STATISTICS          UNIVARIATE INCREMENT
          -----          -----
STEP  PARAMETER  CHI-SQUARE  D.F.  PROBABILITY          CHI-SQUARE  PROBABILITY
-----  -----  -----  -----          -----  -----

*****
      NONE OF THE FREE PARAMETERS IS DROPPED IN THIS PROCESS.
    
```

Fuente: EQS 6.2, con datos propios..

Hasta aquí, fin de los reportes del método, sin embargo, continuamos con los análisis de los resultados de pruebas específicas como lo es la prueba de Wald (*Wald test-for dropping parameters*), considerando parámetros multivariantes (*multivariate Wald test by simultaneous process* (figura 27).

El resultado: *none of the free parameters is dropped in thios process* de la prueba nos indica que no hay ningún parámetro que se pueda considerar en esta relación, dado que sólo hay dos factores que generan covarianza y se requiere para su demostración al menos tres factores. Aún así, es correcto medir con dos factores la asertividad en la toma de decisiones (ATD).

Verificando los reportes EQS, nos encontramos el apartado multiplicador de Lagrange (*Lagrange multiplier test (for adding parameters)*), para datos univariados y multivariados (figuras 28 y 29).

FIGURA 28. EQS. Reporte multiplicador de Lagrange. Datos univariados con el CFA.

```

24-OCT-17      PAGE: 12 EQS      Licensee:
TITLE:  CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
SE DESCARTA ESTA SECCIÓN PARA ANÁLISIS EN LA
MODALIDAD UNIVARIADA. LO IMPORTANTE ES LA
MODALIDAD MULTIVARIADA
ORDERED UNIVARIATE TEST STATISTICS:

```

NO	CODE	PARAMETER	CHI-SQUARE	PROB.	HANCOCK 8 DF PROB.	PARAM. CHANGE	STANDAR- DIZED CHANGE	PREDICTED RMSEA	CFI
1	2 12	V2,F2	4.741	0.029	0.785	0.174	0.354	99.999	1.000
2	2 12	V6,F1	2.098	0.148	0.978	0.181	0.205	99.999	1.000
3	2 12	V3,F2	1.675	0.196	0.989	-0.142	-0.174	0.009	1.000
4	2 12	V1,F2	1.125	0.289	0.997	-0.167	-0.184	0.019	0.998
5	2 12	V4,F1	1.000	0.317	0.998	-0.134	-0.152	0.021	0.998
6	2 12	V5,F1	0.069	0.793	1.000	-0.013	-0.040	0.030	0.995
7	2 0	V4,F2	0.000	1.000	1.000	0.000	0.000	0.031	0.995
8	2 0	V1,F1	0.000	1.000	1.000	0.000	0.000	0.031	0.995

*** NOTE *** IF PREDICTED RMSEA COULD NOT BE CALCULATED, 99.999 IS PRINTED.
IF PREDICTED CFI COULD NOT BE CALCULATED, 9.999 IS PRINTED.

Fuente: EQS 6.2, con datos propios..

Se descarta análisis ya que lo importante son los multivariados (figura 29).

La lectura de que V2 mide a F2 con un incremento de Chi-cuadrado, trae como consecuencia un cambio de incidencia variable-factor, como se indica en la figura 30.

FIGURA 29. EQS. Reporte multiplicador de Lagrange. Datos multivariados con el CFA

24-OCT-17 PAGE: 13 EQS Licensee:
 TITLE: CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES
 MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1

PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:
 PVV PFV PFF PDD GVV GVF GFV GFF BVF BFF

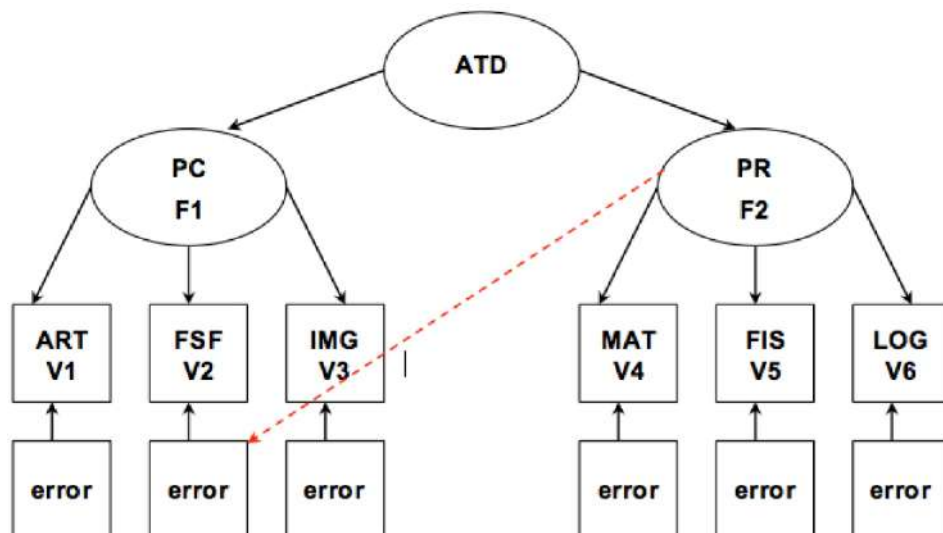
CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT					
STEP	PARAMETER	CHI-SQUARE	D.F.	PROB.	CHI-SQUARE	PROB.	HANCOCK'S SEQUENTIAL D.F. PROB.	PREDICTED RMSEA	CFI
1	V2, F2	4.741	1	0.029	4.741	0.029	8 0.785	99.999	1.000

SE LEE: SI V2 MIDE A F2 ENTONCES CHI CUADRADO SE INCREMENTA EN 4.741

*** NOTE *** IF PREDICTED RMSEA COULD NOT BE CALCULATED, 99.999 IS PRINTED.
 IF PREDICTED CFI COULD NOT BE CALCULATED, 9.999 IS PRINTED.

Fuente: EQS 6.2, con datos propios..

FIGURA 30. Diagrama de trayectoria con cambio de contribución de variable-factor.



Fuente: Elaboración propia.

FIGURA 31. EQS. Reporte solución estandarizada con el CFA

```

24-OCT-17      PAGE:   9  EQS      Licensee:
TITLE:   CFA ATD. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

STANDARDIZED SOLUTION:                                     R-SQUARED

ART  =V1  =   .732 F1   + .682 E1                          .535
FSF  =V2  =   .688*F1  + .725 E2                          .474
IMG  =V3  =   .492*F1  + .871 E3                          .242
MAT  =V4  =   .759 F2   + .651 E4                          .577
FIS  =V5  =   .760*F2  + .650 E5                          .578
LOG  =V6  =   .615*F2  + .789 E6                          .378

```

Fuente: EQS 6.2, con datos propios..

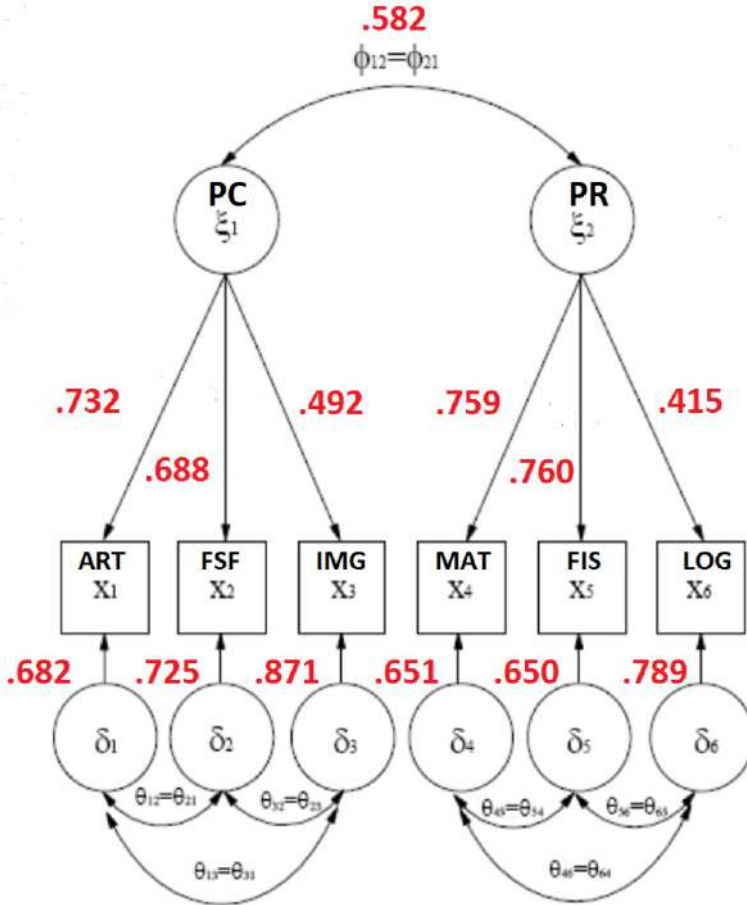
CFA. Discusión

La discusión debe dar una clara exposición de los constructos del modelo y sus implicaciones en función del conocimiento, línea de investigación, incluso sesgo e interés del investigador, de la teoría del modelo vs. la realidad y explicar coincidencias y diferencias con pleno conocimiento del sector abordado. Las posibles causas de resultados diferentes se deben explicar a fin de que el modelo se depure y se haga universal. Hacer observaciones para estudios futuros.

El resultado de nuestro caso de estudio, nos sugiere que para mejorar el modelo, el pensamiento racional (PR), es la causa de la filosofía (FSF) y no del pensamiento creativo (PC). Es decir, V2 al medir a F2 (PR), con el aumento de Chi-cuadrado en vez de F1, implica que el modelo sólo a nivel estadístico requiere este reajuste, sin embargo, NO tiene implicaciones teóricas ya que el modelo fue previamente validado a nivel de constructo en un marco teórico, por lo que es muy probable que la nueva relación NO tenga sustento como ajuste teórico. Puede hacerse la pregunta ¿Puede la filosofía (FSF) medir la asertividad de la toma de decisiones (ATD) en la parte de su pensamiento racional (PR)? Esto exigiría nuevamente una revisión teórica del modelo. Se debe recordar también, que a estas alturas, NO es el objetivo mejorar el modelo ya que incluso, el modelo cumple con las condiciones de los índices de ajuste (*fit indices*, >.9), mostrados en la figura 18.

Asumiendo que el modelo está correctamente respaldado por el marco teórico y con el fin de ser más rigurosos en el análisis, se recurre a una de las condiciones de aplicabilidad emitidas por Bagozzi y Yi (1988), las cuales, para ser tomadas en cuenta, deben ser cada una >.6. Analizando la solución estandarizada de la figura 25, observamos, que esta condición no cumple para la variable (figura 31).

FIGURA 32. Diagrama de trayectoria con los resultados del caso de estudio con solución estandarizada.



Fuente: Jöreskog (1993); Hair et al., 1999); Byrne (2006), con adaptación propia.

CFA y EQS. Re-especificación del modelo

Como se observa, lo logrado hasta el momento implica añadir o eliminar relaciones entre las variables del modelo motivado por dos razones:

- ▶ Mejorar el ajuste.
- ▶ Contrastar alguna hipótesis teórica.

Sin embargo, se recomienda incorporar o eliminar relaciones cuando éstas tengan un soporte teórico que justifique la escala. La mejora del ajuste NO debe ser

el fin en sí mismo.

Lo anterior, es posible realizarlo por medio de:

- ▶ El multiplicador de Lagrange (Lagrange multiplier) para añadir relaciones, parámetro que conduce a un mayor incremento de la Chi-cuadrado.
- ▶ La prueba de Wald (Wald test).

```
/LMTEST  
/WTEST  
  
/END
```

En nuestro caso no aplicó. Importante referirse a autores acreditados para basar su decisión como lo mostrado en la discusión (figuras 27 y 29).

CFA. Confiabilidad

En esta etapa, se debe verificar:

Consistencia. Coherencia interna de los ítems (¿existen correlaciones entre ellos?). ¿Realmente ítems, variables y factores del modelo miden lo planeado?

Si los *items están fuertemente correlacionados entre sí*, significa que miden una misma variable latente (los causa), pero no que esta variable latente que miden, sea el constructo que se pretende medir (validez).

La confiabilidad como condición necesaria pero NO suficiente para la validez. Así, “una escala puede tener confiabilidad pero NO validez, sin embargo, JAMÁS una escala tendrá validez SINO tiene antes, confiabilidad”.

Su medición, suele realizarse por una de tres vías de explicación de la varianza (Bagozzi y Yi, 1988):

Coefficiente alfa de Cronbach > .7

Índice de fiabilidad compuesta (IFC) > .7

Índice de varianza extraída (IVE) > .5

Desde el punto de vista de análisis estructural, se sugiere cumplir con los tres, para lograr confiabilidad de la escala (NO de los datos).

Punto de partida para construir los indicadores de confiabilidad: la matriz de varianzas-covarianzas entre los ítems de la escala.

Recordar que el alfa de Cronbach, es el método más utilizado para medir la confiabilidad cuando ésta se entiende como consistencia interna, analiza la homogeneidad de las cuestiones planteadas de forma simultánea y tiene la ventaja de que para su cálculo, sólo se necesita realizar una medición con un grupo de personas. El proceso consiste en medir la confiabilidad de una suma simple de

medias (ítems) paralelas o *tau-equivalentes* y es la medida de todos los coeficientes de correlación resultantes de correlacionar dos mitades de una escala de todas las formas posibles. Vea nuevamente la ecuación 9 del capítulo 3.

ECUACIÓN. Alfa de Cronbach por correlación

$$\alpha = \frac{k\rho}{1+(k-1)\rho}$$

donde:

ρ es, como ya hemos indicado, la media de los coeficientes de correlación entre todos los ítems que conforman la escala.

El coeficiente alfa de Cronbach debe variar entre 0 y 1; así, conforme se incrementa su valor, menor es la varianza. Así, se tienen dos interpretaciones del valor de alfa de Cronbach:

- ▶ “Es la correlación existente entre la escala y todas las posibles escalas que, midiendo el mismo concepto, tengan igual número de ítems”, o
- ▶ “Es la correlación entre la puntuación que una persona (o personas) obtiene en una escala (valor observado) y la puntuación que debería tener, considerando todos los ítems del universo que definen el concepto”.

El punto de inicio: matriz de varianzas-covarianzas. Si se estandariza, se obtiene la matriz de correlaciones. De esta manera, se afirma que la varianza total de la escala es la suma de los elementos de C (matriz 3.1).

MATRIZ 3.1. Varianzas-covarianzas.

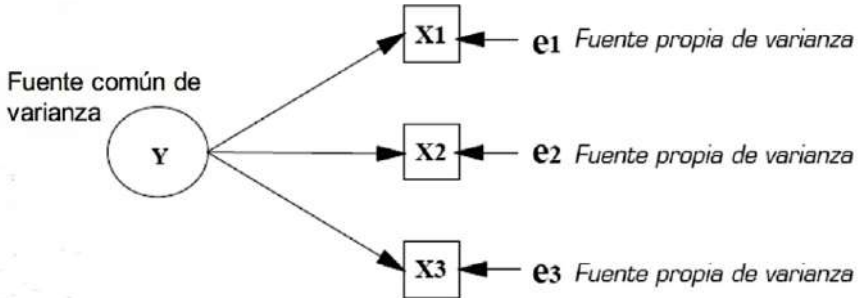
$$C = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Siendo así: Varianza total= Varianza común+ Varianza específica.

Varianza común. Se trata de la varianza en los ítems provocada por la variable latente (que los ítems comparten entre sí). Si varía la variable latente, varían los ítems X_i de la escala.

Varianza específica, única. Varianza causada por los errores de medida de cada ítem, en particular en la diagonal principal de la matriz. Observe la figura 33.

FIGURA 33. Comportamiento de la varianza.



Fuente: Jöreskog (1993); Hair *et al.*, (1999); Byrne (2006), con adaptación propia.

La variable independiente (fuente común de la varianza) y los errores (fuente propia de varianza de cada variable dependiente) afecta a las dependientes.

Alfa de Cronbach= la parte de la varianza total que es atribuible a la variable latente (varianza común), donde:

$$\alpha = \frac{\sigma_y^2}{\sum \sigma_i^2} = 1 - \frac{\sum \sigma_i^2}{\sigma_y^2}$$

VARIANZA TOTAL.
 Suma de los elementos de la Matriz de Varianzas-Covarianzas

VARIANZA ESPECÍFICA.
 Suma de los elementos de la Diagonal de la Matriz de Varianzas-Covarianzas

Fuente: Jöreskog (1993); Hair *et al.*, (1999); Byrne (2006), con adaptación propia.

La expresión anterior, debe ponderarse debido a que el número de elementos necesarios para calcular la varianza total (todos los de la matriz: k^2) no es el mismo que los necesarios para calcular la varianza específica (la diagonal: k).

Por lo tanto, en el numerador hay $k^2 - k^2$ elementos y en el denominador se tienen k elementos, por lo que para que el ratio exprese las magnitudes relativas, más que el número de casos, corregimos por $k^2 / -k^2$ o lo que es lo mismo $k/k - 1$. Así,

la fórmula matemática que queda para el cálculo del alfa de Cronbach, es mostrada en la ecuación 3.8.

Si se requiere expresar el alfa de Cronbach en función de las varianzas-covarianzas, se tiene la ecuación 3.9.

ECUACION 3.9. Alfa de Cronbach por varianza-covarianza.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_y^2} \right)$$

Si se requiere expresar el alfa de Cronbach en función de las correlaciones y no de las varianzas-covarianzas, se puede llegar a la siguiente formulación matemática (Crocker y Algina, 1986), se tiene la ecuación 3.9.

ECUACION 3.10. Alfa de Cronbach por correlación.

$$\alpha = \frac{k \rho}{1 + (k - 1) \rho}$$

donde:

ρ es la media de las correlaciones entre ítems.

Ahora bien, los valores aceptables del alfa de Cronbach:

Nunnally (1987): $\geq .7$

Carmines y Zeller: $\geq .8$

Se debe tener especial atención al número de ítems de la escala y la fase de construcción en la que se encuentra, ya que se sugiere, dependiendo el autor a considerar, se tiene como recomendación:

Alfa de Cronbach, escala en construcción: $.6$ a $.8$

Alfa de Cronbach escala en uso o ajuste normal: $.8$ a $.9$

Alfa de Cronbach escala existente $> .9$ es altamente ajustada

CFA. Cálculo IFC

Como notas importantes, se

- ▶ permite tener en cuenta todos los constructos implicados en la escala y no un análisis uno a uno como Cronbach.
- ▶ requiere la realización de un CFA.

calcula para cada factor implicado (Fornell y Larcker, 1981), (ecuación 3.11).

FIGURA 3.11. Índice de la fiabilidad compuesta.

$$IFC_i = \frac{\left(\sum_j L_{ij}\right)^2}{\left(\sum_j L_{ij}\right)^2 + \sum_j Var(E_{ij})}$$

Fuente: Fornell y Larcker (1981).

donde:

L : es la carga factorial estandarizada de cada uno de los j indicadores que cargan sobre el factor.

$Var(E_{ij})$: es la varianza del término de error. Se calcula como:

$$Var(E_{ij}) = 1 - L_{ij}^2$$

Se debe calcular ya que el software no lo reporta. Siguiendo la tabla 14. Recopilación de información que calcula el IFC, del capítulo 3 para el problema del caso de estudio, tenemos la tabla 5.

Así, los cálculos necesarios para obtener el IFC se ilustran en la tabla 14 del capítulo 3 antes mencionada.

TABLA 5. Recopilación de información que calcula el IFC de ATD del caso de estudio con el CFA.

Constructo e indicadores	Carga factorial Estandarizada L_{ij}	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de fiabilidad compuesta IFC >.7 $\frac{\sum(L_{ij})^2}{[\sum(L_{ij})^2 + \sum Var(E_{ij})]}$
Pensamiento creativo (PC)			
$V_1=ART$.732	.464	.676
$V_2=FSF$.688	.527	
$V_3=IMG$.492	.758	
Total	1.912	1.749	
Pensamiento racional (PR)			
$V_4=MAT$.759	.424	.691
$V_5=FIS$.760	.422	
$V_6=LOG$.415	.828	
Total	1.934	1.674	

Fuente: EQS 6.2, con datos propios..

Observe que en ambos casos es $< .7$. Considerando el apartado CFA. Confiabilidad, y anteriores NO cumple con los requerimientos la escala. Observe que con los promedios de las cargas factoriales $>.6$ enunciados por Hair et al. (1999), SI cumple con el requerimiento de confiabilidad, sin embargo, de acuerdo con los criterios expuestos por Bagozzi y Yi (1988) de cada carga factorial debe ser $>.7$ y con la comprobación de la tabla 5. Recopilación de información que calcula el IFC de ATD del caso de estudio con el CFA, se concluye que la escala diseñada NO cumple los requerimientos de confiabilidad. Observe que las variables que aportan escasamente a la explicación del modelo de ATD, son IMG y LOG.

Acción sugerida: eliminar V3 y realizar de nueva cuenta el análisis.

CFA y EQS. Redefiniendo el modelo

Al realizar la acción sugerida, se renombra el archivo de entrada de datos de programa como: cfa atd sv3.eqs y se documenta en el título como: “CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES SIN V3”. Para eliminar la acción de V3, a ésta se le antepone, en todos los lugares que se mencione, así como su error con el signo “!” (figura 34).

Ejecute nuevamente el programa siguiendo las instrucciones del apartado CFA y EQS. Cómo ejecutar el programa. De los nuevos resultados, observe la sección que las secciones de reporte de matriz de covarianza analizada, de residuos estandarizados y el gráfico de distribución estandarizada de los residuos, permanecen con resultados esperados, sin mayores complicaciones. Sin embargo, observe el apartado Reporte de índices de ajuste (*fit indices*), (figura 35).

En este reporte se encuentran resultados atípicos, por ejemplo:

Nunca debe considerarse índices de ajuste ≥ 1 . Por lo tanto, se descartan el NNFI, CFI; Bollen's; McDonald's.

El RMSEA=0.000 (debe ser: $<.08 < .00$); dado que hay índices de ajuste ≥ 1 , esto validaría el hecho de que hay una alta multicolinealidad de las variables por la supresión de V3. Byrne (2006) y Bentler (1993) condicionan a tomar tres índices, generalmente los primeros tres; como en este caso hay dos que no cumplen, queda en libertad del investigador basarse en los que sí cumplan los requerimientos. En este caso, se tiene la alternativa de utilizar SRMR= 0.019 ya que cumple con estar $<.5$, con la posibilidad de utilizar los índices de ajuste NFI, GFI, AGFI $>.9$ de alto ajuste. Se deberá soportar por marco teórico su uso. (Nota: los índices de ajuste normalizados siempre reportan valores $<$ a los no normalizados ya que están ajustados). Byrne (2006) y Bentler (1993) condicionan a tomar tres índices, generalmente los primeros 3; como en este caso hay dos que

no cumplen, queda en libertad del investigador basarse en los que sí cumplan los requerimientos. Todo lo anterior da a pauta a explicar a mayor detalle las condiciones del modelo dentro de un artículo a publicar.

Asumiendo, que es así, se continúa revisando las secciones de reporte como la mostrada en la figura 36.

¿Son significativos los resultados?, SI, se continúa con el análisis de la sección de reporte de la figura 37.

FIGURA 34. EQS. Codificación de entrada de datos. Caso de estudio 1 con el CFA sin V3.

COPYRIGHT BY P.M. BENTLER VERSION 6.2 (C) 1985 - 2012 (B101).

```
PROGRAM CONTROL INFORMATION

1  /TITLE
2  CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES SIN V3
3  /SPECIFICATIONS
4  VARIABLES= 6; CASES= 275;
5  MATRIX=COR; METHOD=ML; ANAL=COV;
6  /MATRIX
7  1.000
8  0.493 1.000
9  0.401 0.314 1.000
10 0.278 0.347 0.147 1.000
11 0.317 0.318 0.183 0.587 1.000
12 0.284 0.327 0.179 0.463 0.453 1.000
13
14 /STANDARD DEVIATIONS
15 1.0900 0.5900 0.9800 1.1000 0.4100 1.1100
16
17 /LABELS
18 V1=ART; V2=FSF; V3=IMG; V4=MAT; V5=FIS; V6=LOG;
19
20 /EQUATIONS
21 V1=F1+E1;
22 V2=*F1+E2;
23 !V3=*F1+E3;
24 V4=F2+E4;
25 V5=*F2+E5;
26 V6=*F2+E6;
27
28 /VARIANCES
29 E1=*;
30 E2=*;
31 !E3=*;
```

Fuente: EQS 6.2, con datos propios..

FIGURA 35. EQS. Reporte índices de ajuste con el CFA sin V3.

```

26-OCT-17      PAGE:    4  EQS      Licensee:
TITLE:    CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE      =      338.050 ON      10 DEGREES OF FREEDOM

INDEPENDENCE AIC =      318.050    INDEPENDENCE CAIC =      271.882
MODEL AIC =      -4.438            MODEL CAIC =      -22.905

CHI-SQUARE =      3.562 BASED ON      4 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      0.46855

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS      3.559.

FIT INDICES
-----
BENTLER-BONETT      NORMED FIT INDEX =      0.989
BENTLER-BONETT NON-NORMED FIT INDEX =      1.003
COMPARATIVE FIT INDEX (CFI) =      1.000
BOLLEN'S (IFI) FIT INDEX =      1.001
MCDONALD'S (MFI) FIT INDEX =      1.001
JORESJKOG-SORBOM'S GFI FIT INDEX =      0.995
JORESJKOG-SORBOM'S AGFI FIT INDEX =      0.981
ROOT MEAN-SQUARE RESIDUAL (RMR) =      0.017
STANDARDIZED RMR =      0.019
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =      0.000
90% CONFIDENCE INTERVAL OF RMSEA (      0.000,      0.087)
    
```

Fuente: EQS 6.2, con datos y adaptación propios.

¿La covarianza entre las variables independientes es ≥ 1 ? NO, ya que es de .37.

Nota: la covarianza NUNCA debe ser $\geq .8$

¿Es significativa al .5? SI, ya que tienen @ asignada. Se asegura alta probabilidad de validez y se continúa con el análisis de la sección de reporte de la figura 38.

FIGURA 36. EQS. Reporte ecuaciones de medida con el CFA sin V3.

```

26-OCT-17      PAGE:    5  EQS      Licensee:
TITLE:    CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

ART  =V1  =  1.000 F1    +1.000 E1

FSF  =V2  =  .610*F1    +1.000 E2
          .099
          6.139@

MAT  =V4  =  1.000 F2    +1.000 E4

FIS  =V5  =  .368*F2    +1.000 E5
          .039
          9.518@

LOG  =V6  =  .811*F2    +1.000 E6
          .095
          8.582@
    
```

Fuente: EQS 6.2, con datos propios..

FIGURA 37. EQS. Reporte covarianzas entre las variables independientes con el CFA.

```

26-OCT-17      PAGE:    8  EQS      Licensee:
TITLE:    CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

COVARIANCES AMONG INDEPENDENT VARIABLES
-----
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

          V          F
          ---          ---
          I F2 - F2          .370*I
          I F1 - F1          .070 I
          I                    5.292@I
          I                    I
    
```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 38. EQS. Reporte solución estandarizada con el CFA.

```

26-OCT-17      PAGE:    9  EQS      Licensee:
TITLE:    CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

STANDARDIZED SOLUTION:                                R-SQUARED

ART   =V1   =  .661 F1      + .750 E1                .437
FSF   =V2   =  .745*F1     + .667 E2                .556
MAT   =V4   =  .765 F2      + .644 E4                .585
FIS   =V5   =  .755*F2     + .656 E5                .569
LOG   =V6   =  .615*F2     + .789 E6                .378
    
```

Fuente: EQS 6.2, con datos y adaptación propios.

¿Cargas factoriales >.6? SI, se continúa con el cálculo de la fiabilidad compuesta (tabla 6).

TABLA 6. Recopilación de información que calcula el IFC de ATD. Caso de estudio: el CFA sin V3.

Constructo e indicadores	Carga Factorial estandarizada L_{ij}	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Indice de fiabilidad compuesta IFC >.7 $\frac{\sum(L_{ij})^2}{[\sum(L_{ij})^2 + \sum Var(E_{ij})]}$
Pensamiento creativo (PC)			
V ₁ =ART	.661	.563	.662
V ₂ =FSF	.745	.445	
Total	1.406	1.008	
Pensamiento racional (PR)			
V ₄ =MAT	.765	.415	.757
V ₅ =FIS	.755	.430	
V ₆ =LOG	.615	.622	
Total	2.135	1.467	

Fuente: EQS 6.2, con datos y adaptación propios.

Dado que el IFC $>.6$ y partiendo de la tabla 15. Recopilación de información que calcula el IVE del capítulo 3 previa, tenemos para nuestro caso la tabla 7, como sigue:

TABLA 7. Recopilación de información que calcula el IVE de ATD. Caso de estudio: el CFA sin V3.

Constructo e indicadores	Cuadrado de la carga factorial estandarizada $(L_{ij})^2$	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de varianza compuesta IVE $\geq .5$ $\frac{\sum(L_{ij})^2}{[\sum(L_{ij})^2 + \sum Var(E_{ij})]}$
Pensamiento creativo (PC)			
V ₁ =ART	.437	.563	.496
V ₂ =FSF	.555	.445	
Total	.992	1.008	
Pensamiento racional (PR)			
V ₄ =MAT	.585	.415	.511
V ₅ =FIS	.570	.430	
V ₆ =LOG	.378	.622	
Total	1.533	1.467	

Fuente: EQS 6.2, con datos y adaptación propios.

Recuerde que la confiabilidad y la validez se encuentran estrechamente relacionadas, desempeñando papeles complementarios. La confiabilidad trata acerca de la consistencia, exactitud y capacidad de predicción de los hallazgos. La validez, está relacionada con el problema de que estamos realmente midiendo lo que creemos, por lo que es un asunto más amplio y complejo que la confiabilidad.

CFA y EQS. Determinando la validez

Recuerde lo citado en los apartados Validez de versión, Validez relacionada con el criterio y Validez nomológica en que la validez de una escala, sucede cuando la variable latente que están midiendo los indicadores, es realmente lo que se quiere medir. También representa el grado en que un instrumento mide el concepto bajo estudio (Bohrnstedt, 1976).

Recuerde que existen en resumen, los siguientes tipos de validez más utilizados:

Validez de contenido. Grado en que una escala recoge todas las dimensiones teóricas del concepto que se quiere medir (Se refleja en las cargas factoriales $>.6$).

Validez de constructo o concepto.

- ▶ **Convergente.** Los indicadores utilizados para medir un mismo concepto, están fuertemente correlacionados entre sí. (Sucede sólo entre los ítems de cada factor del constructo; es decir, los ítems de un factor sólo se relacionan para dicho factor, o sea, convergen).
- ▶ **Discriminante.** Los indicadores utilizados para medir conceptos distintos están escasamente correlacionados entre sí. (Sucede entre al menos dos o más factores o constructos, midiendo cuestiones diferentes con un mínimo de relación).

Validez nomológica. El constructo medido por los indicadores propuestos guarda relaciones con otros conceptos teóricamente previsibles (concurrente o predictiva). En otras palabras, ¿el modelo teórico inicial es mejor que el ajustado? Se espera que NO. Así, si el modelo teórico sí es mejor que el ajustado, se dice que NO hay validez nomológica. Sin embargo, si el modelo teórico no es mejor al ajustado, se dice que SI hay validez nomológica. Esto es, que el modelo ajustado está más cercano a la explicación de la realidad que el modelo teórico inicial. En nuestro caso, el suprimir la V3 hizo una mejora a los índices del modelo teórico original, por lo que SI se obtuvo validez nomológica. Cuando el modelo teórico coincide con la realidad, NO existe validez nomológica. Al suprimir variables, por baja explicación, se está ajustando el modelo mejorándolo respecto al teórico (SI hay validez nomológica), sin embargo, no se recomienda suprimir más de 10%. Lo ideal es que NO exista validez nomológica (modelo teórico=modelo ajustado= coincide con la realidad). Cuando existe validez nomológica la teoría NO coincide con la realidad, por lo que se deberá buscar explicación de la realidad del porqué no se consideró en la teoría. En nuestro caso, la V3 suprimida no implica que la teoría esté errónea, sino que en la realidad, no se está llevando a cabo su práctica. Esta validez podría medirse haciendo la diferencia de la Chi-cuadrado del modelo teórico y ajustado y el resultado buscar los grados de libertad en qué nivel de error se estarían manejando los ajustes y a que nivel de significatividad. La validez nomológica, simple y llanamente se demuestra al suprimir o no las variables del modelo.

En cuanto a la validez discriminante, recuerde lo citado en el apartado 4.24, de los test más utilizados:

- ▶ Test de intervalos de confianza (Anderson y Gerbing, 1988).
- ▶ Test de la varianza extraída (Fornell y Larcker, 1981).
- ▶ Test de diferencias de las Chi-cuadrado (Anderson y Gerbing, 1988).

Con un solo test que soporte el modelo teórico, basta para demostrar que existe validez discriminante. Se estila hacer los dos primeros.

CFA y EQS. Cálculo de la validez discriminante. Test del intervalo de confianza

Siguiendo el caso y considerando que es validez de los constructos (no deos ítems), se observa que se tienen dos constructos: F1 y F2. Recuerde que de la figura 37.

FIGURA 37. EQS Reporte covarianzas entre las variables independientes.

```

26-OCT-17      PAGE:   8  EQS      Licensee:
TITLE:   CFA ATD SV3. CFA EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES
MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)
COVARIANCES AMONG INDEPENDENT VARIABLES
-----
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

```

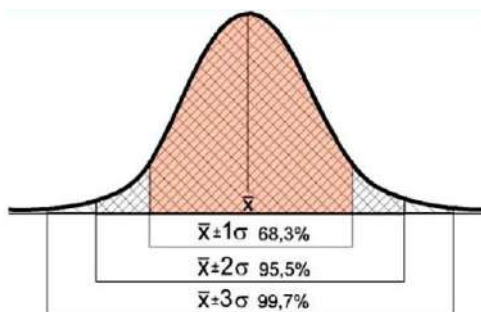
V		F	
	I F2 - F2		.370*I
	I F1 - F1		.070 I
	I		5.292@I
	I		I

Fuente: EQS 6.2, con datos y adaptación propios.

La covarianza de las variables independientes NO debe tener 1, para que tenga validez discriminante. Se debe realizar el cálculo de la siguiente manera:

En EQS se codifica: * F1-F2

FIGURA 38. Campana de Gauss y la desviación estándar.



Fuente: Levin y Rubin (2004).

$$\text{Intervalo límite inferior} = .370 - 2 (.070) = .230$$

$$\text{Intervalo límite superior} = .370 + 2 (.070) = .510$$

El 2 se refiere al nivel de confiabilidad de %95.5 al que se desea trabajar, como el factor más común utilizado en las ciencias de la administración (figura 38).

Del resultado, se lee: “entre el límite inferior y superior entre los factores F1 y F2, NO debe existir la unidad (1) ya que los factores estarían midiendo lo mismo”. En nuestro caso (.230 - .510), por lo tanto SÍ cumple con las condiciones del test, hay validez discriminante. En otras palabras, los factores miden cuestiones diferentes.

Nota: la covarianza NUNCA debe ser $\geq .8$ ya que el cálculo de los límites inferior y superior, al multiplicar por 2 el error, forzarían a que el resultado sea 1.

Si fueran más factores se deben considerar las diversas combinaciones y cumplir con el requerimiento. Ejemplo: cinco factores.

En EQS, se codifica: F1 to F5= *;

Realizando diez combinaciones.

F1-F2; F1-F3; F1-F4; F1-F5; F2-F3; F2-F4; F2-F5; F3-F4; F3-F5; F4-F5 y cumplir las condiciones de los límites.

CFA y EQS. Cálculo de la validez discriminante.

Test de la varianza extraída

Tomando en cuenta la sección de reporte covarianzas en tres las variables independientes (*covariances among the independent variables*) de la figura 37, se debe determinar la covarianza al cuadrado $= (.370)^2 = .137$. La lectura, para nuestro caso de estudio, es:

- ▶ ¿Es el cuadrado de la covarianza de las variables independientes F1-F2 es menor (<) a los IVEs calculados (tabla 7)? o
- ▶ ¿Los valores calculados de los IVEs (tabla 7), son mayores (>) al cuadrado de la covarianza de las variables independientes F1-F2?

Así, de la tabla 7, tenemos los valores de IVE como $IVE_{PC} = .496$ e $IVE_{PR} = .511$. Por lo tanto, SÍ cumple con las condiciones del test, hay validez discriminante.

CFA y EQS. Reportando resultados: consistencia interna y validez convergente

En este punto, es importante tener una referencia del acomodo de los datos, como se sugiere en el ejemplo de la figura 39.

FIGURA 39. Reporte tabla de consistencia interna y validez convergente del modelo teórico sin V3.

Factor	Item V	Variable	Carga Factorial >0.6 (a)	Valor robusto t	Factor promedio de carga factorial	Fiabilidad de la Escala		
						Alfa de Cronbach >=0.7 (b)	IFC >0.7 (b)	IVE >0.5 (c)
F1 (PC)	1	ART	.661***	1.000a	0.703	-	.662	.496
	2	FSF	.745***	6.139				
F2 (PR)	4	MAT	.765***	1.000a	0.712	-	.757	.511
	5	FIS	.755***	9.518				
	6	LOG	.615***	8.582				

Resultados: Chi-cuadrada= 3.562; GL= 4; $p < .468$; NFI = .989; GFI = .995; AGFI = .981; SRMR = .019.
 Conclusión: las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos ya que se presentan con una alta probabilidad de ocurrencia de 99% (significancia .01, marcado con ***). Sin embargo, el $p = .468$ es muy alto, o sea, representa un error del 46.8% (debe ser $< .05$). Esto es, no hay error en el ajuste del modelo, pero SI existe error en los índices (varios son > 1). Aunque existe suficiente evidencia de validación convergente y fiabilidad, que justifica la fiabilidad interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999), los datos reportados, se deben considerar de forma muy cuidadosa ya que presentan error.

Nota:

*** Parámetros constreñidos a ese valor en el proceso de identificación = $p < 0.001$.

(a) De acuerdo con Bagozzi y Yi (1988).

(b) De acuerdo con Hair et al. (1999).

(c) Índice de Varianza Extraída (IVE), de acuerdo con Fornell y Larcker (1981).

Fuente: EQS 6.2 y SPSS 21 IBM, con datos y adaptación propios.

donde las columnas significan:

Factores: listado de los involucrados en el estudio. En nuestro caso pensamiento creativo: F1 (PC) y pensamiento racional: F2 (PR).

Ítem, consecutivo de las variables (V) mostradas, como: V1; V2; V3, etcétera.

Variable, en donde se ubica las siglas que representa cada una de las variables no suprimidas: V1; V2, V4, V5, V6, etc. En nuestro caso: ART, FSE, MAT, FIS, LOG.

Carga factorial $> .6$ (a) (figura 38) donde sólo se reportan los factores no suprimidos y con carga factorial $> .6$. Es posible que el investigador utilice escalas previas que si a su juicio, por el marco teórico, deban asignarse más ítems, es válido realizarlo siempre y cuando, se mencione que la nueva escala es una adaptación de la previa más la añadida como propuesta del investigador dadas las condiciones de la realidad en la que se aplica contra las que originalmente la escala se creó. En esta etapa es donde se valida o no la pertinencia de los nuevos ítems basados en el marco teórico planteado.

Valor Robusto t (figura 36, que se lee por cada variable: “covarianza; error y valor de t significativo a .5), donde se apreciará que hay 1 Ítem de cada factor con coeficiente 1.000a que implica que tiene varianza 1, como condición requerida mínima de que una variable tenga dicho valor de la varianza, para que se realice el cálculo por factor como parte del proceso de identificación; esto implica que no habrá valor de t dado que no se puso asterisco (*) en el arranque del software EQS del programa de entrada y, por lo tanto, no reportará el valor de t”.

Debemos recordar que:

- ▶ Un nivel de 95% de confianza incluirá 47.5% del área que se encuentra a ambos lados de la media de la distribución de muestreo (figura 40 a 0.475. Descubrimos que 0.475 del área bajo la curva normal está contenida entre la media y un punto situado a 1.96 errores estándar a la derecha de la media. Por consiguiente, sabemos que $(2) (0.475) = 0.95$ del área está localizada entre ± 1.96 errores estándar de la media. En consecuencia, 95% del área está contenida entre ± 1.96 errores estándar de la media. ($p=.05$).
- ▶ Un nivel de confianza de 99% incluiría 49.5% del área que se encuentra a

FIGURA 40. Áreas bajo la curva de distribución de probabilidad normal estándar entre la media y los valores positivos de Z.

Ejemplo:	z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
Para encontrar el área bajo la curva entre la media y un punto que está a 2.24 desviaciones estándar a la derecha de la media, busque el valor que se encuentra a la altura del renglón correspondiente a 2.2 y en la columna del 0.04; 0.4875 del área bajo la curva se encuentra entre la media y un valor de z de 2.24.	0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
	0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
	0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
	0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
	0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
	0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
	0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
	0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
	0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
	0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
	1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
	1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
	1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
	1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
	1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
	1.8	0.4641	0.4649	0.4655	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
	1.9	0.4713	0.4719	0.4725	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
	3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Fuente: Levin y Rubin (2004).

cualquier lado de la media de la distribución de muestreo. Si observamos la figura 40, veremos que 0.495 del área bajo la curva normal está localizada entre la media y un punto que se encuentra a ± 2.58 errores estándar de la media. En consecuencia, 99% del área está contenida entre ± 2.58 errores estándar de la media ($p=.01$).

Con lo anterior, es posible afirmar que si los valores de t (ver apartado reporte ecuaciones de medida measurement equations, figura 36) son mayores ($>$) 2.58 errores estándar de la media, entonces se anotarán tres asteriscos (***) por carga factorial, lo cual indica que $p<.0001$. Así se tienen una seguridad que las cargas factoriales aportan sus valores a 99% de confiabilidad.

Alfa de Cronbach, no es posible calcularla directamente dado que se insertaron los datos de nuestro ejemplo a partir de una matriz de correlación (figura 13). Sería factible calcularla separadamente si los datos estuvieran contenidos en una base de datos (como SPSS); sin embargo, no se tuvieron. Lo que se ha calculado con EQS hasta aquí es el alfa de Cronbach= .706 (figura 7.41) de todo el modelo general, sin

FIGURA 41. EQS. Reporte índices de ajuste. Caso de estudio: el CFA sin V3.

```

INDEPENDENCE AIC =      318.050   INDEPENDENCE CAIC =      271.882
MODEL AIC =          -4.438     MODEL CAIC =          -22.905

CHI-SQUARE =           3.562 BASED ON      4 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      0.46855

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS      3.559.

FIT INDICES
-----
BENTLER-BONETT NORMED FIT INDEX =      0.989
BENTLER-BONETT NON-NORMED FIT INDEX =      1.003
COMPARATIVE FIT INDEX (CFI) =      1.000
BOLLEN'S (IFI) FIT INDEX =      1.001
MCDONALD'S (MFI) FIT INDEX =      1.001
JORESKOG-SORBOM'S GFI FIT INDEX =      0.995
JORESKOG-SORBOM'S AGFI FIT INDEX =      0.981
ROOT MEAN-SQUARE RESIDUAL (RMR) =      0.017
STANDARDIZED RMR =      0.019
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =      0.000
90% CONFIDENCE INTERVAL OF RMSEA (      0.000,      0.087)

RELIABILITY COEFFICIENTS
-----
CRONBACH'S ALPHA =      0.706
RELIABILITY COEFFICIENT RHO =      0.771

```

Fuente: EQS 6.2, con datos propios..

embargo, lo que se requiere es realizarlo por factor, para comprobar la confiabilidad de la escala. Se tiene confianza de que el reporte de confiabilidad por factor tienda a ser superior a .7 (se debe calcular teniendo el desglose de la base de datos en SPSS y comprobar).

IFC. Se toman los datos de la figura 6.

IVE. Se toman los datos de la figura 7.

Hasta este momento, tomando en cuenta la figura 39 afirmamos que para maximizar la destreza de la asertividad en la toma de decisiones (ATD) en los ejecutivos de la industria del software, es necesario capacitarlos más en filosofía (FSF) y artes (ART) que en técnicas de imaginación (IMG), para que se fomente el pensamiento creativo (PC) y que esto se debe complementar con una capacitación mayor en matemáticas (MAT), física (FIS), que en lógica (LOG) para reforzar su pensamiento racional (PR). Todo lo anterior basado en las cargas factoriales encontradas con una probabilidad de equivocación de 1/100 o 99% de que lo afirmado ocurra (ver las ***).

donde los renglones por debajo de las cargas factoriales, significan:

-S-B χ^2 Índice Satorra Bentler Chi-cuadrado, es el recomendado cuando se tienen más de dos factores para generar datos robustos. Sin embargo, nuestro ejemplo no corresponde, por lo que es válido reportar la Chi-cuadrado, con sus grados de libertad. Así, se explica la lectura reportada en la base de la figura 39:

“Conclusión:

Resultados: Chi-Cuadrado= 3.562; gl= 4; p < .468; NFI = .989; GFI = .995; AGFI = .981; SRMR = .019

Conclusión: las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos. Sin embargo, el p= .468 es muy alto, o sea, representa un error de 46.8% (debe ser <.05). Esto es, no hay error en el ajuste del modelo, pero sí hay error en los índices (varios son >1). Aunque existe suficiente evidencia de validación convergente y fiabilidad, que justifica la fiabilidad interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999), los datos se deben considerar de forma muy cuidadosa ya que presentan error”.

De no haber tenido el problema p y la falta de alfa de Cronbach por constructo, una posible redacción de los resultados de la figura 39, sería:

“...Por otro lado, para evaluar la fiabilidad y validez de las escalas de medida se realizó un Análisis Factorial Confirmatorio (AFC) utilizando el método de máxima verosimilitud con el software EQS 6.2” (Bentler, 2005; Brown, 2006; Byrne, 2006). Asimismo, la fiabilidad de las escalas de medida se evaluó a partir del coeficiente alfa de Cronbach y el índice de fiabilidad compuesta (IFC) (Bagozzi y Yi, 1988). Todos los valores de la escala excedieron el nivel recomendado de .70 para el alfa de

Cronbach y el IFC que proporciona una evidencia de fiabilidad y justifica la fiabilidad interna de las escalas (Nunnally y Bernstein, 1994; Hair et al., 1995).

Los resultados de la aplicación del AFC se presentan en la tabla 1 y sugiere que el modelo de medición proporciona un buen ajuste de los datos ($S-BX2 = XXX.YYY$; $df = XX$; $p = 0.000$; $NFI > .9XXX$; $NNFI > .9XXX$; $CFI > .9XXX$; y $RMSEA < .08XXX$). Como evidencia de la validez convergente, los resultados del AFC indican que todos los ítems de los factores relacionados son significativos ($p < 0.001$) y el tamaño de todas las cargas factoriales estandarizadas son superiores a 0.60 (Bagozzi y Yi, 1988).

La tabla xxx muestra una alta consistencia interna de los constructos, en cada caso, la α de Cronbach excede el valor de .70 recomendado por Nunnally y Bernstein (1994). El IFC representa la varianza extraída entre el grupo de variables observadas y el constructo fundamental (Fornell y Larcker, 1981). Generalmente un IFC superior a .60 es considerado como deseable (Bagozzi y Yi, 1988), en este estudio este valor es superior. El índice de la varianza extraída (IVE) fue calculado para cada par de constructos, resultando un IVE superior a .5 en todos los factores (Fornell y Larcker, 1981)...”

CFA. Reporte de resultados: validez discriminante del modelo teórico

CON el fin de facilitar la presentación de resultados de la validez discriminante, se sugiere utilizar lo mostrado en la figura 42.

FIGURA 42. Reporte tabla de validez discriminante del modelo teórico.

Variable	F1 Pensamiento creativo (PR)	F2 Pensamiento racional (PR)
F1 Pensamiento creativo (PR)	.496	.137
F2 Pensamiento racional (PR)	.230 -. 510	.511

Fuente: EQS 6.2, con datos propios..

Esta tabla de reporte, se recopila de la siguiente manera:

- ▶ La diagonal es representada por las IVE calculadas (figura 39).
- ▶ Los datos por debajo de la diagonal, son los límites superior-inferior de los factores F1-F2 (ver el apartado Cálculo de la validez discriminante. Test del

intervalo de confianza).

- ▶ Los datos por arriba de la diagonal, son las covarianzas al cuadrado entre las variables independientes (ver el apartado Cálculo de la validez discriminante. Test de la varianza extraída).

La tabla de reporte, se interpreta de la siguiente manera:

- ▶ Los datos por debajo de la diagonal, no deben contener 1. Por lo tanto, cumple con validez convergente y validez discriminante. Test del intervalo de confianza.
- ▶ Los datos por arriba de la diagonal. IVE debe ser mayor a los datos capturados por arriba de la columna y también a los datos que se capturen a la derecha. Por lo tanto, cumple con validez discriminante. Test de la varianza extraída.

Una posible redacción de los resultados de la figura 39, sería:

“Por lo que respecta a la evidencia de la validez discriminante, la medición se proporciona en dos formas que se pueden apreciar en la tabla yyy? Primero, con un intervalo del 95% de confidencialidad, ninguno de los elementos individuales de los factores latentes de la matriz de correlación, contiene el valor 1.0 (Anderson y Gerbing, 1988). Segundo, la varianza extraída entre cada par de constructos es superior que su correspondiente IVE (Fornell y Larcker, 1981). Con base en estos criterios, se puede concluir que las distintas mediciones realizadas en este estudio demuestran suficiente evidencia de fiabilidad y validez convergente y discriminante”.

CFA. Conclusiones

La asertividad en la toma de decisiones (ATD) de los ejecutivos de la industria del software, tiene una mayor explicación en el conocimiento del arte (ART) y de la filosofía (FSF) en su componente de pensamiento creativo (PC) que de la imaginación (IMG). Por lo tanto, invierta menos técnicas de la imaginación (IMG) e invierta más en arte (ART) y filosofía (FSF).

La asertividad en la toma de decisiones (ATD) de los ejecutivos de la industria del software, tiene una mayor explicación en las matemáticas (MAT) y en la física (FIS) que en la lógica (LOG) en su componente de pensamiento racional (PR). Por lo tanto, invierta menos en las técnicas de la lógica (LOG), e invierta más en matemáticas (MAT) y física (FIS).

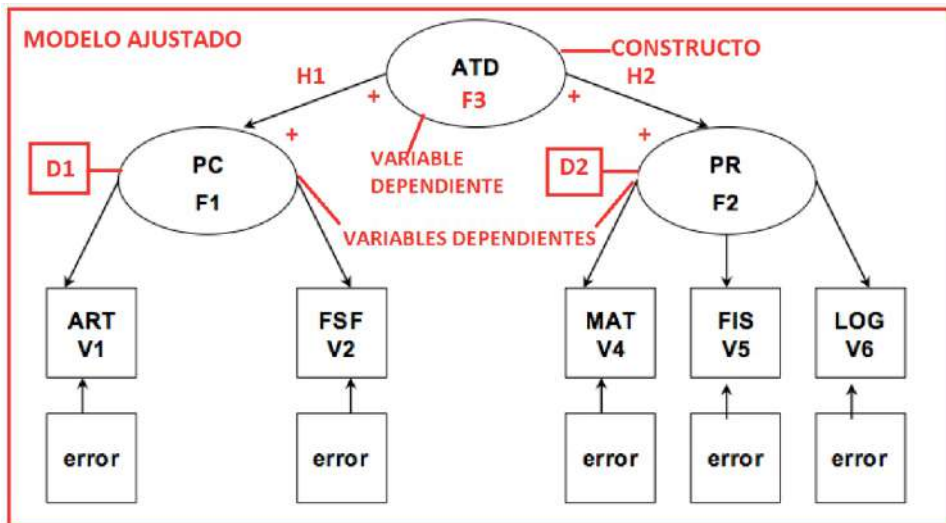
Por lo tanto, la hipótesis: “la ATD de los ejecutivos de la industria del software es explicada mediante las notas del ART; FSF e IMG que componen al PC, así como las notas de MAT; FIS y LOG que componen al PR” sí es explicada.

Dado que existe confiabilidad y validez, se procede a realizar el análisis estructural (SEM. *structural equation modeling*).

SEM y EQS. Calculando la ecuación estructural SEM

Recuerde que en el CFA se realizan análisis de confiabilidad y validez a través de correlaciones entre los ítems y los factores del constructo. Con el análisis estructural, y teniendo como base el CFA, ya se tiene la capacidad de realizar prueba de hipótesis en el constructo, el cual es denominado también, modelo ajustado. Así, partiendo de los resultados mostrados en la figura 43, tenemos:

FIGURA 43. Constructo o modelo ajustado de la asertividad en la toma de decisiones (ATD) para el análisis estructural.



Fuente: Elaboración propia.

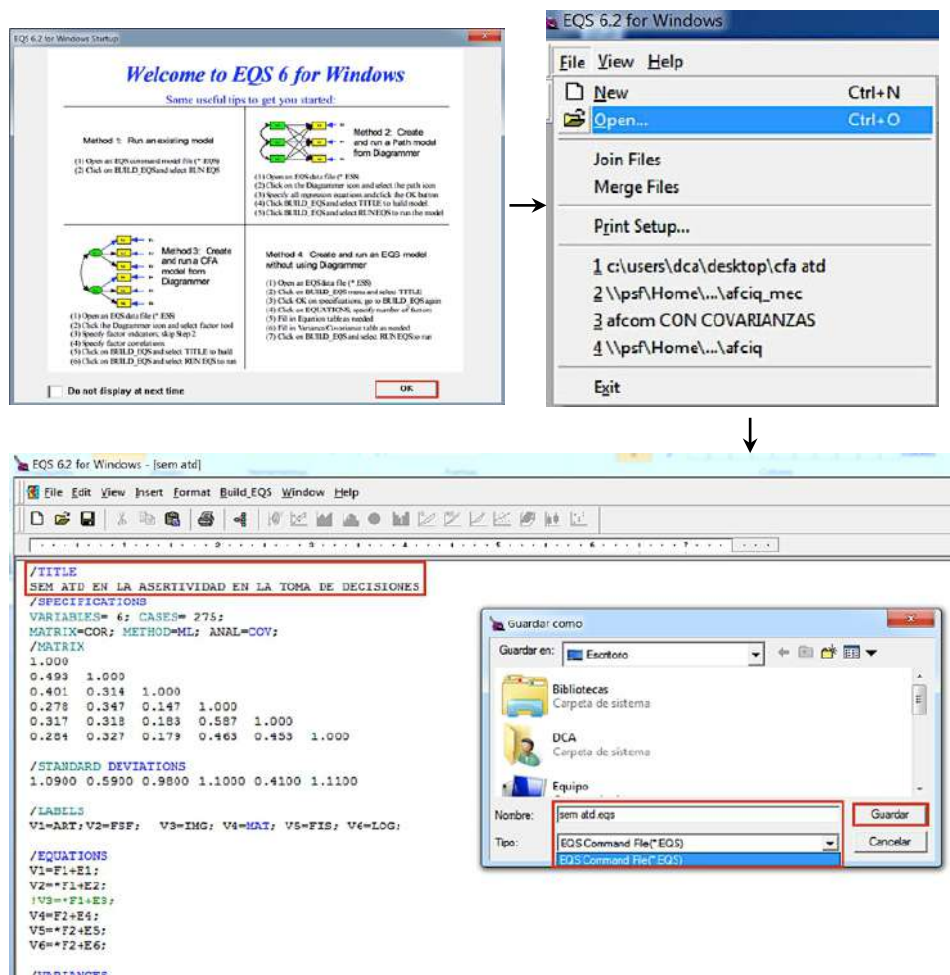
Así, es posible plantear las siguientes hipótesis:

- ▶ H1: "A mayor ATD, mayor PC".
- ▶ H2: "A mayor ATD, mayor PR".

Para verificarlo, se requiere entrar al programa de entrada de datos de EQS y renombrar el archivo como sem atd.eq; se sugiere asignar el título: "sem atd en la asertividad para la toma de decisiones", mediante la siguiente serie de comandos: Icono EQS 6.2 for windows->OK->File->Open->Nombre: sem atd.eq->Tipo: EQS System Data (*ESS)->Abrir->cambio título->Guardar (figura 44).

Así, se deberá realizar el ingreso del nuevo modelo estructural donde se declararán las relaciones del modelo ajustado, tomando en cuenta los factores F1

FIGURA 44 . EQS. Secuencia de comandos para apertura de programa de entrada de datos. Análisis SEM.



Fuente: EQS 6.2, con datos propios.

(PC) y F2 (PR) hacia el nuevo factor o constructo a medir F3 (ATD); también se deberán incluir los efectos provocados por los disturbios (D1 y D2) dado que los factores F1 (PC) y F2 (PR) son tratados ahora como variables dependientes (los errores quedan implícitos en cada uno de los ítems que los conforman), que afectan a F3 (variable independiente), que ya se puede medir en el tratamiento de SEM (en CFA NO era factible medirlo). Nota: recuerde que una ecuación SEM de primer

orden sí se puede medir dado que los factores tienen los ítems; por lo tanto, en una ecuación SEM de segundo orden (ATD, se convierte en un factor de este tipo) dependerá su medición de manera indirecta a través de los factores primer orden. Lo anterior genera expresiones que se deben integrar al programa de entrada en el párrafo /EQUATIONS, tales como: $F1 = *F3 + D1$; $F2 = *F3 + D2$, por lo que se lee: “calcula el valor de F1 respecto a F3, considerando el error generado por F1”.

Por lo anterior, recordar que sólo tienen varianzas las variables independientes y, por lo tanto, al revisar dicho estatus de las nuevas variables dependientes F1(PC) y F2 (PR) éstas NO varían o covarían, por lo que se concluye que ya no debe considerarse sus varianzas en el programa de entrada EQS; se deben suprimir. Al tener sólo un factor independiente F3 (ATD) y no tener otro factor para covariar, se SUPRIME los contenidos en el párrafo /COVARIANCES (figura 45).

FIGURA 45 . EQS. Ingreso de expresiones del SEM al programa de entrada en los párrafos: /EQUATIONS /VARIANCES y /COVARIANCES.

```

/EQUATIONS
V1=F1+E1;
V2=*F1+E2;
!V3=*F1+E3;
V4=F2+E4;
V5=*F2+E5;
V6=*F2+E6;

F1=*F3+D1;
F2=*F3+D2;

/VARIANCES
E1=*;
E2=*;
!E3=*;
E4=*;
E5=*;
E6=*;

!F1 TO F2**; SE CONVIRTIERON EN VARIABLES DEPENDIENTES POR LO TANTO YA NO VARIAN...
!SÓLO LO HACEN LAS VARIABLES INDEPENDIENTES EN PARES
!F3 NO VARIA YA QUE ES LA UNICA QUE EXISTE...
!D1 y D2 SE AGREGAN PORQUE SON ERRORES DE FACTORES DEPENDIENTES
D1 TO D2**;

/COVARIANCES
!F1 TO F2**; SE ANULA; F3 NO SE CONSIDERA PORQUE ES VARIABLE INDEPENDIENTE UNICA

/PRINT
FIT=ALL;

```

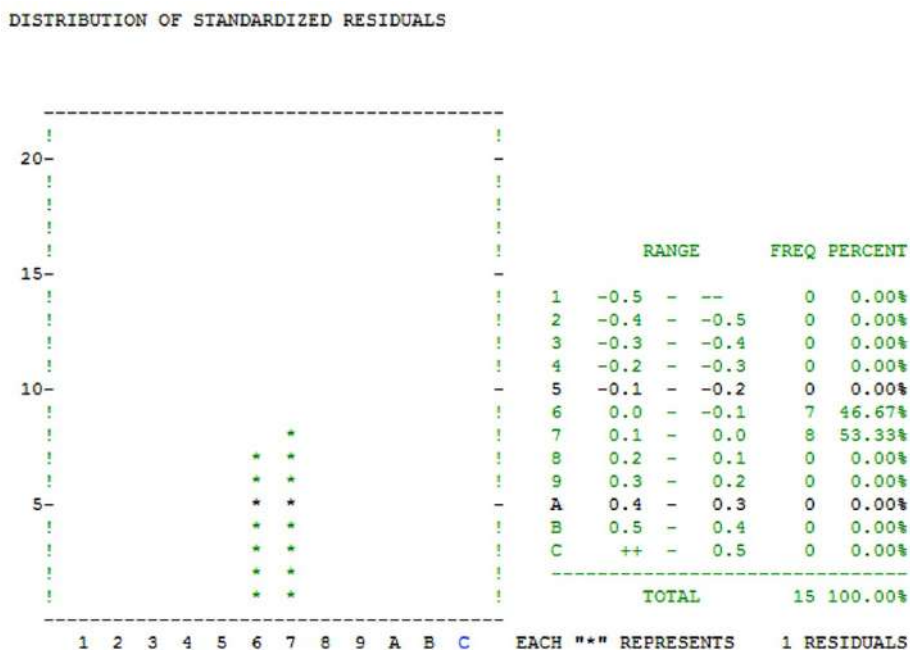
Fuente: EQS 6.2, con datos propios..

Se guardan el archivo y se corre conforme a lo indicado en la figura 14.

SEM y EQS. Análisis de resultados

Verifique los resultados más relevantes de las secciones de reporte, como en la figura 46.

FIGURA 46. EQS. Gráfico distribución estandarizada de los residuos con el SEM.



Fuente: EQS 6.2, con datos y adaptación propios.

Donde se observa, nuevamente que los datos están adecuadamente distribuidos, ya que no existen desviaciones estándar ni a la izquierda ni a la derecha del intervalo. Esto es un indicativo de que el modelo conserva su ajuste.

Continuar revisando hasta verificar la sección de reporte de ajuste de índices (*fit indices*) de la figura 47.

Observe que al haber realizado todos los ajustes del SEM al ingresar en el programa de entrada EQS de factores y errores, así como la supresión-agregados de varianzas y supresión de covarianzas se modifica la distribución de la información a analizar, lo que trae como consecuencia una mejora del modelo, respecto al del CFA, en sus índices de ajuste $>.9 < 1.0$ (ya no aparecen los 1), RMSEA $<.08$. Lo cual se lee:

“Chi-cuadrado= 3.562; gl=2; p= .16849; NFI=.989; NNFI=.976; CFI=.995; RMSEA=.053”.

Se continúa la revisión hasta llegar a la sección de reporte mostrado en la figura 48.

FIGURA 47. EQS. Reporte de índices de ajuste con el SEM.

```

INDEPENDENCE AIC =      318.050   INDEPENDENCE CAIC =      271.882
      MODEL AIC =       -0.438       MODEL CAIC =       -9.672

CHI-SQUARE =          3.562 BASED ON          2 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS          0.16849

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS          3.559.

FIT INDICES
-----
BENTLER-BONETT      NORMED FIT INDEX =          0.989
BENTLER-BONETT NON-NORMED FIT INDEX =          0.976
COMPARATIVE FIT INDEX (CFI) =          0.995
BOLLEN'S (IFI) FIT INDEX =          0.995
MCDONALD'S (MFI) FIT INDEX =          0.997
JORESKOG-SORBOM'S GFI FIT INDEX =          0.995
JORESKOG-SORBOM'S AGFI FIT INDEX =          0.961
ROOT MEAN-SQUARE RESIDUAL (RMR) =          0.017
STANDARDIZED RMR =          0.019
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =          0.053
90% CONFIDENCE INTERVAL OF RMSEA (          0.000,          0.142)

RELIABILITY COEFFICIENTS
-----
CRONBACH'S ALPHA =          0.706

```

Fuente: EQS 6.2, con datos y adaptación propios.

Las tres variables del modelo muestran altos valores de t significativos al .05.

Continúe revisando y observe la sección de reporte de la figura 49.

El cual se lee: “ésta es la medición F1-F3 con una covarianza de F1 en F3 de .663, con un error de .159 así como un valor de t alto y significativo al .05”; del segundo caso, se tiene la siguiente lectura: “ésta es la medición F2-F3 con una covarianza de F2 en F3 de .694, con un error de .000 así como un valor extremadamente alto de t y significativo al .05”.

Analizando más, observamos la sección del reporte de la figura 50.

FIGURA 48. EQS. Reporte: medición de las ecuaciones con el SEM.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

ART	=V1	=	1.000 F1	+	1.000 E1
FSF	=V2	=	.610*F1 .099 6.139@	+	1.000 E2
MAT	=V4	=	1.000 F2	+	1.000 E4
FIS	=V5	=	.368*F2 .039 9.518@	+	1.000 E5
LOG	=V6	=	.811*F2 .095 8.583@	+	1.000 E6

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 49. EQS. Ecuaciones de constructo con errores estándar con el SEM.

28-OCT-17 PAGE: 8 EQS Licensee:
TITLE: SEM ATD EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES
MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)
CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

F1	=F1	=	.663*F3 .159 4.164@	+	1.000 D1
F2	=F2	=	.694*F3 .000 1.0E+38@	+	1.000 D2

Fuente: EQS 6.2, con datos y adaptación propios.

De este reporte, sólo es informativo y se prosigue con la revisión, localizando la sección de reporte mostrada en la figura 52. De esta figura, se aprecian las cargas factoriales que inciden de las variables a los factores y entre factores.

A fin de determinar la significatividad de la relación de factores F1-F3 y F2-F3, se procede a revisar la figura 49 y las consideraciones mostradas de la figura 40, de las que se procede a realizar, lo siguiente:

- ▶ Para la relación F1-F3 y de la figura 49, el valor de $t = 4.164 > 2.58$ (figura 40), por lo tanto se asignan: tres asteriscos (***)
- ▶ Para la relación F2-F3 y de la figura 49, el valor de $t = 1.0E+38@$, representa un valor extremadamente alto que se deberá calcular a partir de calcular el promedio de los valores t de todas las variables relacionadas con F2 (figura 48), de tal forma que:
 - ▶ Valor de t de F2-F3= Promedio de las variables relacionadas con F2= Promedio (V5-F2+ V6-F2) (figura 48).
 - ▶ Valor de t de F2-F3= $(9.518+8.583)/2 = 9.050 > 2.58$ (figura 40), por lo tanto se asignan: tres asteriscos (***)

SEM y EQS. Discusión

Lo anterior, nos permitirá hacer el gráfico correspondiente al diagrama de trayectorias para responder a las hipótesis planteadas como se muestra en la figura 53.

FIGURA 52. EQS. Reporte solución estandarizada con el SEM.

```

28-OCT-17      PAGE: 11 EQS      Licensee:
TITLE:  SEM ATD EN LA ASERTIVIDAD EN LA TOMA DE DECISIONES

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

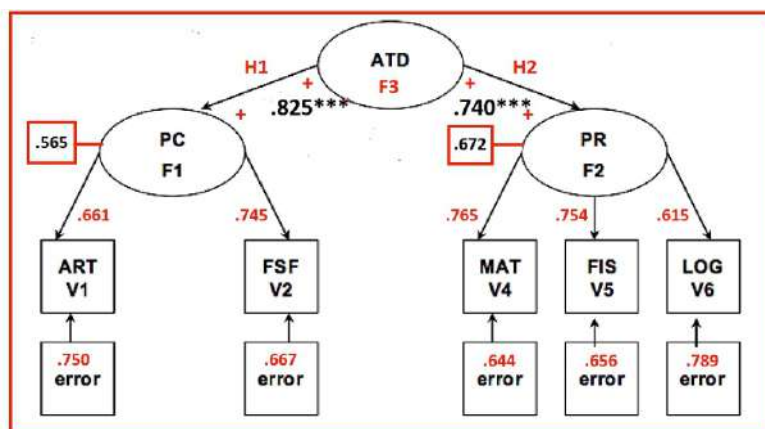
STANDARDIZED SOLUTION:                                     R-SQUARED

ART  =V1  =  .661 F1    + .750 E1                        .438
FSF  =V2  =  .745*F1   + .667 E2                        .555
MAT  =V4  =  .765 F2    + .644 E4                        .585
FIS  =V5  =  .754*F2   + .656 E5                        .569
LOG  =V6  =  .615*F2   + .789 E6                        .378
F1   =F1  =  .825*F3   + .565 D1                        .680
F2   =F2  =  .740*F3   + .672 D2                        .548

```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 53. Constructo o modelo ajustado de la asertividad en la toma de decisiones (ATD) para el análisis estructural con valores de carga factorial del SEM.



Fuente: Elaboración propia.

El resultado del modelo estructural, nos indica que: “para medir la ATD, aporta más en su explicación, el PC que el PR”. Si se decide hacer inversiones al respecto, equivaldría a decir, sólo a manera ejemplificativa y guardando las debidas proporciones, que: “del 100% de recursos, invertir aprox. un 60% a PC y 40% a PR. De lo recursos del 60% de PC y tomando en cuenta la figura 53 asignar un 60% a FSF y un 40% a ART. Del 40% de PR, asignar 30% a MAT, 30% a FIS y 40% de LOG; la r^2 mostradas indican una alta correlación entre cada una las variables y factores del modelo”. Todo esto depende del marco teórico planteado sobre la trealidad estudiada, así como de la experiencia del investigador para evidenciarlo.

Es interesante reflexionar que los estudios SEM, basados en múltiples regresiones permiten hacer este tipo de planteamientos incluyendo el estimado del error, que no son posibles determinar si se hicieran por separado para posteriormente conjuntarlas.

SEM Y EQS. Escribiendo la comprobación de hipótesis

Con el fin de reportar la comprobación de hipótesis con base en los resultados mostrados en la figura 53, se sugiere seguir, el formato mostrado en la figura 54 más que los gráficos por cuestiones de espacio, en el escrito de un artículo. (En algunas editoriales, solicitan entre 20-30 páginas; arial 10, 1, 1.5 o doble espacio o 6000-8000 palabras y/o 150 palabras de resumen, 15 referencias con opción dejar correo para reenvío de las restantes).

FIGURA 54. Reporte: tabla del modelo de ecuaciones estructurales.

Hipótesis	Relación estructural	Coefficiente estandarizado	Valor t robusto
H1: "A mayor ATD, mayor PC"	ATD->PC	.825***	4.164
H2: "A mayor ATD, mayor PR"	ATD->PR	.740***	9.050
Chi-cuadrado= 3.562; gl=2; p= .16849; NFI=.989; NNFI=.976; CFI=.995; RMSEA=.053. Nota: aún hay problemas con $p = .16849 > .05$			

Fuente: EQS 6.2, con datos y adaptación propios.

Conclusión

Una posible redacción de explicación de la figura 54, de tener $p < 0.05$, sería: "Con respecto a la hipótesis H1, los resultados obtenidos (Coeficiente estandarizado = 0.825, $p < 0.001$), indican que la ATD tiene efectos significativos en el PC de los ejecutivos de la industria del software. En cuanto a la hipótesis H2, los resultados obtenidos (coeficiente estandarizado = 0.740, $p < 0.001$), indican la ATD tiene efectos significativos en el PR de los ejecutivos de la industria del software. En resumen, se puede comprobar que los dos factores analizados tienen significancia ($p < 0.001$). La discusión tendrá que versar sobre cómo justificar los hallazgos encontrados empíricamente, el marco teórico y la realidad estudiada.

CFA. Modelo de segundo orden.

Análisis de caso de estudio 2

Tomemos como ejemplo el concepto de "orientación al mercado" y la escala desarrollada por Narver y Slater (1990) para su medición (figura 55). La orientación al mercado es "latente" en cuanto que no es manifiesto, esto es, directamente observable. Además, la orientación al mercado es también "variable" en cuanto que no es constante, es decir, no es fija en el tiempo. Una escala de medición de la orientación al mercado, buscará estimar la magnitud de esta variable latente en el momento y entorno geográfico actual, lo que denominamos, puntuación verdadera.

La noción de variable latente, implica que debe existir cierta relación entre ella y los ítems que componen la escala. La variable latente debe ser, si la escala se ha desarrollado correctamente, la causa de la puntuación obtenida en los distintos ítems de la escala. Existirá una correlación entre la puntuación dada al ítem "X1 nos preocupamos de responder a las exigencias de nuestros clientes" y la orientación al

FIGURA 55. Escala orientación al mercado de Narver y Slater, 1990.

<p><u>ORIENTACIÓN AL CLIENTE</u></p> <p>X1. Nos preocupamos por responder a las exigencias de los clientes. X2. Las acciones de mi empresa van dirigidas a que el cliente obtenga más por el mismo precio. X3. Comprendemos las necesidades de los clientes. X4. Nos fijamos objetivos de satisfacción del cliente. X5. Medimos el grado de satisfacción del cliente. X6. Ofrecemos servicios post-venta.</p> <p><u>ORIENTACIÓN A LA COMPETENCIA</u></p> <p>X7. Poseemos información sobre la cuota de mercado de la competencia. X8. El personal de ventas regularmente comparte información con nuestro negocio en relación a la estrategia de los competidores. X9. Damos una respuesta rápida a las acciones de la competencia. X10. La alta dirección efectúa análisis de las estrategias de la competencia. X11. Vemos como ventajas competitivas las oportunidades de mercado</p> <p><u>COORDINACIÓN INTERFUNCIONAL</u></p> <p>X12. Los distintos departamentos tienen contacto con los clientes. X13. Los departamentos comparten entre sí información relevante relativa a clientes o mercados. X14. Los distintos departamentos integran sus funciones en el marco de una estrategia única general. X15. Los distintos departamentos buscan la satisfacción de los clientes. X16. Todos los departamentos realizan su contribución para que el cliente reciba más sin pagar más. X17. Los departamentos de la empresa efectúan acciones de manera conjunta.</p>
--

Fuente: Narver y Slater (1990).

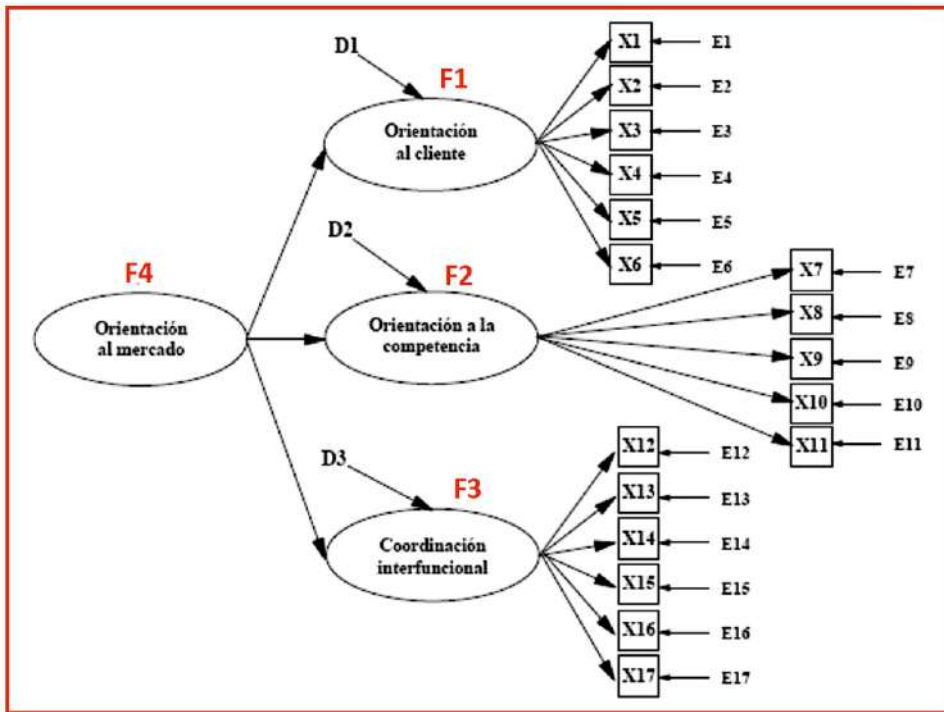
mercado de la empresa que responde, aunque no pueda ser calculado al no ser este concepto directamente medible. La forma más común de representar las relaciones entre los distintos ítems de una escala y el concepto que miden, son los *diagramas de trayectoria* (figura 56).

Recuerde que en un CFA de segundo orden, los factores de primer orden al ser observables se miden de manera directa mientras que los de segundo orden se miden de manera indirecta. Dado lo anterior, al ser el factor de segundo orden una variable independiente (Orientación al mercado. F4) no se considera con errores, sin embargo, los de primer orden como orientación al cliente (F1); Orientación a la competencia (F2) y Coordinación interfuncional (F3), tendrán cada uno su asignación de error (D1 a D3).

CFA, SEM y EQS. Entrada de datos al programa

Se sugiere seguir lo recomendado en la figura 11, accedando el archivo de entrada de programa (*.EQS), llamado cfa om.sav (formato SPSS), el cual, al ser abierto, cambia su formato original *.SAV a *.EES (cfa aom.ess) para su despliegue y manipulación (figura 57).

FIGURA 56. CFA. Diagrama de trayectoria del constructo orientación al mercado.



Fuente: Narver y Slater (1990); Byrne (2006), con adaptación propia.

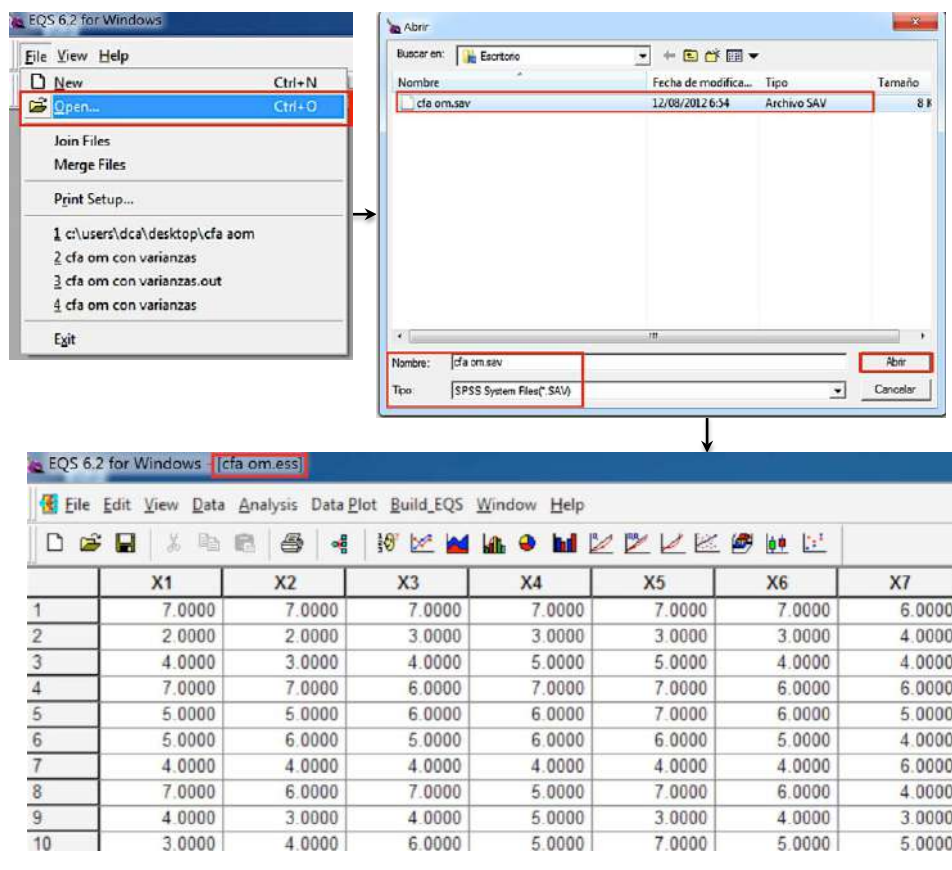
donde:

- F1. Orientación al cliente;
- F2. Orientación a la competencia;
- F3. Coordinación interfuncional;
- F4. Orientación al mercado.

Así también es posible verificar el archivo de programa de entrada EQS, del que resaltamos el nombre del archivo en el párrafo /TITLE: “CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN”. En el párrafo: /SPECIFICATIONS, la declarativa de analizar 17 variables con 300 casos.

NOTA: EQS, para el caso de sistema operativo WINDOWS debe especificarse, en un archivo externo, la trayectoria del archivo de programa de entrada; por ejemplo: DATAFILE='E://cfa om.eqsl'. Si es IOS, sólo se menciona el nombre del archivo dentro

FIGURA 57. EQS. Selección de archivo externo. Caso de estudio 2. Cambio de formato SPSS (*.SAV) a EQS (*.ESS).

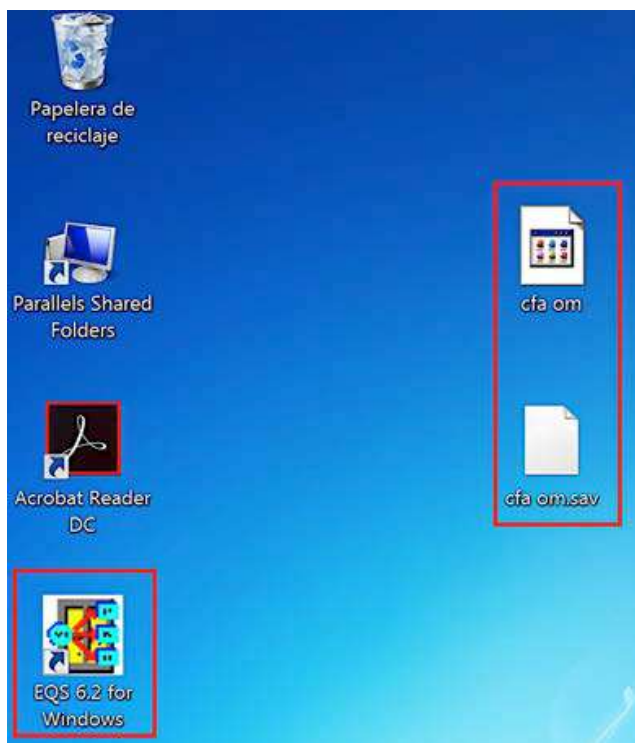


Fuente: EQS 6.2, con datos y adaptación propios.

de la computadora a nivel de escritorio con los archivos apareciendo en el emulador de Windows (Parallels, en nuestro caso: DATAFILE='cfa_om.eqs'). (figura 58).

Otra instrucción es MATRIX=RAW, que indica que la matriz de datos será generada por el mismo software. La instrucción METHOD=ML que señala que se aplicará máxima verosimilitud (ML, maximum likelihood), la instrucción ROBUST en la que se da el indicativo que los datos tienen filtros para garantizar mínimos errores (no se recomienda trabajar con los datos originales) y finalmente, la

FIGURA 58 Vista de archivos desde emulador Windows (parallels). Manejo de EQS con IOS.



Fuente: EQS 6.2, con datos y adaptación propios.

instrucción ANALYSIS= COV, sobre la aplicación de covarianzas (es mejor a la matriz de correlaciones).

En el párrafo /EQUATIONS, se tienen las ecuaciones de todas y cada una de las variables con sus errores señalando con la supresión de asterisco (*) al menos una variable que tenga varianza 1. Observe en particular las expresiones que involucran a los constructos o factores que identifican al modelo de segundo orden: $F1=*F4+D1$; $F2=*F4+D2$; $F3=*F4+D3$, en las que al incluir al F4 en las mismas, se da la idea de que se correrá al mismo tiempo el CFA con el SEM, situación que es válida y que facilita el análisis, dado lo complejo del modelo.

Nota: ¿Qué sucedería si se introdujera otro constructo de segundo orden, por ejemplo, innovación (F8) con tres factores como variables dependientes como proceso (F5), producto (F6) y tecnología (F7), que se relacionen con el modelo de

FIGURA 59. EQS. Codificación de entrada de datos. Caso de estudio 2.

```

/TITLE
CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.
NOMBRE DEL ARCHIVO: cfa om.eqs
/SPECIFICATIONS
VARIABLES=17; CASES=300; DATAFILE='cfa om.ess';
MATRIX=RAW; METHOD=ML, ROBUST; ANALYSIS=COV;

/EQUATIONS
V1=F1+E1;
V2=*F1+E2;
V3=*F1+E3;
V4=*F1+E4;
V5=*F1+E5;
V6=*F1+E6;

V7=F2+E7;
V8=*F2+E8;
V9=*F2+E9;
V10=*F2+E10;
V11=*F2+E11;

V12=F3+E12;
V13=*F3+E13;
V14=*F3+E14;
V15=*F3+E15;
V16=*F3+E16;
V17=*F3+E17;

F1=*F4+D1;
F2=*F4+D2;
F3=*F4+D3;

/VARIANCES
E1=*;
E2=*;
E3=*;
E4=*;
E5=*;
E6=*;
E7=*;
E8=*;
E9=*;
E10=*;
E11=*;
E12=*;
E13=*;
E14=*;
E15=*;
E16=*;
E17=*;

F4=1;
D1 TO D3=*;

/LMTEST

/PRINT
FIT=ALL;

/END

```

Fuente: EQS 6.2, con datos propios..

orientación de mercado (F4)? Se relacionaría con una flecha de F4-F8, se generaría hipótesis inicial de “a mayor orientación del mercado (F4), mayor innovación (F8)”. Se asignarían D5 a D7 para calcular a F5 a F7. Para su análisis dentro del CFA EQS, se deberían agregar, además, la relación de F5-F8; F6-F8; F7-F8, así como los ítems que involucran a cada factor, numerados a partir del 18 en adelante (17 contenidos previamente hasta orientación de mercado (F4). Hay modelos de tres factores de segundo orden, por ejemplo, competitividad (F12) con factores de primer orden (F9-F10-F11) con D9 a D11 con tres factores, muy complejos; se generarían hipótesis tales como (dependiendo el sentido de las flechas): “a mayor orientación de mercado mayor competitividad” y “a mayor innovación, mayor competitividad”. Por lo anterior, se recomienda hacer estudios de primer orden, a partir de calcular promedios (sólo en regresiones lineales) para calcular a F4; sin embargo, resta información importante. Por cierto, un ítem sólo abona a un factor de primer orden y de forma indirecta a un factor de segundo orden; nunca un ítem se propone a dos factores al mismo tiempo. Cabe anotar que si se llegara el caso de tener más de 1 000 muestras bajo estudio, en regresión lineal sería mucho más

FIGURA 60. CFA, SEM y EQS. Vista del programa de salida.

```

EQS, A STRUCTURAL EQUATION PROGRAM                                MULTIVARIATE SOFTWARE, INC.
COPYRIGHT BY P.M. BENTLER                                         VERSION 6.2 (C) 1985 - 2012 (B101).

PROGRAM CONTROL INFORMATION

1 /TITLE
2 CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.
3 NOMBRE DEL ARCHIVO: cfa om.eqs
4 /SPECIFICATIONS
5 VARIABLES=17; CASES=300; DATAFILE='cfa om.ess';
6 MATRIX=RAW; METHOD=ML, ROBUST; ANALYSIS=COV;
7
8 /EQUATIONS
9 V1=F1+E1;
10 V2=*F1+E2;
11 V3=*F1+E3;
12 V4=*F1+E4;
13 V5=*F1+E5;
14 V6=*F1+E6;
15
16 V7=F2+E7;
17 V8=*F2+E8;
18 V9=*F2+E9;
19 V10=*F2+E10;
20 V11=*F2+E11;

```

Fuente: EQS 6.2, con datos y adaptación propios.

sensible que en un estudio de segundo orden, es decir, todas las relaciones son altamente significativas. En los estudios factoriales esto no afecta ya que realiza cientos de regresiones lineales al mismo tiempo. En el párrafo de /VARIANCES se asigna a $F4=1$; dado que no se puede medir la variable independiente $F4$, y $D1$ TO $D3=*F4$, por su posicionamiento como parte de las variables dependientes de los factores $F1$ a $F3$. No se incluye el párrafo /COVARIANCES (ya que sólo corren como variables independientes). Es importante hacer notar que tanto los nombres de la base de datos proveniente de SPSS (.SAV), tanto el nombre de variables y su aparición sea estrictamente como corresponde al modelo teórico, con lectura de izquierda a derecha, ya que esto es una condición necesaria para que el modelo EQS, pueda correr correctamente y/o subdividir las bases de datos. En otras palabras, debe haber correspondencia en nombre y orden de aparición de las variables de la base

FIGURA 61. CFA, SEM y EQS. Reporte de estadísticos univariados, curtosis multivariada y curtosis elíptica. Caso de estudio 2

30-OCT-17 PAGE: 3 EQS Licensee:
TITLE: CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.

SAMPLE STATISTICS BASED ON COMPLETE CASES

UNIVARIATE STATISTICS

VARIABLE	X1 V1	X2 V2	X3 V3	X4 V4	X5 V5
MEAN	5.4300	5.1267	5.4200	4.9800	5.1233
SKEWNESS (G1)	-1.0011	-0.7650	-0.8937	-0.5964	-0.7291
KURTOSIS (G2)	1.0529	0.0403	0.9292	-0.1302	0.3503
STANDARD DEV.	1.3801	1.4690	1.3023	1.5560	1.4840

MULTIVARIATE KURTOSIS

MARDIA'S COEFFICIENT (G2, P) = 104.7620
NORMALIZED ESTIMATE = 35.6959

ELLIPTICAL THEORY KURTOSIS ESTIMATES

MARDIA-BASED KAPPA = 0.3243 MEAN SCALED UNIVARIATE KURTOSIS = 0.0334
MARDIA-BASED KAPPA IS USED IN COMPUTATION. KAPPA= 0.3243

CASE NUMBERS WITH LARGEST CONTRIBUTION TO NORMALIZED MULTIVARIATE KURTOSIS:

CASE NUMBER	158	172	191	206	223
ESTIMATE	997.7348	889.0390	1818.8817	701.6374	1114.6302

Fuente: EQS 6.2, con datos y adaptación propios.

de datos, con el modelo teórico, con el fin de que EQS ejecute correctamente las instrucciones con las variables correspondientes y produzca los resultados esperados (figura 59).

Realizado lo anterior, ejecute el programa EQS, siguiendo lo mostrado en las instrucciones del apartado CFA y EQS. Cómo ejecutar el programa.

CFA, SEM y EQS. Análisis de resultados

Se procede a revisar y analizar las diversas secciones de reporte que se presentan en el archivo: cfa om.out. Observe que contienen los datos del programa de entrada (cfa om.eqs), entrada a diferencia de que se encuentran numeradas de forma consecutiva y se anexa el reporte de resultados por secciones en el archivo de salida: cfa om.out (figura 60).

Así, la primera sección de reporte, describe a los estadísticos univariados (univariate statistics), que muestra la media, la asimetría, la curtosis, y la desviación estándar variable por variable; asimismo, se presenta el resultado de la curtosis multivariados (multivariate kurtosis) que, por cierto, a través del coeficiente de Mardia (Mardia's coefficient) mientras más grande se presente, da el indicativo de problemas de normalidad. Sin embargo, esto no es de preocupar ya que el sistema EQS, resuelve y aporta los resultados a nivel de solución normalizada o estandarizada correspondiente y a consideración del investigador, en el proceso de generación de la matriz de datos a analizar. Así también se reportan los números de casos y estimaciones (figura 61).

FIGURA 62. CFA, SEM y EQS. Matriz de covarianza. Caso de estudio 2.

29-OCT-17 PAGE: 4 EQS Licensee:
TITLE: CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.

COVARIANCE MATRIX TO BE ANALYZED: 17 VARIABLES (SELECTED FROM 17 VARIABLES)
BASED ON 300 CASES.

		X1	X2	X3	X4	X5
		V1	V2	V3	V4	V5
X1	V1	1.905				
X2	V2	1.256	2.158			
X3	V3	1.096	1.171	1.696		
X4	V4	0.708	1.518	0.821	2.421	
X5	V5	0.559	1.078	0.811	1.350	2.202
X6	V6	0.807	1.311	0.942	1.511	1.345
X7	V7	0.749	1.108	0.819	1.129	0.996
X8	V8	0.670	1.214	0.775	1.406	1.102
X9	V9	0.626	1.239	0.922	1.552	1.294
X10	V10	0.612	1.245	0.757	1.440	1.104
X11	V11	0.448	0.728	0.857	0.858	0.752
X12	V12	0.508	0.970	0.603	1.074	0.914
X13	V13	0.645	1.298	0.844	1.323	1.161
X14	V14	0.338	0.883	0.570	1.085	0.897
X15	V15	0.543	1.175	0.535	1.392	0.904
X16	V16	0.360	0.903	0.515	1.116	0.947
X17	V17	0.459	1.017	0.421	1.191	1.068

Fuente: EQS 6.2, con datos y adaptación propios.

La siguiente sección de reporte, llamada matriz de covarianza (covariance matrix), en la que se observan las 17 variables analizadas de los 300 casos recopilados en la base de datos, que se construye en el proceso EQS (figura 62).

Se observa la subsección de reporte estructural de Bentler-Weeks, que presenta la cantidad de variables dependientes e independientes, parámetros libres y parámetros fijos $<>0$ (figura 63).

Se observa la sección de reporte de matriz de covarianzas residuales (*residual covariance matrix*) y estandarizada (*standardized residual matrix*). Sólo son informativos (figura 64).

Al revisar la sección de reporte del gráfico de distribución estandarizada, observamos que hay tres datos desviados a la izquierda de la media y uno desviado a la derecha de la media (figura 65).

La sección de reporte de índices de ajuste (*fit indices*) reporta que se usó el método de máxima verosimilitud (ML. *maximum likelihood*) para proponer la solución normal. Así también, se observa que el resultado de Chi-cuadrado= 554.577, con $gl= 116$, valor de $p \leq 0.01$ (altamente significativo). Así también, los índices de ajuste $NFI=.853$; $NNFI=.859$ y $CFI=.880$, cumplen si se considerara que

FIGURA 63. SEM y EQS. Reporte estructural de Bentler-Weeks.

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 20

DEPENDENT V'S :	1	2	3	4	5	6	7	8	9	10
DEPENDENT V'S :	11	12	13	14	15	16	17			
DEPENDENT F'S :	1	2	3							

NUMBER OF INDEPENDENT VARIABLES = 21

INDEPENDENT F'S :	4									
INDEPENDENT E'S :	1	2	3	4	5	6	7	8	9	10
INDEPENDENT E'S :	11	12	13	14	15	16	17			
INDEPENDENT D'S :	1	2	3							

NUMBER OF FREE PARAMETERS = 37

NUMBER OF FIXED NONZERO PARAMETERS = 24

*** WARNING MESSAGES ABOVE, IF ANY, REFER TO THE MODEL PROVIDED.
CALCULATIONS FOR INDEPENDENCE MODEL NOW BEGIN.

*** WARNING MESSAGES ABOVE, IF ANY, REFER TO INDEPENDENCE MODEL.
CALCULATIONS FOR USER'S MODEL NOW BEGIN.

3RD STAGE OF COMPUTATION REQUIRED
PROGRAM ALLOCATED 200000000 WORDS

300104 WORDS OF MEMORY.

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 64. CFA, SEM y EQS. Matriz de covarianza residual generada. Caso de estudio 2.

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

		X1 V1	X2 V2	X3 V3	X4 V4	X5 V5
X1	V1	0.000				
X2	V2	0.354	0.000			
X3	V3	0.475	0.179	0.000		
X4	V4	-0.235	0.014	-0.215	0.000	
X5	V5	-0.217	-0.160	-0.042	0.057	0.000
X6	V6	-0.075	-0.096	-0.027	0.040	0.135
X7	V7	0.031	-0.039	0.029	-0.070	0.010
X8	V8	-0.139	-0.077	-0.113	0.057	-0.008
X9	V9	-0.264	-0.182	-0.057	0.067	0.072
X10	V10	-0.135	0.053	-0.064	0.195	0.079
X11	V11	-0.037	-0.045	0.324	0.050	0.087
X12	V12	-0.090	0.017	-0.054	0.078	0.094
X13	V13	-0.125	0.070	-0.002	0.039	0.104
X14	V14	-0.271	-0.089	-0.100	0.070	0.061
X15	V15	-0.163	0.047	-0.242	0.213	-0.066
X16	V16	-0.245	-0.063	-0.150	0.107	0.116
X17	V17	-0.176	0.004	-0.277	0.132	0.196

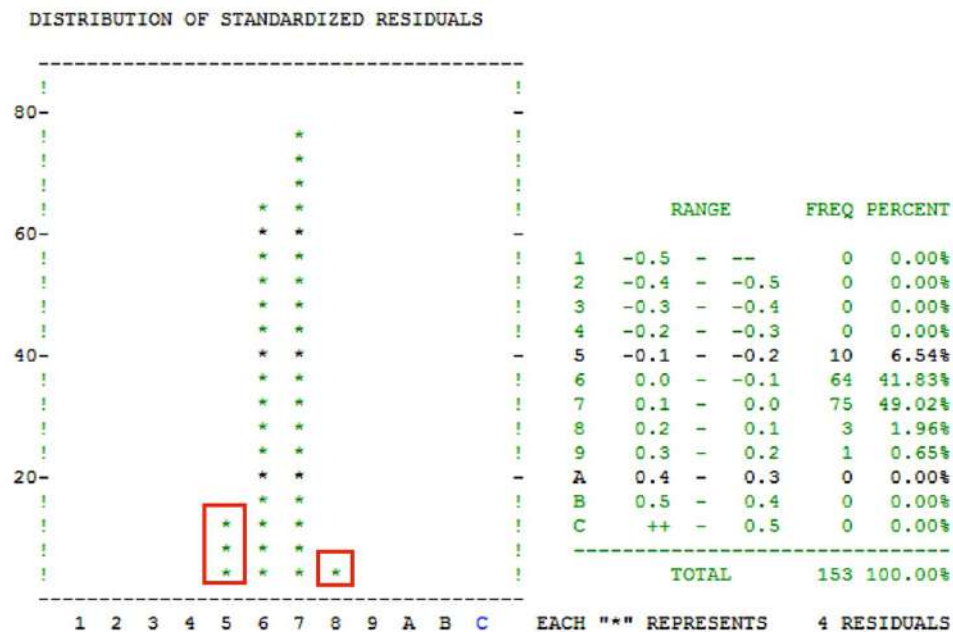
Fuente: EQS 6.2, con datos y adaptación propios.

STANDARDIZED RESIDUAL MATRIX:

		X1 V1	X2 V2	X3 V3	X4 V4	X5 V5
X1	V1	0.000				
X2	V2	0.175	0.000			
X3	V3	0.264	0.094	0.000		
X4	V4	-0.109	0.006	-0.106	0.000	
X5	V5	-0.106	-0.073	-0.022	0.025	0.000
X6	V6	-0.037	-0.045	-0.014	0.018	0.062
X7	V7	0.015	-0.018	0.015	-0.029	0.004
X8	V8	-0.064	-0.033	-0.055	0.023	-0.004
X9	V9	-0.113	-0.073	-0.026	0.025	0.029
X10	V10	-0.063	0.023	-0.032	0.081	0.034
X11	V11	-0.020	-0.023	0.181	0.023	0.042
X12	V12	-0.046	0.008	-0.029	0.036	0.045
X13	V13	-0.058	0.031	-0.001	0.016	0.045
X14	V14	-0.146	-0.045	-0.057	0.033	0.030
X15	V15	-0.079	0.021	-0.123	0.091	-0.030
X16	V16	-0.133	-0.032	-0.087	0.051	0.059
X17	V17	-0.091	0.002	-0.151	0.061	0.094

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 65. CFA, SEM y EQS. Gráfico distribución estandarizada. Caso de estudio 2.



Fuente: EQS 6.2, con datos y adaptación propios.

esta escala ha sido muy utilizada de manera efectiva. Recordar que para casos de nuevas escalas, los índices de ajuste deben ser $>.9$. Por otro lado, el $RMSEA = 0.112 > 0.08$, no cumple el requerimiento. Por último, el alfa de Cronbach = $.943$, no es de interés a nuestro estudio ya que es el del modelo en general, EQS, no lo reporta a nivel de factor o variable de manera individual, que es lo que se necesita reportar (figura 66).

Más adelante se encuentra la sección de reporte de índices de ajuste (fit indices), con el método robusto (robust). Esta sección es la que se sugiere tomar en cuenta para los análisis ya que presentan mínimos errores. Así, por ejemplo, el valor Chi-cuadrado de Satorra-Bentler = 369.7671 ; $gl = 116$; valor $p < .01$ ((Nota: $369.77671 / 116 = 3.18 > 2$ Tendencia a que los índices de ajuste sean $<.9$), cumple con el requerimiento de significatividad. En el reporte, los índices de ajuste $NFI = .851$; $NNFI = .874$ y $CFI = .892$, a pesar de ser $>.8$, el promedio = $(.851 + .874 + .892) / 3 = .872 < .9$, no cumple para la creación de una escala nueva. Si embargo, $RMSEA = .086 > .08$, sí cumple como indicador. Cabe señalar, que en la sección de método robusto (robust), NO se encuentran presentes los índices de

ajuste GFI, AGFI; SRMR (sólo en los normales). Se considera así, dado que Satorra y Bentler, creadores de EQS y no considera otros (que aparecen por ejemplo en LISREL, donde Joreskog y Sorbom son los creadores), (figura 67).

Como ya es de su conocimiento, las siguientes secciones de reporte, se deberán revisar que cumplan con sus condiciones de aplicabilidad. Por ejemplo, en la sección de medición de ecuaciones (*measurement equations*) en la parte de variables, se debe revisar que los valores de t sean significativas. En las 17 variables cumple. (figura 68).

Al revisar la sección de reporte de ecuaciones de constructo (*construct equations*), en su relación entre los factores, debe cumplir con valores t significativos. En este caso cumple en la relación de F1-F4; F2-F4; F3-F4 (figura 69).

Un dato importante, normalmente a verificar es la covarianza entre variables independientes (*covariance among independent variables*), que en este caso de estudio NO se presentan ya que es sólo existe una variable independiente F4.

FIGURA 66. CFA, SEM y EQS. Reporte de índices de ajuste con valores normales. Caso de estudio 2.

```

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE          =   3778.176 ON   136 DEGREES OF FREEDOM

INDEPENDENCE AIC =   3506.176   INDEPENDENCE CAIC =   2866.461
MODEL AIC =       322.577       MODEL CAIC =       -223.061

CHI-SQUARE =   554.577 BASED ON   116 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS   0.00000

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS   577.028.

FIT INDICES
-----
BENTLER-BONETT      NORMED FIT INDEX =   0.853
BENTLER-BONETT NON-NORMED FIT INDEX =   0.859
COMPARATIVE FIT INDEX (CFI) =   0.880
BOLLEN'S           (IFI) FIT INDEX =   0.880
MCDONALD'S        (MFI) FIT INDEX =   0.481
JORESKOG-SORBOM'S GFI FIT INDEX =   0.815
JORESKOG-SORBOM'S AGFI FIT INDEX =   0.756
ROOT MEAN-SQUARE RESIDUAL (RMR) =   0.122
STANDARDIZED RMR =   0.060
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) =   0.112
90% CONFIDENCE INTERVAL OF RMSEA (   0.103,   0.122)

RELIABILITY COEFFICIENTS
-----
CRONBACH'S ALPHA =   0.943
RELIABILITY COEFFICIENT RHO =   0.956

```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 67. CFA, SEM y EQS. Reporte de índices de ajuste valores robustos. Caso de estudio 2.

```

GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

ROBUST INDEPENDENCE MODEL CHI-SQUARE = 2488.294 ON 136 DEGREES OF FREEDOM

INDEPENDENCE AIC = 2216.294 INDEPENDENCE CAIC = 1576.580
MODEL AIC = 137.767 MODEL CAIC = -407.872

SATORRA-BENTLER SCALED CHI-SQUARE = 369.7671 ON 116 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

MEAN- AND VARIANCE-ADJUSTED CHI-SQUARE = 91.624 ON 29 D.F.
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

MEAN-SCALED AND SKEWNESS-ADJUSTED STATISTIC = 21.143 ON 7 D.F.
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00356

RESIDUAL-BASED TEST STATISTIC = 531.896
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC = 190.995
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00001

YUAN-BENTLER RESIDUAL-BASED F-STATISTIC = 2.822
DEGREES OF FREEDOM = 116, 184
PROBABILITY VALUE FOR THE F-STATISTIC IS 0.00000

```

```

FIT INDICES
-----
BENTLER-BONETT NORMED FIT INDEX = 0.851
BENTLER-BONETT NON-NORMED FIT INDEX = 0.874
COMPARATIVE FIT INDEX (CFI) = 0.892
BOLLEN'S (IFI) FIT INDEX = 0.893
MCDONALD'S (RFI) FIT INDEX = 0.655
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.086
90% CONFIDENCE INTERVAL OF RMSEA ( 0.076, 0.095)

```

Fuente: EQS 6.2, con datos y adaptación propios.

Avanzando en el análisis de los reportes, encontramos el de soluciones estandarizadas (standardized solutions), del cual se realiza la pregunta: ¿son todas las cargas factoriales $>.6$? Para nuestro caso de estudio, se observa que NO cumple en $V1 = .545$ y $r^2 = .297$, lo que significa que este ítem aporta muy poco al modelo de orientación al mercado. Además, si se observan las cargas factoriales $F1 = .898$; $F2 = .858$; $F3 = .868$, y sus r^2 de: $.807$; $.736$; $.753$, son muy similares, no tienen una diferencia notable, lo que se traduce que las tres orientaciones son igual de importantes para explicar el modelo de orientación al mercado (figura 70).

FIGURA 68. CFA, SEM y EQS. Medición de ecuaciones de medición. Caso de estudio 2.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
 (ROBUST STATISTICS IN PARENTHESES)

```

X1  =V1 = 1.000 F1 + 1.000 E1

X2  =V2 = 1.596*F1 + 1.000 E2
      .166
      9.626@
      (.218)
      ( 7.320@)

X3  =V3 = 1.099*F1 + 1.000 E3
      .132
      8.335@
      (.132)
      ( 8.304@)

X4  =V4 = 1.668*F1 + 1.000 E4
      .174
      9.561@
      (.283)
      ( 5.903@)
    
```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 69. CFA, SEM y EQS. Medición de ecuaciones de constructo. Caso de estudio 2.

```

30-OCT-17 PAGE: 10 EQS Licensee:
TITLE: CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)
CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
(ROBUST STATISTICS IN PARENTHESES)

F1  =F1 = .675*F4 + 1.000 D1
      .074
      9.181@
      (.097)
      ( 6.971@)

F2  =F2 = 1.064*F4 + 1.000 D2
      .077
      13.821@
      (.077)
      ( 13.802@)

F3  =F3 = .885*F4 + 1.000 D3
      .072
      12.270@
      (.077)
      ( 11.475@)
    
```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 70. CFA, SEM y EQS. EQS. Reporte solución estandarizada

```

30-OCT-17      PAGE: 14  EQS      Licensee:
TITLE:   CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

STANDARDIZED SOLUTION:                                     R-SQUARED

X1   =V1   = .545 F1   + .839 E1   .297
X2   =V2   = .817*F1  + .577 E2   .667
X3   =V3   = .634*F1  + .773 E3   .403
X4   =V4   = .806*F1  + .592 E4   .649
X5   =V5   = .695*F1  + .719 E5   .484
X6   =V6   = .806*F1  + .592 E6   .649
X7   =V7   = .819 F2  + .574 E7   .671
X8   =V8   = .880*F2  + .475 E8   .774
X9   =V9   = .908*F2  + .419 E9   .824
X10  =V10  = .833*F2  + .554 E10  .693
X11  =V11  = .609*F2  + .793 E11  .371
X12  =V12  = .727 F3  + .687 E12  .529
X13  =V13  = .846*F3  + .533 E13  .716
X14  =V14  = .771*F3  + .637 E14  .595
X15  =V15  = .802*F3  + .597 E15  .643
X16  =V16  = .776*F3  + .631 E16  .602
X17  =V17  = .772*F3  + .635 E17  .596
F1   =F1   = .898*F4  + .440 D1   .807
F2   =F2   = .858*F4  + .514 D2   .736
F3   =F3   = .868*F4  + .497 D3   .753

```

Fuente: EQS 6.2, con datos y adaptación propios.

Sobre la sección de reporte de multiplicadores de Lagrange multivariado (multivariate Lagrange multiplier), tenemos las siguientes lecturas. “que V10, debiera impactar en F3; que V1 a F4; V3 a F3; V7 a F3; V2 a F2; V7 a F4”, de hacerlo, se incrementarían las chi-cuadrados en: 18.555; 16.399; 12.466, etc. Así, se debe buscar NO aumentar las Chi-cuadrados, sino bajarlo, ya que baja el error, por lo que aumentan los índices de ajuste. Se debe buscar disminuir el error, es decir, la chi-cuadrado. Cabe hacer notar que las sugerencias del reporte multiplicador de multivariado, son sólo estadísticas, NO proceden si sugieren orientar a otro factor diferente del modelo (figura 71).

FIGURA 70. CFA, SEM y EQS. EQS. Test de los multiplicadores de Lagrange.

```

30-OCT-17      PAGE: 16  EQS  License:
TITLE:  CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.
MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1
PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:
PVV PVF PFF PDD GVV GVF GFF BVF BFF

-----
CUMULATIVE MULTIVARIATE STATISTICS      UNIVARIATE INCREMENT
-----
STEP  PARAMETER      CHI-SQUARE      D.F.  PROB.      CHI-SQUARE      PROB.
-----
1      V10,F3      18.555      1  0.000      18.555  0.000
2      V1,F4      34.954      2  0.000      16.399  0.000
3      V3,F3      47.420      3  0.000      12.466  0.000
4      V7,F3      58.171      4  0.000      10.751  0.001
5      V2,F2      66.529      5  0.000      8.358  0.004
6      V7,F4      70.665      6  0.000      4.137  0.042
-----
HANCOCK'S SEQUENTIAL      ROBUST
D.F.  PROB.      RMSEA  CFI
-----
116  1.000      0.083  0.900
115  1.000      0.080  0.906
114  1.000      0.079  0.911
113  1.000      0.077  0.915
112  1.000      0.076  0.918
111  1.000      0.076  0.920
-----
*** NOTE *** IF PREDICTED RMSEA COULD NOT BE CALCULATED, 99.999 IS PRINTED.
IF PREDICTED CFI COULD NOT BE CALCULATED, 9.999 IS PRINTED.

```

Fuente: EQS 6.2, con datos y adaptación propios.

CFA, SEM y EQS eliminando V1

Así, y dado que la variable V1= .545 (figura 72) aporta escasamente al modelo y de que se requiere además, disminuir Chi-cuadrado, para que disminuya el error, se sugiere retirar del modelo EQS a la variable V1; cambiar el asterico (*) a V2 para conservar los requerimientos de análisis de EQS de que una de las variables de cada factor debe tener varianza 1 (figura 72).

FIGURA 70. CFA, SEM y EQS. EQS. Reporte solución estandarizada

```

/TITLE
CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.SIN V1
NOMBRE DEL ARCHIVO: cfa om sin v1.eq5

/SPECIFICATIONS
VARIABLES=17; CASES=300; DATAFILE='cfa om.ess';
MATRIX=RAW; METHOD=ML, ROBUST; ANALYSIS=COV;

/EQUATIONS
!V1=F1+E1;
V2= F1+E2;
V3=*F1+E3;
V4=*F1+E4;
V5=*F1+E5;
V6=*F1+E6;

V7=F2+E7;
V8=*F2+E8;
V9=*F2+E9;
V10=*F2+E10;
V11=*F2+E11;

V12=F3+E12;
V13=*F3+E13;
V14=*F3+E14;
V15=*F3+E15;
V16=*F3+E16;
V17=*F3+E17;

F1=*F4+D1;
F2=*F4+D2;
F3=*F4+D3;

/VARIANCES
!E1=*;

/VARIANCES
E1=*;
E2=*;
E3=*;
E4=*;
E5=*;
E6=*;
E7=*;
E8=*;
E9=*;
E10=*;
E11=*;
E12=*;
E13=*;
E14=*;
E15=*;
E16=*;
E17=*;

F4=1;
D1 TO D3=*;

/LMTEST

/PRINT
FIT=ALL;

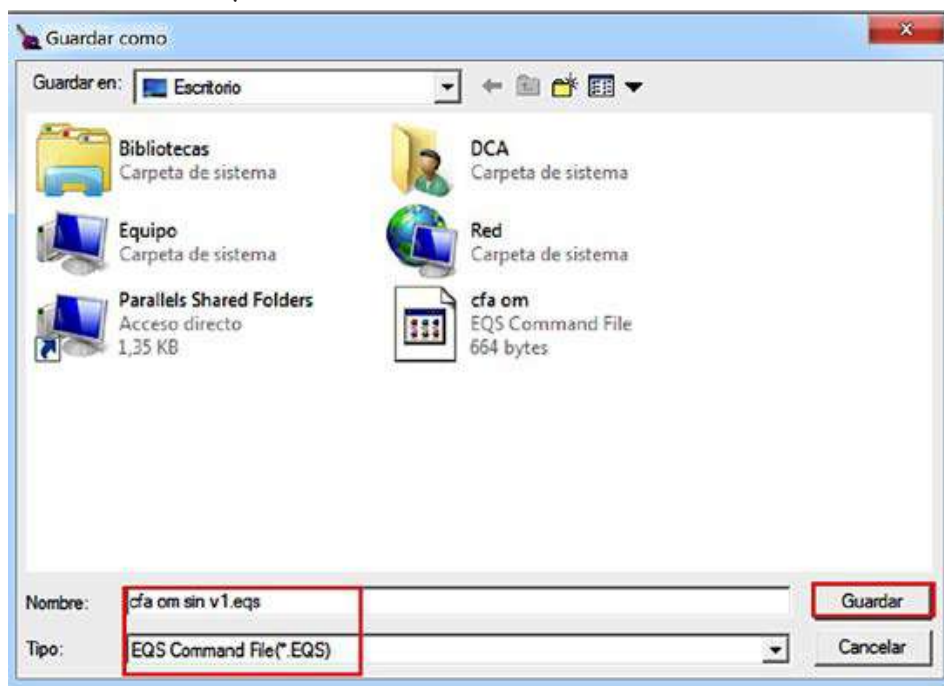
/END

```

Fuente: EQS 6.2, con datos y adaptación propios.

y realizar la ejecución del nuevo modelo, para analizar los resultados. En nuestro caso se renombra el archivo EQS como: cfa om sin V1.eqs (figura 73).

FIGURA 73. CFA, SEM y EQS. EQS. Renombrando archivo de entrada. Caso de estudio 2 sin V1.



Fuente: EQS 6.2, con datos y adaptación propios.

Se guardan el archivo y se corre conforme a lo indicado en la figura 14.

Se procede a revisar rápidamente las secciones de reporte de resultados como se ve en la figura 74.

CFA, SEM y EQS. Discusión rápida de resultados sin V1

De los reportes emitidos, podemos observar y concluir de cada sección de reporte lo siguiente:

- ▶ Gráfico de distribución de residuos (*distribution of standardized residuals*). Cumple.
- ▶ Índices de ajuste (*fit indices*) robustos (robust), valores a analizar:

FIGURA 74. CFA, SEM y EQS. EQS. Proceso de revisión rápida de resultados. Caso estudio 2 sin V1.

EQS, A STRUCTURAL EQUATION PROGRAM
COPYRIGHT BY P.M. BENTLER

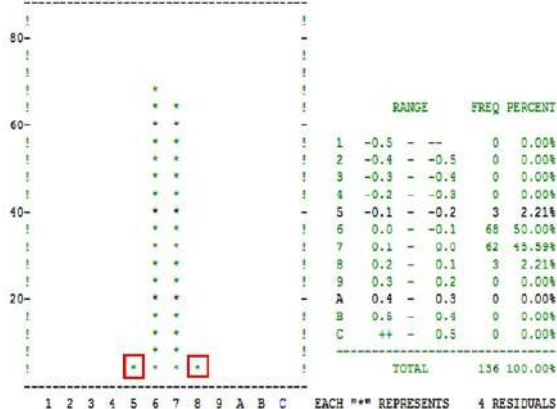
MULTIVARIATE SOFTWARE, INC.
VERSION 6.2 (C) 1985 - 2012 (B101).

PROGRAM CONTROL INFORMATION

```

1 /TITLE
2 CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.SIN V1
3 NOMBRE DEL ARCHIVO: cfa om sin v1.eq$
4 /SPECIFICATIONS
5 VARIABLES=17; CASES=300; DATAFILE='cfa om.ess';
6 MATRIX=RAW; METHOD=ML, ROBUST; ANALYSIS=COV;
7
8 /EQUATIONS
9 !V1=F1+E1;
10 V2= F1+E2;
11 V3==F1+E3;
12 V4==F1+E4;
13 V5==F1+E5;
14 V6==F1+E6;
15
16 V7=F2+E7;
17 V8==F2+E8;
18 V9==F2+E9;
19 V10==F2+E10;
20 V11==F2+E11;
21
22 V12=F3+E12;
23 V13==F3+E13;
24 V14==F3+E14;
25 V15==F3+E15;
26 V16==F3+E16;
27 V17==F3+E17;
    
```

DISTRIBUTION OF STANDARDIZED RESIDUALS



GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

ROBUST INDEPENDENCE MODEL CHI-SQUARE = 2344.224 ON 120 DEGREES OF FREEDOM
 INDEPENDENCE AIC = 2104.224 INDEPENDENCE CAIC = 1539.770
 MODEL AIC = 84.115 MODEL CAIC = -390.967

SATORRA-BENTLER SCALED CHI-SQUARE = 286.1146 ON 101 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

FIT INDICES

BENTLER-BONETT NORMED FIT INDEX = 0.878
 BENTLER-BONETT NON-NORMED FIT INDEX = 0.901
 COMPARATIVE FIT INDEX (CFI) = 0.917
 BOLLEN'S (IFI) FIT INDEX = 0.917
 MCDONALD'S (MFI) FIT INDEX = 0.735
 ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.078
 90% CONFIDENCE INTERVAL OF RMSEA (0.068, 0.089)

Fuente: EQS 6.2, con datos y adaptación propios.

Finalizando la escala. Análisis factorial confirmatorio y ecuación estructural

CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH §.
(ROBUST STATISTICS IN PARENTHESES)

F1 =F1 = 1.058*F4 + 1.000 D1
 .076
 13.890§
 (.077)
 §13.662§

F2 =F2 = 1.060*F4 + 1.000 D2
 .077
 13.814§
 (.076)
 §14.010§

F3 =F3 = .887*F4 + 1.000 D3
 .072
 12.319§
 (.077)
 §11.505§

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH §.
(ROBUST STATISTICS IN PARENTHESES)

X2 =V2 = 1.000 F1 + 1.000 E2

X3 =V3 = .673*F1 + 1.000 E3
 .064
 10.444§
 (.085)
 §7.935§

X4 =V4 = 1.109*F1 + 1.000 E4
 .073
 15.241§
 (.075)
 §14.699§

X5 =V5 = .914*F1 + 1.000 E5
 .071
 12.700§
 (.085)
 §10.753§

→

30-OCT-17 PAGE: 14 EQS License:
TITLE: CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.SIN V1
MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

STANDARDIZED SOLUTION:

X2 =V2 =	.786 F1	+ .618 E2	.618
X3 =V3 =	.697*F1	+ .902 E3	§.357§
X4 =V4 =	.823*F1	+ .568 E4	.677
X5 =V5 =	.711*F1	+ .703 E5	.506
X6 =V6 =	.813*F1	+ .353 E6	.469
X7 =V7 =	.813 F2	+ .176 E7	.668
X8 =V8 =	.879*F2	+ .476 E8	.773
X9 =V9 =	.909*F2	+ .417 E9	.826
X10 =V10 =	.833*F2	+ .353 E10	.694
X11 =V11 =	.609*F2	+ .793 E11	.370
X12 =V12 =	.726 F3	+ .687 E12	.528
X13 =V13 =	.845*F3	+ .535 E13	.714
X14 =V14 =	.772*F3	+ .638 E14	.596
X15 =V15 =	.807*F3	+ .588 E15	.643
X16 =V16 =	.777*F3	+ .630 E16	.604
X17 =V17 =	.773*F3	+ .634 E17	.598
F1 =F1 =	.817*F4	+ .460 D1	.840
F2 =F2 =	.856*F4	+ .517 D2	.733
F3 =F3 =	.870*F4	+ .453 D3	.767

R-SQUARED

↓

LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)

↓

ORDERED UNIVARIATE TEST STATISTICS:

NO	CODE	PARAMETER	CHI-SQUARE	PROB.	HANCOCK 101 DF	PARAM. CHANGE	STANDARDIZED CHANGE	ROBUST PREDICTED RMSEA	CFI
--	----	-----	-----	-----	-----	-----	-----	-----	-----
1	2 20	V10,F3	18.491	0.000	1.000	0.406	0.257	0.075	0.925
2	2 20	V7,F3	16.678	0.000	1.000	-0.387	-0.250	0.075	0.924
3	2 12	V10,F4	13.646	0.000	1.000	0.564	0.364	0.076	0.922
4	2 20	V3,F3	6.572	0.010	1.000	-0.328	-0.247	0.077	0.919
5	2 12	V7,F4	6.472	0.011	1.000	-0.389	-0.257	0.078	0.919
6	2 20	V10,F1	6.460	0.011	1.000	0.241	0.135	0.078	0.919
7	2 20	V4,F3	3.669	0.055	1.000	0.250	0.158	0.078	0.918
8	2 20	V13,F2	3.019	0.082	1.000	0.138	0.072	0.078	0.918
9	2 12	V3,F4	2.828	0.093	1.000	-0.483	-0.371	0.078	0.918
10	2 20	V6,F3	2.524	0.112	1.000	-0.195	-0.132	0.078	0.917
11	2 12	V4,F4	2.478	0.115	1.000	0.479	0.308	0.078	0.917

Fuente: EQS 6.2, con datos y adaptación propios.

Chi-cuadrado de Satorra-Bentler= 286.1146; $gI= 101$ (Nota: $286.1146/101=2.83>2$ Tendencia a que alguno de los índices de ajuste sea $<.9$); valor $p<0.01$. Cumple.

NFI= .878; NNFI= .901; CFI= .917. NO Cumple en NFI. RMSEA=.078 $<.08$. Cumple.

- ▶ Ecuaciones de medición (measurement equations), valores de $t (@) = .05$. Cumple.
- ▶ Ecuaciones de constructo (construct equations), valores de $t (@) = .05$. Cumple.
- ▶ Varianzas de las variables independientes (variances of independent variables), valores de $t (@) = .05$. Cumple.
- ▶ Solución estandarizada (standardized solutions), ¿todas las cargas factoriales son $>.6$? En $V3=.597$. Recuerde que Hair (et al., 1999) afirma que el promedio

FIGURA 75. CFA, SEM y EQS. EQS. Recodificación de entrada de datos. Caso de estudio 2 sin V1 y V3.

```

/TITLE
CFA ORIENTACIÓN AL MERCADO DE SEGUNDO ORDEN.SIN V1 y V3
NOMBRE DEL ARCHIVO: cfa om sin v1 y v3.eqs
/SPECIFICATIONS
VARIABLES=17; CASES=300; DATAFILE='cfa om.ess';
MATRIX=RAW; METHOD=ML, ROBUST; ANALYSIS=COV;

/EQUATIONS
!V1=F1+E1;
V2= F1+E2;
!V3=*F1+E3;
V4=*F1+E4;
V5=*F1+E5;
V6=*F1+E6;

V7=F2+E7;
V8=*F2+E8;
V9=*F2+E9;
V10=*F2+E10;
V11=*F2+E11;

V12=F3+E12;
V13=*F3+E13;
V14=*F3+E14;
V15=*F3+E15;
V16=*F3+E16;
V17=*F3+E17;

F1=*F4+D1;
F2=*F4+D2;
F3=*F4+D3;

/VARIANCES
!E1=*;
!E2=*;

/VARIANCES
E1=*;
E2=*;
E3=*;
E4=*;
E5=*;
E6=*;
E7=*;
E8=*;
E9=*;
E10=*;
E11=*;
E12=*;
E13=*;
E14=*;
E15=*;
E16=*;
E17=*;

F4=1;
D1 TO D3=*;

/LMTEST

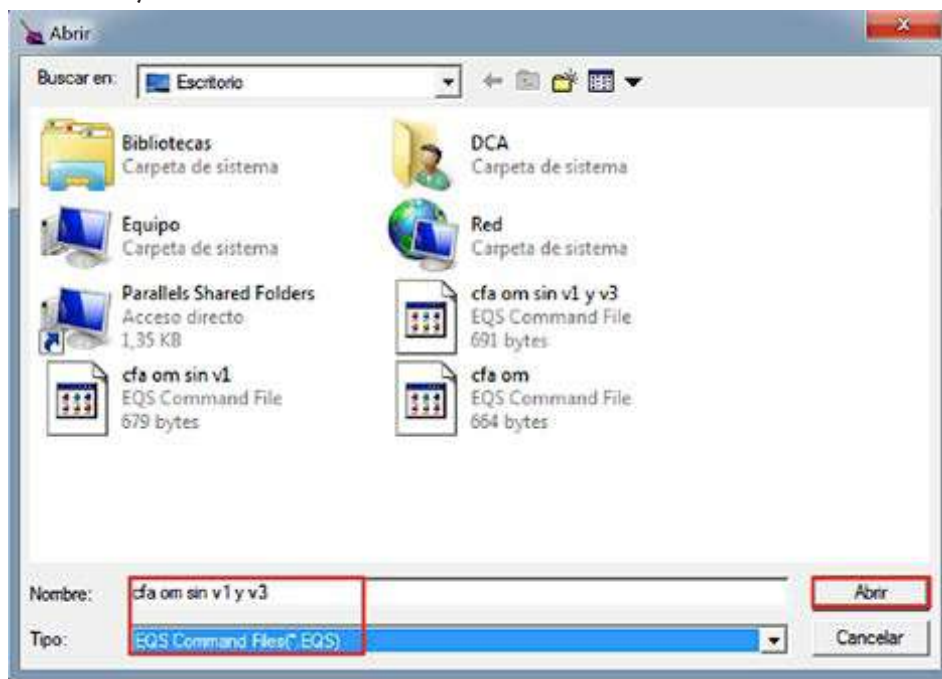
/PRINT
FIT=ALL;

/END

```

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 76. CFA, SEM y EQS. EQS. Renombrando archivo de entrada. Caso de estudio 2 sin V1 y V3.



Fuente: EQS 6.2, con datos y adaptación propios.

de todas las cargas factoriales $>.6$ vs. Bagozzi y Yi (1988) que cada carga factorial debe ser $>.6$. Tomando en cuenta este último criterio, NO cumple.

- Prueba del multiplicador de Lagrange (*Lagrange multiplier test*), se observa la importancia del ajuste estadístico del reacomodo de variables y no se justifica hacer reacomodo de variables, como sugiere la prueba del multiplicador de Lagrange, cambiar V3 a F3, NO se debe proceder si se hizo un EFA donde V3 abona a F1, a menos que tenga un sustento teórico y evidencia suficiente para explicarlo. Por lo tanto, considerar la posible supresión de variables, que tengan un Chi-cuadrado bajo para bajar el error y aumentar los índices de ajuste, como en nuestro caso NO cumple en NFI (.878). Así, se percibe que si se anula V3 con su error E3, se obtendría una mejora del índice de ajuste. Por otro lado, recuerde que sólo se recomienda suprimir un máximo de 20% de los indicadores, ya que esto da aún la posibilidad de explicar el modelo sin mayores problemas, de lo contrario, se estaría obteniendo un *modelo ateorico*.

CFA, SEM y EQS eliminando V3

Así, y dado que la variable V3= .597 (figura 76) aporta escasamente al modelo y de que se requiere además, disminuir chi-cuadrado, para que disminuya el error, se sugiere retirar del modelo EQS a la variable V3; se conserva el asterico (*) a V2 para conservar los requerimientos de análisis de EQS de que una de las variables de cada factor debe tener varianza 1 (figura 75).

Y realizar la ejecución del nuevo modelo, para analizar los resultados. En nuestro caso se renombra el archivo EQS como: cfa om sin v1 y v3.eqs (figura 76).

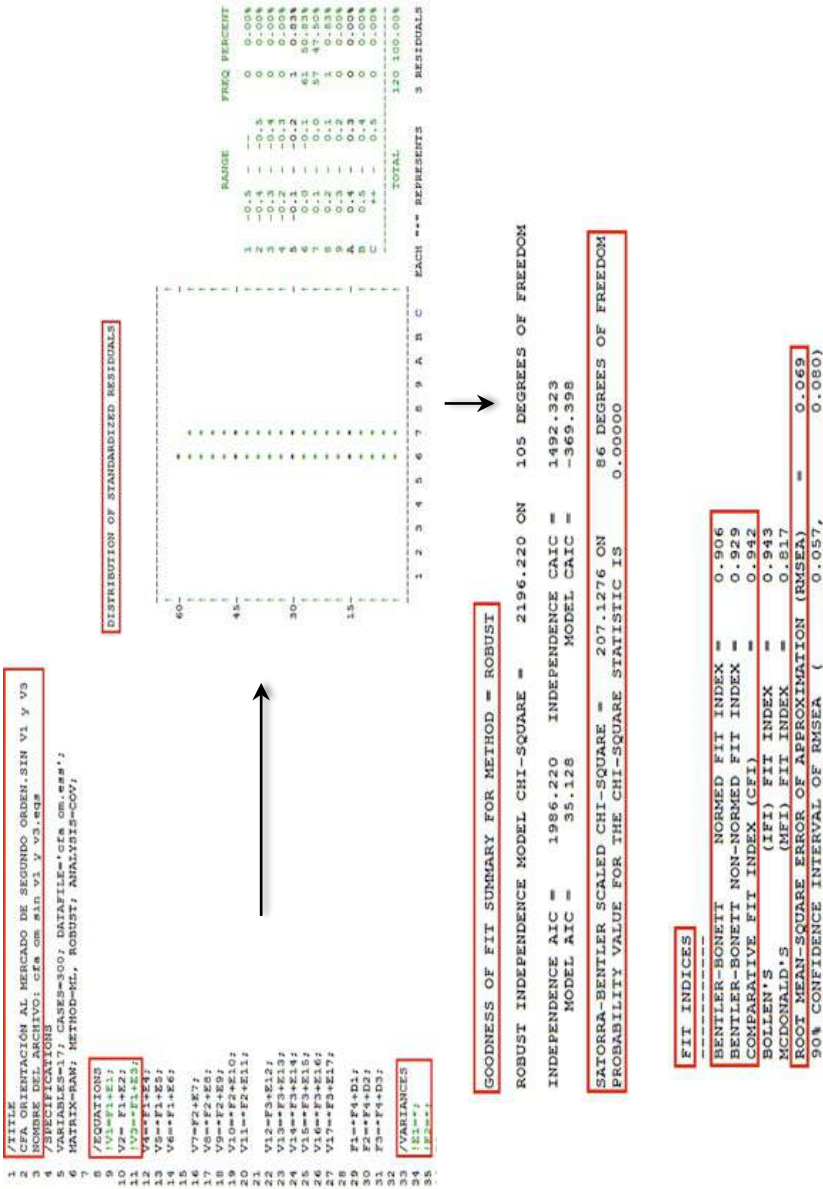
Se guardan el archivo y se corre conforme a lo indicado en la figura 14. Se procede a revisar rápidamente las secciones de reporte de resultados como se ve en la figura 77.

CFA, SEM y EQS. Discusión rápida de resultados, sin V1 y V3

De los reportes emitidos, podemos observar y concluir de cada sección de reporte lo siguiente:

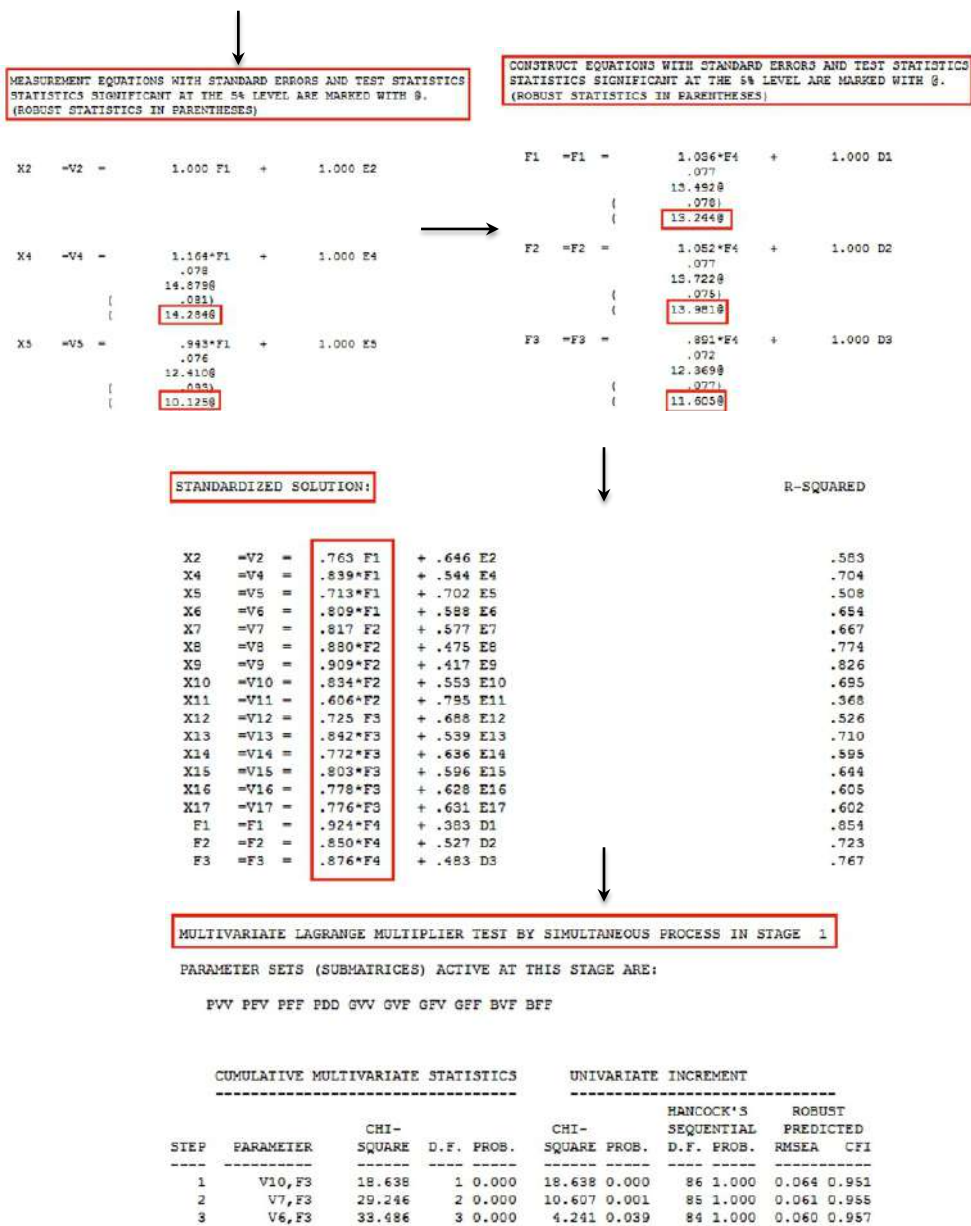
- ▶ Gráfico de distribución de residuos (distribution of standardized residuals). Cumple.
- ▶ Índices de ajuste (fit indices) robustos (robust), valores a analizar:
Chi-cuadrado de Satorra-Bentler= 207.1276; gl= 86 (Nota: $207.1276/86=0.2727 < 2$ Tendencia a que los índices de ajuste SÍ sean .9); valor $p < 0.01$. Cumple.
NFI= .906; NNFI= .929; CFI= .942. Cumple $> .9$. RMSEA= .069 $< .08$. Cumple.
- ▶ Ecuaciones de medición (measurement equations), valores de t (@) = .05. Cumple.
- ▶ Ecuaciones de constructo (construct equations), valores de t (@) = .05. Cumple.
- ▶ Varianzas de las variables independientes (variances of independent variables), valores de t (@) = .05. Cumple.
Solución estandarizada (standardized solutions), ¿todas las cargas factoriales son $> .6$? Bagozzi y Yi (1988) que cada carga factorial debe ser $> .6$. Cumple.
- ▶ Ya no es necesario verificar la prueba del multiplicador de Lagrange (Lagrange multiplier test), ya que cumplen los requisitos anteriores. Con estos resultados, ya estamos, tenemos posibilidades de hacer aprobar o desaprobar, tres hipótesis de manera directa, tales como:
 - H1: “A mayor orientación al mercado, mayor orientación al cliente”.
 - H2: “A mayor orientación al mercado, mayor orientación a la competencia”.

FIGURA 77. CFA, SEM y EQS. EQS. Proceso de revisión rápida de resultados. Caso estudio 2 sin V1 y V3.



Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 77. CFA, SEM y EQS. Proceso de revisión rápida de resultados. Caso estudio 2 sin V1 y V3.



Fuente: EQS 6.2, con datos y adaptación propios.

H3: “A mayor orientación al mercado, mayor coordinación interfuncional”.

La respuesta a los tres es afirmativa en un valor correspondiente a (.924); (.850); y (.876). Lo anterior significa también, que el empresario con orientación a mercado, debe enfatizar más su atención a la orientación al cliente, seguido de la orientación de infraestructura y, finalmente, en la orientación a la competencia para optimizar sus recursos con mejores resultados.

Con los datos ajustados, se procederá a calcular la confiabilidad de cada constructo con el alfa de Cronbach, para posteriormente compararlo con el índice de fiabilidad compuesta IFC. Si cumplen el requisito, se realizará la prueba de validez calculando el índice de varianza extraída (IVE) correspondiente.

CFA, SEM y EQS. Cálculo del alfa de Cronbach sin V1 y V3

Dado que se requiere el alfa de Cronbach por constructo (NO por modelo, que es lo que se reporta en EQS), requerimos calcularlo. Este cálculo deberá proceder de la base de datos SPSS, en nuestro caso cfa om.sav y proceder únicamente con las variables que quedaron del modelo estructura, o sea, NO incluir las variables V1 ni V3. Es importante considerar que si está utilizando SPSS soportado con IOS o Windows (con un emulador como Parallels), ubique su archivo .SAV, de acuerdo con el sistema operativo con el cual va a trabajar. En nuestro caso es de forma directa en sistema operativo IOS.

Así, la secuencia de comandos para ubicar y abrir el archivo cfa om.sav, es:

Teclear en ícono SPSS desde aplicaciones o barra de herramientas-> ¿Qué desea hacer? Abrir un archivo de datos existente:... cfa om.sav->Aceptar->Nombre de archivo: cfa om.sav; Archivos de tipo: SPSS Statistics (*.sav)->Abrir ->Despliegue de datos en modo: vista de datos (figura 78).

Así, se procederá a realizar el cálculo de alfa de Cronbach, para cada uno de los ítems de los factores, con la siguiente secuencia inicial de comandos:

Analizar->Escala->Análisis de fiabilidad->Modelo: Alfa->Elementos: V2; V4; V5; V6 (en el caso del archivo, los marcados como: P02ACL; P04ACL; P05ACL; P06ACL)->Aceptar (figura 79).

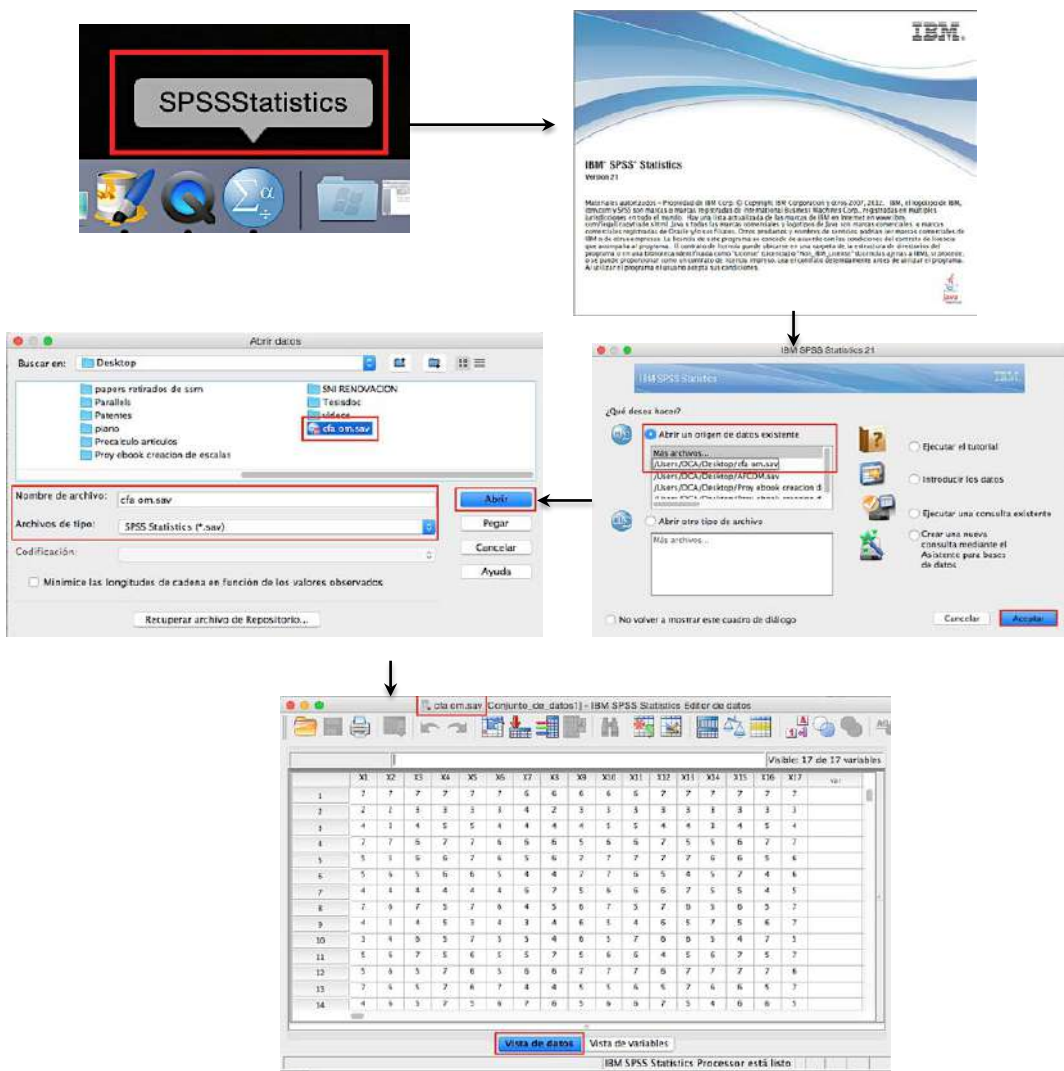
En consecuencia, el alfa de Cronbach por factor, es:

F1 (V2; V4; V5; V6)= .861

F2 (V7; V8; V9; V10; 11)= .906

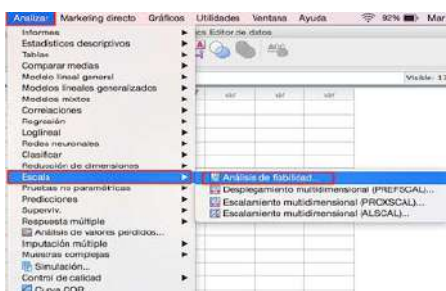
F3 (V12; V13; V14; V15; V16; V17)= .904

FIGURA 78. Secuencia de comandos apertura de archivo base de datos SPSS. Caso de estudio 2.



Fuente: EQS 6.2, con datos y adaptación propios.

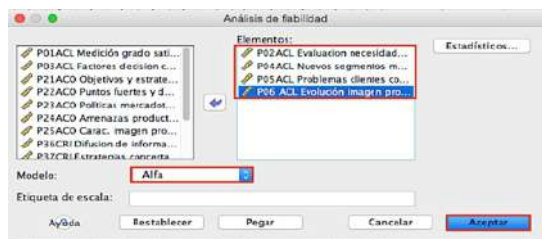
FIGURA 79. Secuencia de comando. Cálculo alfa de Cronbach con SPSS. Caso de estudio 2.



➔ Análisis de fiabilidad

[Conjunto_de_datos1] /Users/DCA/Desktop/cfa_om.sav

Escala: TODAS LAS VARIABLES



Resumen del procesamiento de los casos

Casos	Válidos	N	%
	Válidos	300	100,0
	Excluidos ^a	0	,0
	Total	300	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,861	4

➔ Análisis de fiabilidad

[Conjunto_de_datos1] /Users/DCA/Desktop/cfa_om.sav

Escala: TODAS LAS VARIABLES



Resumen del procesamiento de los casos

Casos	Válidos	N	%
	Válidos	300	100,0
	Excluidos ^a	0	,0
	Total	300	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

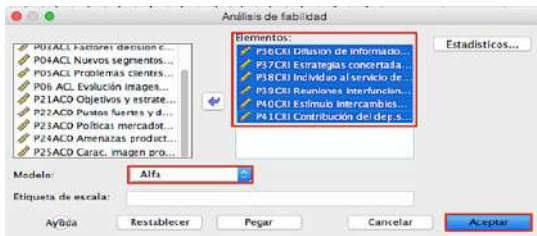
Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,908	5

➔ Análisis de fiabilidad

[Conjunto_de_datos1] /Users/DCA/Desktop/cfa_om.sav

Escala: TODAS LAS VARIABLES



Resumen del procesamiento de los casos

Casos	Válidos	N	%
	Válidos	300	100,0
	Excluidos ^a	0	,0
	Total	300	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,904	6

Fuente: EQS 6.2, con datos y adaptaciones propios.

CFA, SEM y EQS. Cálculo de IFC sin V1 y V3

Normalmente, en la investigación científica los modelos teóricos desarrollados en la literatura no aparecen implicados en un solo constructo, sino varios, por lo cual el alfa de Cronbach para cada factor por separado no considera la influencia sobre la fiabilidad del resto de constructos. Por esta razón, Fornell y Larcker (1981) proponen el cálculo del índice de la fiabilidad compuesta (IFC) para cada factor que, interpretándose exactamente igual que el alpha de Cronbach, sí que tiene en cuenta las interrelaciones. Por lo tanto, para determinarlo partimos del reporte de solución estandarizada (*standardized solution*) de la figura 77.

Calculando el reporte IFC, como se observa en la figura 80.

STANDARDIZED SOLUTION:

R-SQUARED

X2	=V2	=	.763 F1	+	.646 E2	.583
X4	=V4	=	.839*F1	+	.544 E4	.704
X5	=V5	=	.713*F1	+	.702 E5	.508
X6	=V6	=	.809*F1	+	.588 E6	.654
X7	=V7	=	.817 F2	+	.577 E7	.667
X8	=V8	=	.880*F2	+	.475 E8	.774
X9	=V9	=	.909*F2	+	.417 E9	.826
X10	=V10	=	.834*F2	+	.553 E10	.695
X11	=V11	=	.606*F2	+	.795 E11	.368
X12	=V12	=	.725 F3	+	.688 E12	.526
X13	=V13	=	.842*F3	+	.539 E13	.710
X14	=V14	=	.772*F3	+	.636 E14	.595
X15	=V15	=	.803*F3	+	.596 E15	.644
X16	=V16	=	.778*F3	+	.628 E16	.605
X17	=V17	=	.776*F3	+	.631 E17	.602
F1	=F1	=	.924*F4	+	.383 D1	.854
F2	=F2	=	.850*F4	+	.527 D2	.723
F3	=F3	=	.876*F4	+	.483 D3	.767

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 80. CFA, SEM y EQS. Recopilación de información que calcula el IFC. Caso de estudio 2 sin V1 y V3.

Constructo e indicadores	Carga factorial estandarizada L_{ij}	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de fiabilidad compuesta $IFC > .7$ $\frac{\sum(L_{ij})^2 / [\sum(L_{ij})^2 + \sum Var(E_{ij})]}{\sum Var(E_{ij})}$
Orientación al cliente (F₁)			
V ₂	.763	.418	.863
V ₄	.839	.296	
V ₅	.713	.492	
V ₆	.809	.346	
Total	3.124	1.552	
Orientación a la competencia (F₂)			
V ₇	.817	.333	.907
V ₈	.880	.226	
V ₉	.909	.174	
V ₁₀	.833	.306	
V ₁₁	.606	.633	
Total	4.045	1.672	
Coordinación interfuncional (F₃)			
V ₁₂	.725	.474	.905
V ₁₃	.842	.291	
V ₁₄	.772	.404	
V ₁₅	.803	.355	
V ₁₆	.778	.395	
V ₁₇	.776	.398	
Total	4.696	2.317	
Orientación al mercado			
F ₁	.924	.146	.915
F ₂	.850	.277	
F ₃	.876	.233	
Total	2.650	.656	

Fuente: EQS 6.2, con datos y adaptación propios.

CFA, SEM y EQS. Cálculo de IVE sin V1 y V3

Fornell y Larcker (1981) presentan el índice IVE como la relación entre la varianza que es capturada por un factor i en relación con la varianza total debida al error de medida de ese factor. Nótese que la única diferencia con el IFC es que cada carga factorial estandarizada es primero elevada al cuadrado antes de ser sumadas. Fornell y Larcker (1981) sugieren que es deseable que el constructo tenga valores de $IVE \geq .5$, es decir, que sea superior la varianza capturada por el factor que la debida al error de medida (figura 81).

FIGURA 81. CFA, SEM y EQS. EQS. Recopilación de información que calcula el IVE del caso de estudio 2 sin V1 y V3.

Constructo e indicadores	Cuadrado de la carga factorial estandarizada $(L_{ij})^2$	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de varianza compuesta $IVE \geq .5$ $\frac{\sum (L_{ij})^2}{[\sum (L_{ij})^2 + \sum Var(E_{ij})]}$
Orientación al cliente (F₁)			
V ₂	.582	.418	.612
V ₄	.704	.296	
V ₅	.508	.492	
V ₆	.654	.346	
Total	2.448	1.552	
Orientación a la competencia (F₂)			
V ₇	.667	.333	.666
V ₈	.774	.226	
V ₉	.826	.174	
V ₁₀	.694	.306	
V ₁₁	.367	.633	
Total	3.328	1.672	
Coordinación interfuncional (F₃)			
V ₁₂	.526	.474	.614
V ₁₃	.709	.291	
V ₁₄	.596	.404	
V ₁₅	.645	.355	
V ₁₆	.605	.395	
V ₁₇	.602	.398	
Total	3.683	2.317	
Orientación al mercado			
F ₁	.854	.146	.806
F ₂	.723	.277	
F ₃	.767	.233	
Total	2.334	.565	

Fuente: EQS 6.2, con datos y adaptación propios.

CFA, SEM y EQS. Análisis de la validez de la escala sin V1 y V3

Recordando a Bohrnstedt (1976): “validez es el grado en que un instrumento mide el concepto bajo estudio”. Desafortunadamente, el concepto tiene múltiples dimensiones que deben explicarse y analizarse por separado.

Validez de contenido

La figura 77 muestra los estadísticos de bondad de ajuste proporcionados por EQS.

GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST			
ROBUST INDEPENDENCE MODEL CHI-SQUARE =	2196.220	ON	105 DEGREES OF FREEDOM
INDEPENDENCE AIC =	1986.220	INDEPENDENCE CAIC =	1492.323
MODEL AIC =	35.128	MODEL CAIC =	-369.398
SATORRA-BENTLER SCALED CHI-SQUARE =	207.1276	ON	86 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS		0.00000	

Fuente: EQS 6.2, con datos y adaptación propios.

De su análisis se desprende que el valor de la Chi-cuadrado y el número de grados de libertad es superior a 2 ($207.1276/86 = 2.408$). Parece indicar que el modelo *Orientación al mercado* no tiene un buen ajuste, pero es posible seguir analizando otros estadísticos. Entre el conjunto de indicadores de ajuste que proporciona EQS, se recomienda analizar aquellos no normalizados, dado que tienen menor tendencia a ofrecer resultados sesgados en casos de pequeñas muestras (Bentler y Wu, 1993). Se observa que, tanto el NNFI (Bentler y Bonnet, 1980) como el CFI (Bentler y Wu, 1993), que cumplen este requisito, tienen valores $> .9$, indicando un ajuste razonable. RMSEA = .069 cumple con el requisito de ser $< .08$. De la misma figura 77, ver el apartado Reporte ajuste de índices (*fit indices*).

FIT INDICES			
BENTLER-BONETT	NORMED FIT INDEX =	0.906	
BENTLER-BONETT	NON-NORMED FIT INDEX =	0.929	
COMPARATIVE FIT INDEX (CFI)		= 0.942	
BOLLEN'S	(IFI) FIT INDEX =	0.943	
MCDONALD'S	(MFI) FIT INDEX =	0.817	
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)		=	0.069
90% CONFIDENCE INTERVAL OF RMSEA (0.057,	0.080)

Fuente: EQS 6.2, con datos y adaptación propios.

Ya hemos estimado el modelo de medida y hemos comprobado que ofrece un buen ajuste como base de la validez de contenido. Pues bien, sobre este modelo de medida, ya es posible analizar la validez convergente y la validez discriminante de las escalas implicadas. Para analizar la validez nomológica, será necesario estimar el modelo teórico.

Validez de constructo o de concepto

La validez de constructo o de concepto se divide en dos: *validez convergente* y *validez discriminante*. Una escala tiene validez de constructo, cuando tiene validez convergente y validez discriminante.

Validez convergente

La validez convergente existe cuando se emplean distintos instrumentos para medir un mismo constructo (distintos ítems para una misma variable latente) y estos instrumentos están fuertemente correlacionados. Si analizamos de la misma figura 77, las secciones de reporte las ecuaciones de medida (*measurement equations*) se deberá observar que las cargas factoriales sean < 1 significativas a 5%.

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
(ROBUST STATISTICS IN PARENTHESES)

```

X2    =V2 =          1.000 F1    +          1.000 E2

X4    =V4 =          1.164*F1    +          1.000 E4
                .078
                14.879@
                (
                (.081)
                ( 14.284@)

X5    =V5 =          .943*F1    +          1.000 E5
                .076
                12.410@
                (
                (.093)
                ( 10.125@)

```

Fuente: EQS 6.2, con datos y adaptación propios.

La sección de reporte de soluciones estandarizadas (standardized solutions), así como el modelo de medida, observamos que todas las cargas factoriales son significativas, como se desprende de sus respectivas t para todas las variables.

Asimismo, si calculáramos el promedio de los mismos veríamos que superan claramente .7 y de la misma figura 77 en su sección de prueba del multiplicador de Lagrange (*Lagrange multiplier test*) el ajuste del modelo mejoraría significativamente con la consideración, con suficiente marco teórico y el EFA validado de cambiar la V10 a F3, ambas como intento de mejora del modelo.

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1

PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:

PVV PFV PFF PDD GVV GVF GFV GFF BVF BFF

CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT						
STEP	PARAMETER	CHI-SQUARE	D. F.	PROB.	CHI-SQUARE	PROB.	HANCOCK'S SEQUENTIAL D. F.	PROB.	ROBUST PREDICTED RMSEA	CFI
1	V10, F3	18.638	1	0.000	18.638	0.000	86	1.000	0.064	0.951
2	V7, F3	29.246	2	0.000	10.607	0.001	85	1.000	0.061	0.955
3	V6, F3	33.486	3	0.000	4.241	0.039	84	1.000	0.060	0.957

Fuente: EQS 6.2, con datos y adaptación propios.

Validez discriminante

Por otro lado, la *validez discriminante* se consigue cuando, si diversos instrumentos de medida están diseñados para medir distintas variables latentes, entonces las correlaciones entre dichos instrumentos son bajas. Una escala tiene, por lo tanto, validez discriminante cuando no mide un constructo para el que no se diseñó. En nuestro caso NO es posible medir la validez discriminante porque estamos utilizando un factor de segundo orden (F4), que NO TIENE COVARIANZA con otro factor de segundo orden (es el único).

Validez nomológica

El modelo teórico tendrá validez nomológica si, efectivamente, no hay diferencias significativas entre los ajustes del modelo de medida y el teórico, dado que las escalas habrán sido capaces de establecer relaciones predictivas de otras variables tan sustantivas que, siendo menos, igualan la bondad del modelo.

En este momento, es posible hacer un comparativo de las Chi-cuadrado del modelo teórico (figura 67) vs. el de medida de la figura 77.

SATORRA-BENTLER SCALED CHI-SQUARE = 369.7671 ON 116 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

SATORRA-BENTLER SCALED CHI-SQUARE = 207.1276 ON 86 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

Fuente: EQS 6.2, con datos y adaptación propios.

Nos muestran los estadísticos de bondad de ajuste que son necesarios para determinar la validez nomológica. Así, es posible comprobar que las respectivas Chi cuadrado del modelo teórico y de medida (que obtuvimos al analizar la validez discriminante) son:

Modelo teórico: 369.7671; grados de libertad, 116

Vs.

Modelo de medida: 207.1276; grados de libertad, 86

FIGURA 82. Área correspondiente al extremo derecho de una distribución Chi-cuadrado.

0.20	0.10	Área en el extremo derecho				Grados de libertad
		0.05	0.25	0.01		
1.642	2.706	3.841	5.024	6.635	1	
3.219	4.605	5.991	7.378	9.210	2	
4.642	6.251	7.815	9.348	11.345	3	
5.989	7.779	9.488	11.143	13.277	4	
7.289	9.236	11.070	12.833	15.086	5	
8.558	10.645	12.592	14.449	16.812	6	
9.803	12.017	14.067	16.013	18.475	7	
11.030	13.362	15.507	17.535	20.090	8	
12.242	14.684	16.919	19.023	21.666	9	
13.442	15.987	18.307	20.483	23.209	10	
14.631	17.275	19.675	21.920	24.725	11	
15.812	18.549	21.026	23.337	26.217	12	
16.985	19.812	22.362	24.736	27.688	13	
18.151	21.064	23.685	26.119	29.141	14	
19.311	22.307	24.996	27.488	30.578	15	
20.465	23.542	26.296	28.845	32.000	16	
21.615	24.769	27.587	30.191	33.409	17	
22.760	25.989	28.869	31.526	34.805	18	
23.900	27.204	30.144	32.852	36.191	19	
25.038	28.412	31.410	34.170	37.566	20	
26.171	29.615	32.671	35.479	38.932	21	
27.301	30.813	33.924	36.781	40.289	22	
28.429	32.007	35.172	38.076	41.638	23	
29.553	33.196	36.415	39.364	42.980	24	
30.675	34.382	37.652	40.647	44.314	25	
31.795	35.563	38.885	41.923	45.642	26	
32.912	36.741	40.113	43.194	46.963	27	
34.027	37.916	41.337	44.461	48.278	28	
35.139	39.087	42.557	45.722	49.588	29	
36.250	40.256	43.773	46.979	50.892	30	

Fuente: EQS 6.2, con datos y adaptación propios.

La validez nomológica se establecería calculando un test de diferencias entre las Chi cuadrado, tal como hicimos para la validez discriminante:

Diferencias entre las Chi-Cuadrado: $(369.7671 - 207.1276) = 162.6395$

Diferencias en grados de libertad: $(116 - 86) = 30$

Valor crítico $p < 0.01$: 50.892 (figura 82).

En otras palabras, el modelo teórico ofrece un ajuste significativamente mejor que el modelo de medida (diferencia inferior al valor crítico), lo que permitiría constatar la validez nomológica de las escalas.

CFA, SEM y EQS. Reportando resultados: consistencia interna y validez convergente

En este punto, es importante tener una referencia del acomodo de los datos, como se sugiere en el ejemplo de la figura 83.

donde las columnas significan:

Factores, listado de los involucrados en el estudio. En nuestro caso:

Orientación al cliente (F1);

Orientación a la competencia (F2);

Coordinación interfuncional (F3) y

Orientación al mercado.

Ítem, consecutivo de las variables (V) no suprimidas mostradas, como: V1; V2; V3, etc.

- ▶ *Carga factorial* $> .6$ (a) (figura 77) donde sólo se reportan los factores no suprimidos y con carga factorial $> .6$. Es posible que el investigador utilice escalas previas que si a su juicio, por el marco teórico, deban asignarse más ítems, es válido realizarlo siempre y cuando se mencione que la nueva escala es una adaptación de la previa más la añadida como propuesta del investigador dadas las condiciones de la realidad en la que se aplica contra las que originalmente la escala se creó. En esta etapa es donde se valida o no la pertinencia de los nuevos ítems basados en el marco teórico planteado.
- ▶ *Valor robusto t* (figura 40 para mayor detalle).
Con lo anterior, es posible afirmar que si los valores de t (sección de reporte ecuaciones de medida/*measurement equations* y ecuaciones de constructo/*construct equations para los factores* (figura 79) son mayores ($>$) 2.58 errores estándar de la media, entonces se anotarán tres asteriscos (***) por carga factorial, lo cual indica que $p < .0001$. Así se tiene una seguridad que las cargas factoriales aportan sus valores a 99% de confiabilidad:

- ▶ *Alfa de Cronbach*, calculada y reportada de acuerdo con la figura 81 mediante el uso de base de datos SPSS (cfa om.sav)
- ▶ IFC. Se toman los datos de la figura 82 y es muy similar al alfa de Cronbach.
- ▶ IVE. Con base en los datos de la figura 83.

FIGURA 83. Reporte tabla de consistencia interna y validez convergente del modelo teórico sin V1 y V3.

Factor	Item V	Carga Factorial >0.6 (a)	Valor Robustot	Factor Promedio de Carga Factorial	Fiabilidad de la Escala		
					Alfa de Cronbach >=0.7 (b)	IFC >0.7 (b)	IVE >0.5 (c)
Orientación al cliente (F ₁)	V ₂	.763***	1.000a	0.781	.861	.863	.612
	V ₄	.839***	14.284				
	V ₅	.713***	10.125				
	V ₆	.809***	13.394				
Orientación a la competencia (F ₂)	V ₇	.817***	1.000a	.809	.906	.907	.666
	V ₈	.880***	19.924				
	V ₉	.909***	17.996				
	V ₁₀	.833***	14.630				
Coordinación interfuncional (F ₃)	V ₁₁	.606***	9.886	.782	.904	.905	.614
	V ₁₂	.725***	1.000a				
	V ₁₃	.842***	15.555				
	V ₁₄	.772***	11.990				
	V ₁₅	.803***	13.575				
	V ₁₆	.778***	12.409				
Orientación al mercado	F ₁	.924***	13.244	.883	-	.915	.806
	F ₂	.850***	13.981				
	F ₃	.876***	11.605				
Conclusión Resultados: Chi-cuadrado= 207.1276; gl= 86; $p < .01$; NFI = .906; NNFI = .929; CFI = .942; RMSEA = .069 Conclusión: las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos. Esto es, no hay error en el ajuste del modelo. Aunque existe suficiente evidencia de validación convergente y fiabilidad, que justifica la fiabilidad interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999), no se tiene validez discriminante porque sólo existe un actor F4 (sólo una variable independiente). Por lo tanto, no es posible determinar un Reporte tabla de validez discriminante del modelo teórico".							

Nota:

*** Parámetros constreñidos a ese valor en el proceso de identificación = $p < 0.001$.

(a) De acuerdo con Bagozzi y Yi (1988).

(b) De acuerdo con Hair et al. (1999).

© Índice de Varianza Extraída (IVE), de acuerdo con Fornell y Larcker, 1981.

Fuente: EQS 6.2, con datos y adaptación propios.

Conclusión

Hasta este momento, tomando en cuenta la figura 85 afirmamos que para maximizar la orientación a mercado deberá ser atendido en mayor medida, cada una de las cargas factoriales con valor absoluto alto y hacer una explicación de la realidad, con base en el marco teórico del porqué debe ser así (o no), según los argumentos del investigador. Todo lo anterior basado en las cargas factoriales encontradas con una probabilidad de equivocación de 1/100 o 99% de que lo afirmado ocurra (ver las ***) donde los renglones por debajo de las cargas factoriales, significan:

S-B χ^2 Índice Satorra Bentler Chi-cuadrada, es el recomendado cuando se tienen más de dos factores para generar datos robustos. Así, se explica la lectura reportada en la base de la figura 39.

“Conclusión:

Resultados: Chi-Cuadrado= 207.1276; gl= 86; $p < .01$; NFI = .906; NNFI = .929; CFI = .942; RMSEA = .069.

“Conclusión: las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos. Esto es, no hay error en el ajuste del modelo. Aunque existe suficiente evidencia de validación convergente y fiabilidad, que justifica la fiabilidad interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999), no se tiene validez discriminante porque sólo existe un factor F4 (sólo una variable independiente). Por lo tanto, no es posible determinar un Reporte tabla de validez discriminante del modelo teórico”.

Así, una posible redacción de los resultados de la figura 85, sería:

“...Por otro lado, para evaluar la fiabilidad y validez de las escalas de medida se realizó un análisis factorial confirmatorio (AFC), utilizando el método de máxima verosimilitud con el software EQS 6.2 (Bentler, 2005; Brown, 2006; Byrne, 2006). Asimismo, la fiabilidad de las escalas de medida se evaluó a partir del coeficiente alfa de Cronbach y el índice de fiabilidad compuesta (IFC) (Bagozzi y Yi, 1988). Todos los valores de la escala excedieron el nivel recomendado de .70 para el alfa de Cronbach y el IFC que proporciona una evidencia de fiabilidad y justifica la fiabilidad interna de las escalas (Nunnally y Bernstein, 1994; Hair et al., 1995).

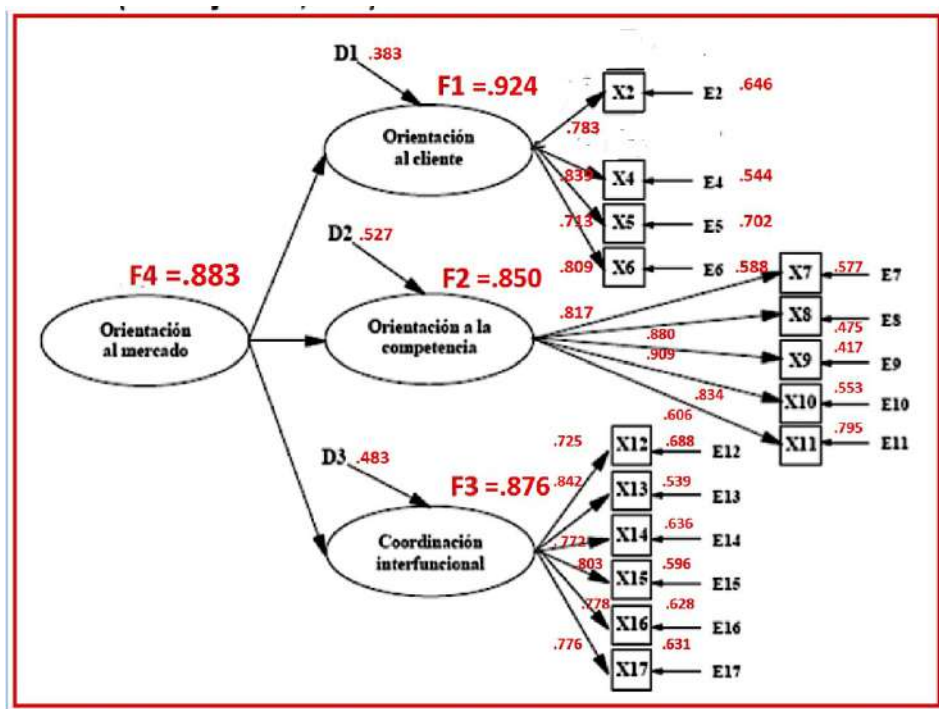
Los resultados de la aplicación del AFC se presentan en la tabla 1 y sugiere que el modelo de medición proporciona un buen ajuste de los datos ($S-BX^2 = XXX.YYY$; $df = XX$; $p = 0.000$; $NFI > .9XXX$; $NNFI > .9XXX$; $CFI > .9XXX$; y $RMSEA < .08XXX$). Como evidencia de la validez convergente, los resultados del AFC indican que todos los ítems de los factores relacionados son significativos ($p < 0.001$) y el tamaño de

todas las cargas factoriales estandarizadas son superiores a 0.60 (Bagozzi y Yi, 1988).

La tabla xxx muestra una alta consistencia interna de los constructos, en cada caso, la α de Cronbach excede el valor de .70 recomendado por Nunnally y Bernstein (1994). El IFC representa la varianza extraída entre el grupo de variables observadas y el constructo fundamental (Fornell y Larcker, 1981). Generalmente un IFC superior a .60 es considerado como deseable (Bagozzi y Yi, 1988), en este estudio este valor es superior. El índice de la varianza extraída (IVE) fue calculado para cada par de constructos, resultando un IVE superior a .5 en todos los factores (Fornell y Larcker, 1981)..."

El diagrama de trayectoria final, es el mostrado en la figura 84.

FIGURA 84. Diagrama de trayectoria final. Caso de estudio 2.



Fuente: Elaboración propia.

CFA. Modelo de segundo orden. Análisis de caso de estudio 3

Tomemos como ejemplo el concepto de *marketing mix*, propuesto por el doctor Gonzalo Maldonado Guzmán (2016), el cual ha sido un concepto muy tratado y que a primera vista es interesante actualizar llevando a cabo una búsqueda a través de las múltiples herramientas de bases de datos que existen hoy en día, con base en, por cierto, en palabras clave o por fases completas mediante el empleo de operadores lógicos. Éste fue aplicado en la ciudad de Aguascalientes, México, a empresarios PyME manufactureros, con resultados reflejados en la base de datos marketing mix. Sav.

Adjunto mostramos la propuesta de escala de nuestro estudio de caso, como se aprecia en la figura 85.

FIGURA 85. Marketing mix.

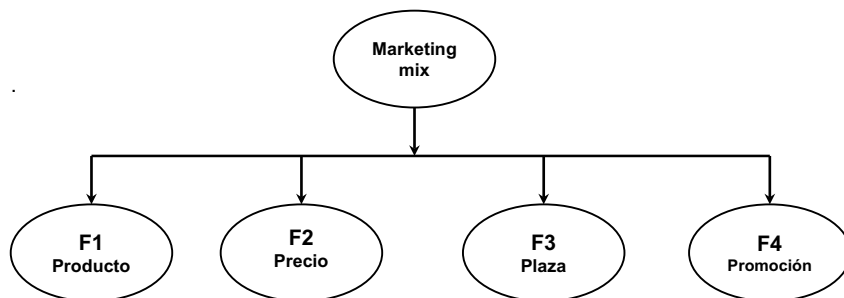
Item	Var	Indicador	Total Desacuerdo			Total Acuerdo	
			1	2	3	4	5
Indique si su empresa a nivel de producto:							
MPP1	V1	Tiene una identidad o marca de sus productos					
MPP2	V2	Desarrolla constantemente nuevos productos					
MPP3	V3	Desarrolla nuevas líneas de productos y servicios					
MPP4	V4	Ha modificado líneas de productos o servicios por emergencia					
MPP5	V5	En comparación con la competencia, mi empresa es a menudo la primera en introducir nuevos productos o servicios					
MPP6	V6	Se distingue por la calidad de sus productos					
MPP7	V7	Se distingue por la especialización en sus productos					
MPP8	V8	Toma en cuenta las necesidades del mercado para producir sus productos					
MPP9	V9	Se centra en la maximización de las necesidades de sus clientes en cuento a los requerimientos de sus productos					
MPP10	V10	Invierte recursos necesarios en el desarrollo de nuevos productos o servicios					
MPP11	V11	Realiza estudios de mercado para desarrollar nuevos productos o servicios					
MPP12	V12	Es muy sensible a cómo evalúa el cliente sus productos y servicios, por lo que si se requieren modificaciones se realizan inmediatamente.					
MPP13	V13	Tiene un diseño de la marca, logotipo, símbolo, lema, el embalaje, etc. de sus productos y servicios para maximizar su imagen y comercialización.					
Indique si su empresa a nivel de precio:							
MPP1	V14	Optimiza el precio, el costo y la calidad del producto / servicio que ofrecen para satisfacer o exceder las expectativas de los clientes.					
MPP2	V15	Los precios de nuestros productos son inferiores a los de la competencia					
MPP3	V16	Los precios de nuestros productos son los adecuados de acuerdo a los costos					

Finalizando la escala. Análisis factorial confirmatorio y ecuación estructural

Item	Var	Indicador	Total Desacuerdo			Total Acuerdo	
Indique si su empresa a nivel de producto:			1	2	3	4	5
MPR4	V17	Los precios de nuestros productos varían en función de la cantidad de productos que nos compren.					
MPR5	V18	Aplicamos una política de descuento por pronto pago					
MPR6	V19	Aplicamos una estrategia de precios					
MPR7	V20	Comúnmente negociamos el precio de nuestros productos con nuestros clientes					
Indique si su empresa a nivel de plaza:			1	2	3	4	5
MPL1	V21	Influye o controla los canales de distribución de sus productos					
MPL2	V22	Desarrolla o implementa técnicas innovadoras de distribución					
MPL3	V23	Utiliza agentes de ventas altamente calificados y eficientes					
MPL4	V24	Cuenta con productos que son muy aceptados por los intermediarios del canal de distribución.					
MPL5	V25	Resuelve eficientemente sus problemas de logística					
MPL6	V26	Tiene una flexibilidad en sus procesos de logística					
MPL7	V27	Gestiona adecuadamente la cadena de suministro					
MPL8	V28	Permanentemente se mantiene en contacto con sus distribuidores					
MPL9	V29	Utiliza un software para controlar los pedidos y las entregas					
MPL10	V30	Tiene un sistema para controlar los la percepción de valor de marca de sus productos por parte de los intermediarios y distribuidores.					
MPL11	V31	Subcontrata frecuentemente las actividades de distribución y logística					
Indique si su empresa a nivel de promoción:			1	2	3	4	5
MPO1	V32	Aprovecha cada herramienta de comunicación para promocionar sus productos o servicios.					
MPO2	V33	Tiene una persona responsable capaz de monitorear la promoción de sus productos o servicios.					
MPO3	V34	La publicidad que realizamos es mejor que la que realiza la competencia					
MPO4	V35	Los medios de comunicación que utilizamos son los adecuados					
MPO5	V36	La inversión que realizamos en publicidad es la adecuada					
MPO6	V37	La publicidad que realizamos está dirigida a nuestro mercado meta					
MPO7	V38	Las ventas de la empresa se han incrementado gracias a la publicidad que realizamos					
MPO8	V39	Realizamos constantemente campañas promocionales de nuestros productos					

Cuyo diagrama de trayectoria se muestra en la figura 86.

FIGURA 86. Diagrama de trayectoria modelo marketing mix.



Fuente: Maldonado-Guzman (2016).

CFA, SEM y EQS. Entrada de datos al programa

Se sugiere seguir lo recomendado en las figuras 11, accediendo el archivo de entrada de programa (*.EQS), llamado marketing mix.sav (formato SPSS), el cual, al ser abierto, cambia su formato original *.SAV a *.EES (marketing mix.ess) para su despliegue y manipulación (figura 87).

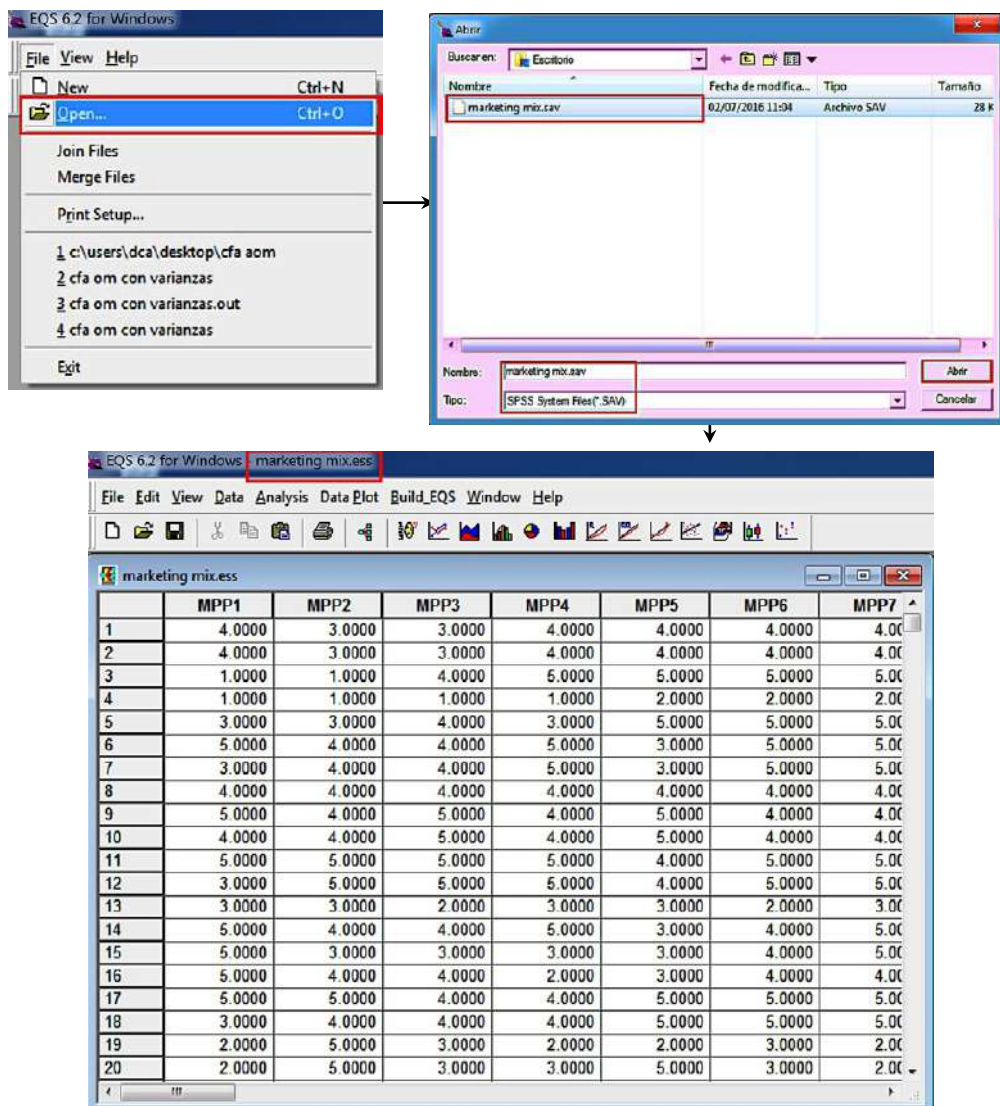
Así también es posible verificar el archivo de programa de entrada EQS, del que resaltamos el nombre del archivo en el párrafo /TITLE: “CFA y SEM de Marketing MIX Nombre: marketing mix”. En el párrafo: /SPECIFICATIONS, la declarativa de analizar 39 variables con 400 casos.

NOTA: EQS, para el caso de sistema operativo WINDOWS debe especificarse, en un archivo externo, la trayectoria del archivo de programa de entrada; por ejemplo: DATAFILE='E://cfa om.eqs'. Si es IOS, sólo se menciona el nombre del archivo dentro de la computadora a nivel de escritorio con los archivos apareciendo en el emulador de Windows (“Parallels”, en nuestro caso: DATAFILE='marketing mix.ess');. (figura 7.88).

Otra instrucción es MATRIX=RAW, que indica que la matriz de datos será generada por el mismo software. La instrucción METHOD=ML que indica que se aplicará máxima verosimilitud (*ML. maximum likelihood*), la instrucción ROBUST en la que se da el indicativo que los datos tienen filtros para garantizar mínimos errores (no se recomienda trabajar con los datos originales) y finalmente, la instrucción ANALYSIS= COV, sobre la aplicación de covarianzas (es mejor a la matriz de correlaciones).

En el párrafo /EQUATIONS, se tienen las ecuaciones de todas y cada una de las variables con sus errores señalando con la supresión de asterisco (*) al menos una variable que tenga varianza 1. En el párrafo de /VARIANCES, observe en particular las expresiones que involucran a los constructos o factores que identifican al modelo de

FIGURA 87. EQS. Selección de archivo externo. Caso de estudio 2. Cambio de formato SPSS (*.SAV) a EQS (*.ESS).



Fuente: EQS 6.2, con datos y adaptación propios.

segundo orden: F1=*, F2=*, F3=*, y F4=* en las que al incluir al F4 en las mismas, se da la idea de que se correrá al mismo tiempo el CFA con el SEM, situación que es válida y que facilita el análisis, dado lo complejo del modelo. Se incluye el párrafo

/COVARIANCES, ya que se consideran que corren como variables independientes, en todas las combinaciones cubiertas como:

F1, F2 =*;

F1, F3 =*;

F1, F4 =*;

F2, F3 =*;

F2, F4 =*;

F3, F4 =*;

Es importante hacer notar que tanto los nombres de la base de datos proveniente de SPSS (.SAV), tanto el nombre de variables y su aparición sea estrictamente como corresponde al modelo teórico, con lectura de izquierda a derecha, ya que esto es una condición necesaria para que el modelo EQS, pueda correr correctamente y/o subdividir las bases de datos. En otras palabras, debe haber correspondencia en nombre y orden de aparición de las variables de la base de datos, con el modelo teórico, con el fin de que EQS ejecute correctamente las instrucciones con las variables correspondientes y produzca los resultados esperados (figura 89).

FIGURA 88. Vista de archivos desde emulador Windows (parallels) manejo de EQS con IOS.



Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 89. EQS. Codificación de entrada de datos. Caso de estudio 2.

```

/TITLE
CFA y SEM de Marketing MIX
Nombre: marketing mix

/SPECIFICATIONS
VARIABLES= 39; CASES= 400;
DATAFILE='marketing mix.ess'; MATRIX=RAW;
METHOD=ML, ROBUST; ANALYSIS = COV;

/EQUATIONS

V1 = F1 +E1 ;
V2 = *F1 +E2 ;
V3 = *F1 +E3 ;
V4 = *F1 +E4 ;
V5 = *F1 +E5 ;
V6 = *F1 +E6 ;
V7 = *F1 +E7 ;
V8 = *F1 +E8 ;
V9 = *F1 +E9 ;
V10 = *F1 +E10 ;
V11 = *F1 +E11 ;
V12 = *F1 +E12 ;
V13 = *F1 +E13 ;

V14 = F2 +E14 ;
V15 = *F2 +E15 ;
V16 = *F2 +E16 ;
V17 = *F2 +E17 ;
V18 = *F2 +E18 ;
V19 = *F2 +E19 ;
V20 = *F2 +E20 ;

V25 = F3 +E25 ;
V26 = *F3 +E26 ;
V27 = *F3 +E27 ;
V28 = *F3 +E28 ;
V29 = F3 +E29 ;
V30 = *F3 +E30 ;
V31 = *F3 +E31 ;

V32 = F4 +E32 ;
V33 = *F4 +E33 ;
V34 = *F4 +E34 ;
V35 = *F4 +E35 ;
V36 = *F4 +E36 ;
V37 = *F4 +E37 ;
V38 = *F4 +E38 ;
V39 = *F4 +E39 ;

/VARIANCES

E1 ==*;
E2 ==*;
E3 ==*;
E4 ==*;
E5 ==*;
E6 ==*;
E7 ==*;
E8 ==*;
E9 ==*;
E10 ==*;
E11 ==*;
E12 ==*;
E13 ==*;
E14 ==*;
E15 ==*;
E16 ==*;

E33 ==*;
E34 ==*;
E35 ==*;
E36 ==*;
E37 ==*;
E38 ==*;
E39 ==*;

F1 ==*;
F2 ==*;
F3 ==*;
F4 ==*;

/COVARIANCES

F1, F2 ==*;
F1, F3 ==*;
F1, F4 ==*;
F2, F3 ==*;
F2, F4 ==*;
F3, F4 ==*;

/LMTEST

/WTEST

/PRINT
FIT = ALL;

/END

```

Fuente: EQS 6.2, con datos y adaptación propios.

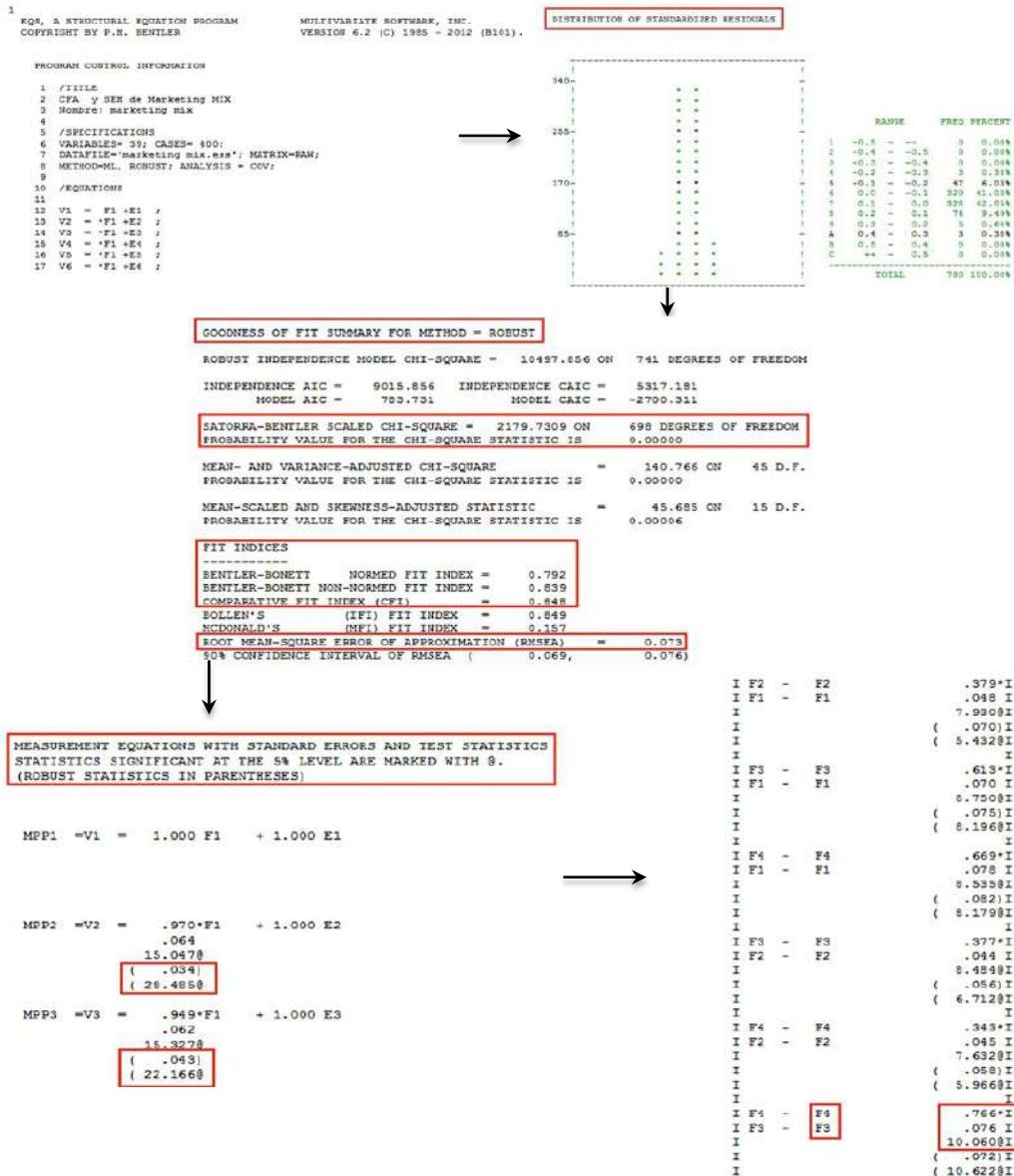
Realizado lo anterior, ejecute el programa EQS, siguiendo lo mostrado en las instrucciones del apartado CFA y EQS. Cómo ejecutar el programa.

Se procede a revisar rápidamente las secciones de reporte de resultados como se ve en la figura 90.

CFA, SEM y EQS. Discusión rápida de resultados.

De los reportes emitidos en la figura 90, podemos observar y concluir de cada sección de reporte lo siguiente:

FIGURA 90. CFA, SEM y EQS. Proceso de revisión rápida de resultados. Caso de estudio 3.



Fuente: EQS 6.2, con datos y adaptación propios.

STANDARDIZED SOLUTION:				R-SQUARED
MPP1	=V1	= .762 F1	+ .648 E1	.580
MPP2	=V2	= .720*F1	+ .694 E2	.518
MPP3	=V3	= .731*F1	+ .682 E3	.534
MPP4	=V4	= .618*F1	+ .786 E4	.382
MPP5	=V5	= .735*F1	+ .679 E5	.539
MPP6	=V6	= .723*F1	+ .691 E6	.522
MPP7	=V7	= .733*F1	+ .680 E7	.537
MPP8	=V8	= .803*F1	+ .596 E8	.645
MPP9	=V9	= .820*F1	+ .572 E9	.673
MPP10	=V10	= .810*F1	+ .587 E10	.656
MPP11	=V11	= .729*F1	+ .684 E11	.532
MPP12	=V12	= .778*F1	+ .629 E12	.605
MPP13	=V13	= .695*F1	+ .719 E13	.483
MPR1	=V14	= .592 F2	+ .806 E14	.350
MPR2	=V15	= .595*F2	+ .804 E15	.354
MPR3	=V16	= .601*F2	+ .800 E16	.361
MPR4	=V17	= .773*F2	+ .634 E17	.598
MPR5	=V18	= .768*F2	+ .640 E18	.590
MPR6	=V19	= .819*F2	+ .573 E19	.671
MPR7	=V20	= .699*F2	+ .715 E20	.488
MPL1	=V21	= .725 F3	+ .689 E21	.525
MPL2	=V22	= .792*F3	+ .611 E22	.627
MPL3	=V23	= .765*F3	+ .643 E23	.586
MPL4	=V24	= .806*F3	+ .592 E24	.650
MPL5	=V25	= .815 F3	+ .579 E25	.664
MPL6	=V26	= .815*F3	+ .580 E26	.664
MPL7	=V27	= .805*F3	+ .593 E27	.648
MPL8	=V28	= .703*F3	+ .711 E28	.494
MPL9	=V29	= .713 F3	+ .701 E29	.508
MPL10	=V30	= .813*F3	+ .582 E30	.661
MPL11	=V31	= .728*F3	+ .686 E31	.529
EMO1	=V32	= .776 F4	+ .630 E32	.603
MPO2	=V33	= .765*F4	+ .644 E33	.585
MPO3	=V34	= .884*F4	+ .467 E34	.782
MPO4	=V35	= .862*F4	+ .507 E35	.743
MPO5	=V36	= .893*F4	+ .450 E36	.797
MPO6	=V37	= .835*F4	+ .550 E37	.697
MPO7	=V38	= .871*F4	+ .491 E38	.759

STANDARDIZED SOLUTION:				R-SQUARED
MPO8	=V39	= .729*F4	+ .685 E39	.531

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1

PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:

FFV FFV FFF FDD GVV GVF GFV GFF BVF BFF

CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT			
STEP	PARAMETER	CHI-SQUARE	D.F. PROB.	CHI-SQUARE	PROB.	HANCOCK'S SEQUENTIAL D.F. PROB.	ROBUST PREDICTED RMSEA CFI
1	V14, F1	52.315	1 0.000	52.315	0.000	698 1.000	0.072 0.853
2	V11, F4	87.292	2 0.000	34.976	0.000	697 1.000	0.071 0.857
3	V32, F2	111.710	3 0.000	24.419	0.000	696 1.000	0.070 0.859
4	V9, F3	135.570	4 0.000	23.859	0.000	695 1.000	0.070 0.862
5	V24, F2	159.242	5 0.000	18.672	0.000	694 1.000	0.069 0.863
6	V16, F1	169.282	6 0.000	15.040	0.000	693 1.000	0.069 0.865

Fuente: EQS 6.2, con datos y adaptación propios.

- ▶ Gráfico de distribución de residuos (distribution of standardized residuals). Tiene problemas de datos alejados de la media. NO cumple.
- ▶ Índices de ajuste (fit Indices) robustos (robust), valores a analizar: Chi-cuadrado de Satorra-Bentler= 2179.7309; $gl= 698$ (Nota: $2179.7309/698=3.122>2$ Tendencia a que los índices de ajuste sean $<.9$); valor $p<0.01$. NO cumple.
NFI= .792; NNFI= .839; CFI= .848. NO Cumple $>.9$ (para nueva escala). RMSEA=.073<.08. Cumple.
- ▶ Ecuaciones de medición (*measurment equations*), valores de $t(@)=.05$. Cumple
- ▶ NO genera Ecuaciones de constructo (*construct equations*).
- ▶ Covarianzas de las variables independientes (covariances among independent variables), valores de $t(@)=.05$, pero hay casos con correlación $>.6$ (nunca $>.8$) como $F3-F4=.766$. NO cumple (Nota: si cambia la matriz de COV a COR baja la correlación; de lo que se trata es generar la mayor correlación a partir de las covarianzas) y tiene potencial problema de validez discriminante, la cual se consigue cuando, si diversos instrumentos de medida están diseñados para medir distintas variables latentes, entonces las correlaciones entre dichos instrumentos son bajas. Una escala tiene, pues, validez discriminante solamente cuando no mide un constructo para el que no se diseñó. Generalmente se realiza a través de tres tests: 1. Test de las diferencias entre las Chi-cuadrado 2. Test del intervalo de confianza 3. Test de la varianza extraída.

Se deben aplicar entre cada par de factores Se suele aplicar sólo al par de factores que más problema pueden causar, es decir, los que tienen una covarianza más elevada entre sí. En este caso $F3-F4=.766$, siendo el Test de las diferencias entre las Chi cuadrado (Bagozzi y Phillips, 1982), de la siguiente forma:

- ▶ Se estima el modelo de medida pero fijando a 1 las covarianzas entre los dos factores cuya validez discriminante suscita dudas ($F3$ y $F4$).
- ▶ Se restan las Chi cuadrado de ambos modelos y los grados de libertad.
- ▶ El valor del estadístico (la resta) debe ser superior al valor crítico de la Chi cuadrado para ese nivel de grados de libertad.
- ▶ Esto querrá decir que el modelo en el que los factores están relacionados es mejor que aquel en que esos factores son el mismo (ausencia de validez discriminante).

Esto se verá más adelante al correr nuevamente el programa sin las variables problema.

- ▶ Solución estandarizada (standardized solutions), ¿todas las cargas factoriales son $>.6$? Bagozzi y Yi (1988) que cada carga debe ser factorial debe ser $>.6$. NO Cumple en $V14 (.592)$; $V15 (.595)$; $V16 (.601)$

- ▶ Prueba del multiplicador de Lagrange multivariado (Lagrange multiplier test), se observa que V14 y V16 tienen problemas.

CFA, SEM y EQS. Discusión rápida de resultados sin V14, V15 y V16

Por lo anterior, se sugiere eliminar las variables V14, V15 junto a V16 (recuerde colocar en el programa de entrada marketing mix.eqqs el símbolo ! a las ecuaciones que se encuentran en párrafo /EQUATIONS para V14, V15, V16, así como en el párrafo VARIANCES los errores E14, E15 y E16 y hacer revisión rápida (figura 91).

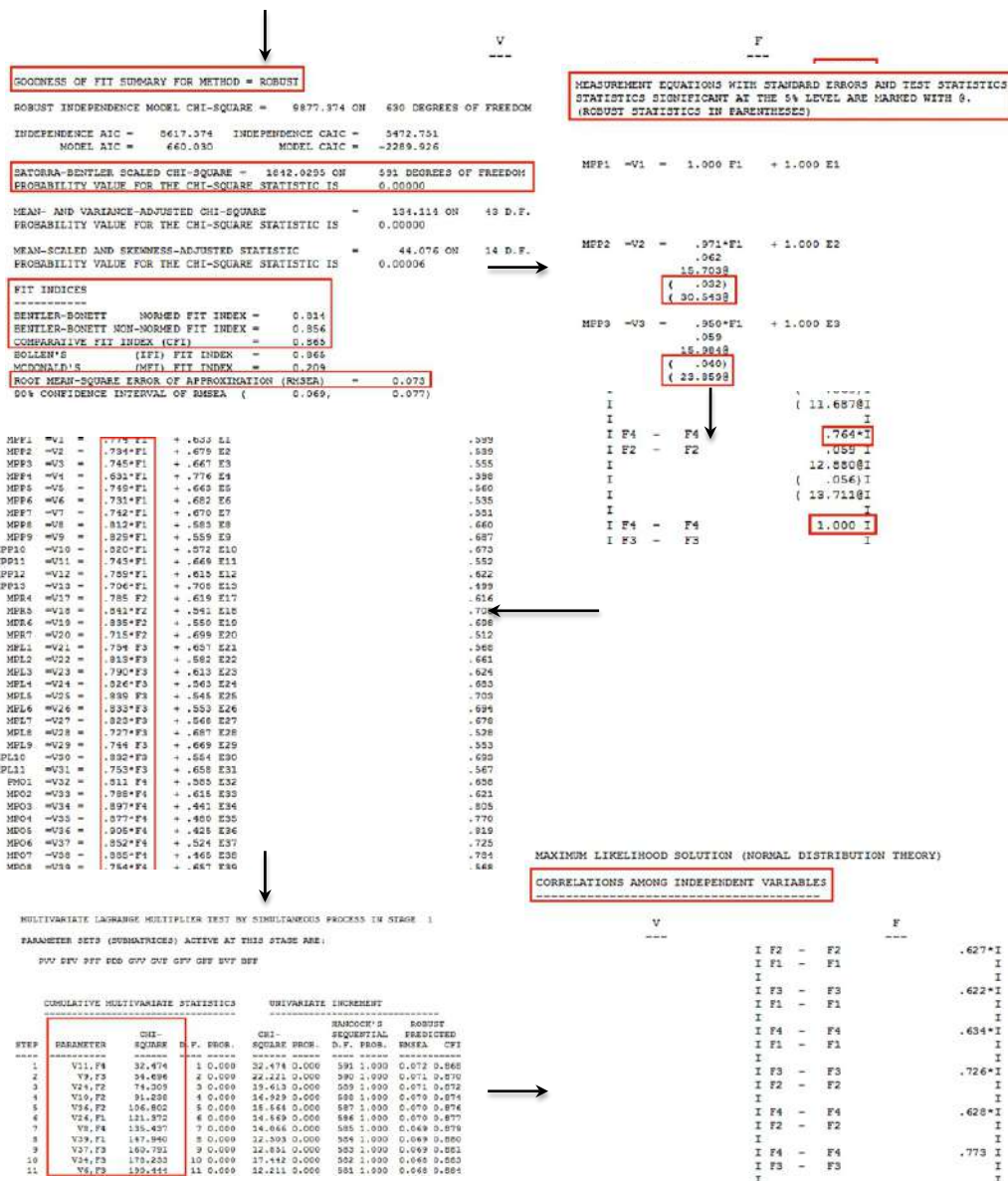
Nuevamente, de los reportes emitidos podemos observar y concluir de cada sección de reporte lo siguiente:

- ▶ Gráfico de distribución de residuos (*distribution of standardized residuals*). Mejor que el anterior aunque aún tiene problemas de datos alejados de la media. NO cumple.
- ▶ Índices de ajuste (*fit indices*) robustos (*robust*), valores a analizar:
Chi-cuadrado de Satorra-Bentler= 1842.0295; gl= 591
(Nota: $1842.0295/591=3.116>2$ Tendencia a que los índices de ajuste sean $<.9$);
valor $p<0.01$. NO cumple. NFI= .814; NNFI= .856; CFI= .865. NO cumple $>.9$ (para nueva escala). RMSEA=.073 $<.08$. Cumple.
- ▶ Ecuaciones de medición (measurement equations), valores de $t(@)=.05$. Cumple NO genera ecuaciones de constructo (construct equations).
- ▶ Covarianzas de las variables independientes (covariances among independent variables), valores de $t(@)=.05$, y dado que condicionamos $F3-F4=1$ el resto de los casos incrementan sus correlaciones aunque <1 (normalmente nunca $>.8$). Cabe destacar que sería un máximo de 3.9 variables=4 (10%) del total que podrían ser eliminadas. Se anima al lector a decidir cuál otra sería la candidata a partir de observar el multiplicador de Lagrange multivariado. Por lo pronto V14, V15, V16 hablan de precios; se tiene que revisar de la teoría el porque estos tres indicadores aportan muy poco a la escala; aparentemente generan confusión en los encuestados, por lo que también se debe revisar su redacción y/o presentación. Para fines didácticos, se acuerda que el modelo en este momento está ajustado.
- ▶ Solución estandarizada (standardized solutions), ;todas las cargas factoriales son $>.6$? Bagozzi y Yi (1988) que cada carga factorial debe ser $>.6$. Cumple.

CFA, SEM y EQS. Cálculo del alfa de Cronbach, sin V14, V15 y V16

Dado que se requiere el alfa de Cronbach por constructo (NO por modelo, que es lo que se reporta en EQS), requerimos calcularlo. Este cálculo deberá proceder de la base

FIGURA 91. CFA y SEM. Caso de estudio 3 sin V14, V15 y V16.



Fuente: EQS 6.2, con datos y adaptación propios.

de datos SPSS, en nuestro caso marketing mix.sav y proceder únicamente con las variables que quedaron del modelo estructura, o sea, NO incluir las variables V14, V15 ni V16. Es importante considerar que si está utilizando SPSS soportado con IOS o Windows (con un emulador como Parallels), ubique su archivo .SAV, de acuerdo con el sistema operativo con el cual va a trabajar. En nuestro caso es de forma directa en el sistema operativo IOS.

Así, la secuencia de comandos para ubicar y abrir el archivo marketing mix.sav, es:

- ▶ Teclar en ícono SPSS desde aplicaciones o barra de herramientas>Archivo->Abrir (ir a subdirectorio correspondiente)->Nombre de archivo: marketing mix.sav->Abrir (figura 92).
- ▶ Así, se procederá a realizar el cálculo de alfa de Cronbach, para cada uno de los ítems de los factores, con la siguiente secuencia inicial de comandos:
Analizar->Escala->Análisis de fiabilidad->Modelo: Alfa->Elementos: (F1= V1 hasta V13; F2=V17 hasta V20; F3=V21 hasta V31 y F4=V32 hasta V39, según el caso) >Aceptar (figura 93).

En consecuencia, el alfa de Cronbach por factor, es:

F1 (V2; V3; V4; V5; V6; V7; V8; V9; V10; V11; V12; V13)= .938.

F2 (V17; V18; V19; V20)= .862.

F3 (V21; V22; V23; V24; V25; V26; V27; V28; V29; V30; V31)= .940.

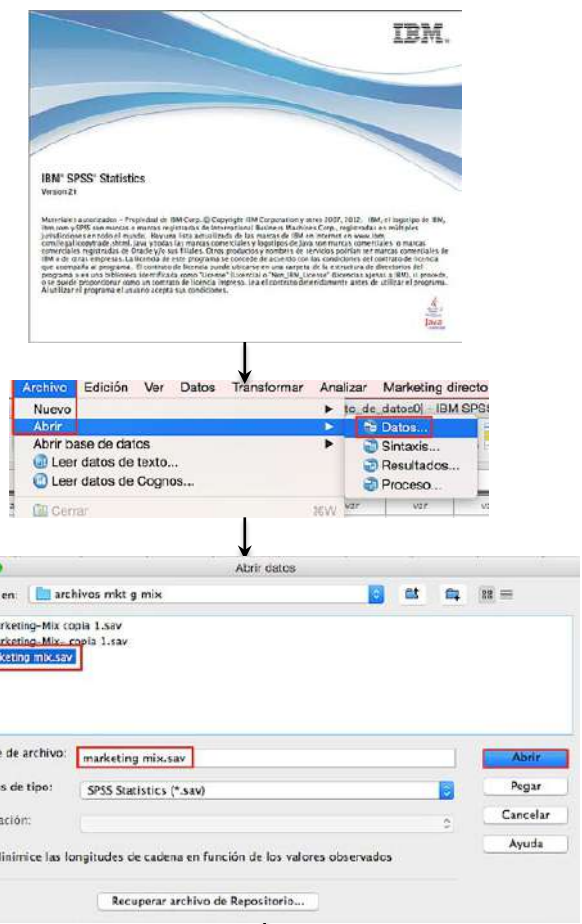
F4 (V32; V33; V34; V35; V36; V37; V38; V39)= .944.

CFA, SEM y EQS. Cálculo de IFC sin V14, V15 y V16

Normalmente, en la investigación científica los modelos teóricos desarrollados en la literatura no aparecen implicados un solo constructo, sino varios, por lo cual el alfa de Cronbach para cada factor por separado no considera la influencia sobre la fiabilidad del resto de constructos. Por esta razón, Fornell y Larcker (1981) proponen el cálculo del índice de la fiabilidad compuesta (IFC) para cada factor que, interpretándose exactamente igual que el alpha de Cronbach, sí que tiene en cuenta las interrelaciones. Por lo tanto, para determinarlo partimos del reporte de solución estandarizada (*standardized solution*) de la figura 91.

Calculando el reporte IFC, como se observa en la figura 94.

FIGURA 92. Secuencia de comandos. Apertura de archivo base de datos SPSS. Caso de estudio 3.



	MPP1	MPP2	MPP3	MPP4	MPP5	MPP6	MPP7	MPP8	MPP9	MPP10	MPP11	MPP12	MPP13	MPR1	MPR2	MPR3
1	4	3	3	4	4	4	4	4	4	3	4	4	5	4	5	5
2	4	3	3	4	4	4	4	4	4	3	4	4	5	5	5	5
3	1	1	4	5	5	5	5	5	5	5	5	5	5	5	5	5
4	1	1	1	1	2	2	2	2	2	4	3	1	5	2	1	1
5	3	3	4	3	5	5	5	5	5	5	2	5	5	5	1	5
6	5	4	4	5	3	5	5	5	5	4	2	4	5	4	1	4
7	3	4	4	5	3	5	5	5	5	4	2	4	5	4	3	5
8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	5
9	5	4	5	4	5	4	4	4	5	4	4	4	4	4	3	3
10	4	4	5	4	5	4	4	5	4	5	4	4	4	4	3	4

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 93. Secuencia de comandos. Cálculo alfa de Cronbach con SPSS. Caso de estudio 3.

The image shows the SPSS 'Análisis de fiabilidad' (Reliability) menu with the following options:

- Análisis de fiabilidad F1 [Conjunto_de_datos1] /libres/DCM/Desk
- Análisis de fiabilidad F2 [Conjunto_de_datos1] /libres/DCM/Desk
- Análisis de fiabilidad F3 [Conjunto_de_datos1] /libres/DCM/Desk
- Análisis de fiabilidad F4 [Conjunto_de_datos1] /libres/DCM/Desk

Below the menu, three summary tables are displayed, each representing a step in the process:

Step 1: Análisis de fiabilidad F1

Resumen del procesamiento de los casos	
Casos	%
Válidos	400 100,0
Excluidos ^a	0 ,0
Total	400 100,0

^a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
,933	13

Step 2: Análisis de fiabilidad F2

Resumen del procesamiento de los casos	
Casos	%
Válidos	400 100,0
Excluidos ^a	0 ,0
Total	400 100,0

^a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
,862	4

Step 3: Análisis de fiabilidad F3

Resumen del procesamiento de los casos	
Casos	%
Válidos	400 100,0
Excluidos ^a	0 ,0
Total	400 100,0

^a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
,940	11

Step 4: Análisis de fiabilidad F4

Resumen del procesamiento de los casos	
Casos	%
Válidos	400 100,0
Excluidos ^a	0 ,0
Total	400 100,0

^a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	N de elementos
,942	8

Fuente: EQS 6.2, con datos y adaptación propios.

MPP1	=V1	=	.779 F1	+	.633 E1	.599
MPP2	=V2	=	.734*F1	+	.679 E2	.539
MPP3	=V3	=	.745*F1	+	.667 E3	.555
MPP4	=V4	=	.631*F1	+	.776 E4	.398
MPP5	=V5	=	.749*F1	+	.663 E5	.560
MPP6	=V6	=	.731*F1	+	.682 E6	.535
MPP7	=V7	=	.742*F1	+	.670 E7	.551
MPP8	=V8	=	.812*F1	+	.583 E8	.660
MPP9	=V9	=	.829*F1	+	.559 E9	.687
MPP10	=V10	=	.820*F1	+	.572 E10	.673
MPP11	=V11	=	.743*F1	+	.669 E11	.552
MPP12	=V12	=	.789*F1	+	.615 E12	.622
MPP13	=V13	=	.706*F1	+	.708 E13	.499
MPR4	=V17	=	.785 F2	+	.619 E17	.616
MPR5	=V18	=	.841*F2	+	.541 E18	.708
MPR6	=V19	=	.835*F2	+	.550 E19	.698
MPR7	=V20	=	.715*F2	+	.699 E20	.512
MPL1	=V21	=	.754 F3	+	.657 E21	.568
MPL2	=V22	=	.813*F3	+	.582 E22	.661
MPL3	=V23	=	.790*F3	+	.613 E23	.624
MPL4	=V24	=	.826*F3	+	.563 E24	.683
MPL5	=V25	=	.839 F3	+	.545 E25	.703
MPL6	=V26	=	.833*F3	+	.553 E26	.694
MPL7	=V27	=	.823*F3	+	.568 E27	.678
MPL8	=V28	=	.727*F3	+	.687 E28	.528
MPL9	=V29	=	.744 F3	+	.669 E29	.553
MPL10	=V30	=	.832*F3	+	.554 E30	.693
MPL11	=V31	=	.753*F3	+	.658 E31	.567
PMO1	=V32	=	.811 F4	+	.585 E32	.658
MPO2	=V33	=	.788*F4	+	.615 E33	.621
MPO3	=V34	=	.897*F4	+	.441 E34	.805
MPO4	=V35	=	.877*F4	+	.480 E35	.770
MPO5	=V36	=	.905*F4	+	.425 E36	.819
MPO6	=V37	=	.852*F4	+	.524 E37	.725
MPO7	=V38	=	.885*F4	+	.465 E38	.784
MPO8	=V39	=	.754*F4	+	.657 E39	.568

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 94. CFA, SEM y EQS. Recopilación de información que calcula el IFC. Caso de estudio 3 sin V14, V15 y V16.

Constructo e indicadores	Carga factorial estandarizada L_{ij}	Varianza del término de error $\text{Var}(E_{ij}) = 1 - (L_{ij})^2$	Índice de fiabilidad compuesta IFC > .7 $\frac{\sum(L_{ij})^2}{[\sum(L_{ij})^2 + \sum \text{Var}(E_{ij})]}$
Producto (F₁)			
V ₁	.774	0.401	.945
V ₂	.734	0.461	
V ₃	.745	0.445	
V ₄	.631	0.602	
V ₅	.749	0.439	
V ₆	.731	0.466	
V ₇	.742	0.449	
V ₈	.812	0.341	
V ₉	.829	0.313	
V ₁₀	.820	0.328	
V ₁₁	.743	0.448	
V ₁₂	.789	0.377	
V ₁₃	.706	0.502	
Total	9.805	5.571	
Precio (F₂)			
V ₁₇	.785	0.384	.873
V ₁₈	.841	0.293	
V ₁₉	.835	0.303	
V ₂₀	.715	0.489	
Total	3.176	1.468	
Paza (F₃)			
V ₂₁	.754	0.431	.950
V ₂₂	.813	0.339	
V ₂₃	.790	0.376	
V ₂₄	.826	0.318	
V ₂₅	.839	0.296	
V ₂₆	.833	0.306	
V ₂₇	.823	0.323	
V ₂₈	.727	0.471	
V ₂₉	.744	0.446	
V ₃₀	.832	0.308	
V ₃₁	.753	0.433	
Total	8.734	4.048	
Promoción (F₄)			
V ₃₂	.811	0.342	.953
V ₃₃	.788	0.379	
V ₃₄	.897	0.195	
V ₃₅	.877	0.231	
V ₃₆	.905	0.181	
V ₃₇	.852	0.274	
V ₃₈	.885	0.217	
V ₃₉	.754	0.431	
Total	6.769	2.251	

Fuente: EQS 6.2, con datos y adaptación propios.

CFA, SEM y EQS. Cálculo de IVE sin V14, V15 y V16

Fornell y Larcker (1981) presentan el índice IVE como la relación entre la varianza que es capturada por un factor i en relación con la varianza total debida al error de medida de ese factor. Nótese que la única diferencia con el IFC es que cada carga factorial estandarizada es primero elevada al cuadrado antes de ser sumadas. Fornell y Larcker (*idem*) sugieren que es deseable que el constructo tenga valores de IVE $\geq .5$, es decir, que sea superior la varianza capturada por el factor que la debida al error de medida (figura 95).

CFA, SEM Y EQS. Análisis de la validez de la escala sin V14, V15 y V16

Recordando a Bohrnstedt (1976): “validez es el grado en que un instrumento mide el concepto bajo estudio”. Desafortunadamente, el concepto tiene múltiples dimensiones que deben explicarse y analizarse por separado.

Validez de contenido

La figura 91 contiene la sección de reporte índices de ajuste (fit indices) proporcionados por EQS. De su análisis se desprende que el valor de la Chi-cuadrado y el número de grados de libertad es superior a 2 ($1842.0295/591 = 3.11$). Esto nos indica que el modelo de Marketing mix no tiene un buen ajuste, en sus índices ya que varios serán $<.9$, sin embargo, esto no es limitante para analizar el resto de sats estadísticos. Entre el conjunto de indicadores de ajuste que proporciona EQS, se recomienda analizar aquellos no normalizados, dado que tienen menor tendencia a ofrecer resultados sesgados en casos de pequeñas muestras (Bentler y Wu, 1993). Se observa que, tanto el NNFI (Bentler y Bonnet, 1980) como el CFI (Bentler y Wu, 1993), que cumplen este requisito, tienen valores $>.8$ pero $<.9$, indicando un ajuste razonable. Se debe recordar que estos ajustes, ($>.8$ y $<.9$ escalas probadas; $>.9$ escalas por crear) aunque no cumplen el nivel deseado de ser $>.9$, es posible continuar con el análisis de los reportes estadísticos. Por ejemplo, RSMEA= 0.073, ya que debe ser $<.08$. De a misma figura 91, ver la sección de reporte ajuste de índices (*fit indices*).

Hasta este momento, se ha estimado el modelo de medida y se ha comprobado que ofrece un buen ajuste, para dar base de validez de contenido. Pues bien, sobre este modelo de medida, ya es posible analizar la validez convergente y la validez discriminante de las escalas implicadas. Para analizar la validez nomológica, será necesario estimar el modelo teórico.

FIGURA 95. CFA, SEM y EQS. Recopilación de información que calcula el IVE. Caso de estudio 3 sin V14, V15 y V16.

Constructo e indicadores	Cuadrado de la carga factorial estandarizada $(L_{ij})^2$	Varianza del término de error $Var(E_{ij}) = 1 - (L_{ij})^2$	Índice de varianza compuesta $IVE \geq .5$ $\frac{\sum (L_{ij})^2}{[\sum (L_{ij})^2 + \sum Var(E_{ij})]}$
Producto (F1)			
V ₁	0.599	0.401	.571
V ₂	0.539	0.461	
V ₃	0.555	0.445	
V ₄	0.398	0.602	
V ₅	0.561	0.439	
V ₆	0.534	0.466	
V ₇	0.551	0.449	
V ₈	0.659	0.341	
V ₉	0.687	0.313	
V ₁₀	0.672	0.328	
V ₁₁	0.552	0.448	
V ₁₂	0.623	0.377	
V ₁₃	0.498	0.502	
Total	7.429	5.571	
Precio (F2)			
V ₁₇	0.616	0.384	.633
V ₁₈	0.707	0.293	
V ₁₉	0.697	0.303	
V ₂₀	0.511	0.489	
Total	2.532	1.468	
Plaza (F3)			
V ₂₁	0.569	0.431	.632
V ₂₂	0.661	0.339	
V ₂₃	0.624	0.376	
V ₂₄	0.682	0.318	
V ₂₅	0.704	0.296	
V ₂₆	0.694	0.306	
V ₂₇	0.677	0.323	
V ₂₈	0.529	0.471	
V ₂₉	0.554	0.446	
V ₃₀	0.692	0.308	
V ₃₁	0.567	0.433	
Total	6.952	4.048	
Promoción (F4)			
V ₃₂	0.658	0.342	.718
V ₃₃	0.621	0.379	
V ₃₄	0.805	0.195	
V ₃₅	0.769	0.231	
V ₃₆	0.819	0.181	
V ₃₇	0.726	0.274	
V ₃₈	0.783	0.217	
V ₃₉	0.569	0.431	
Total	5.749	2.251	

Fuente: EQS 6.2, con datos y adaptación propios.

GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

ROBUST INDEPENDENCE MODEL CHI-SQUARE = 9877.374 ON 630 DEGREES OF FREEDOM
 INDEPENDENCE AIC = 8617.374 INDEPENDENCE CAIC = 5472.751
 MODEL AIC = 660.030 MODEL CAIC = -2289.926

**SATORRA-BENTLER SCALED CHI-SQUARE = 1842.0295 ON 591 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000**

MEAN- AND VARIANCE-ADJUSTED CHI-SQUARE = 134.114 ON 43 D.F.
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

MEAN-SCALED AND SKEWNESS-ADJUSTED STATISTIC = 44.076 ON 14 D.F.
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00006

FIT INDICES

 BENTLER-BONETT NORMED FIT INDEX = 0.814
 BENTLER-BONETT NON-NORMED FIT INDEX = 0.856
 COMPARATIVE FIT INDEX (CFI) = 0.865

BOLLEN'S (IFI) FIT INDEX = 0.865
 MCDONALD'S (RFI) FIT INDEX = 0.209

ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.073
 90% CONFIDENCE INTERVAL OF RMSEA (0.069, 0.077)

Fuente: EQS 6.2, con datos y adaptación propios.

**MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
 (ROBUST STATISTICS IN PARENTHESES)**

$$\text{MPP1} = \text{V1} = 1.000 \text{ F1} + 1.000 \text{ E1}$$

$$\begin{aligned} \text{MPP2} = \text{V2} &= .971 * \text{F1} + 1.000 \text{ E2} \\ &.062 \\ &15.703@ \\ &(\ .032) \\ &(\ 30.543@) \end{aligned}$$

$$\begin{aligned} \text{MPP3} = \text{V3} &= .950 * \text{F1} + 1.000 \text{ E3} \\ &.059 \\ &15.984@ \\ &(\ .040) \\ &(\ 23.859@) \end{aligned}$$

Fuente: EQS 6.2, con datos y adaptación propios.

Validez de constructo o de concepto

La validez de constructo o de concepto se divide en dos: validez convergente y validez discriminante. Una escala tiene validez de constructo, cuando tiene validez convergente y validez discriminante.

Validez convergente

La validez convergente existe cuando se emplean distintos instrumentos para medir un mismo constructo (distintos ítems para una misma variable latente) y estos instrumentos están fuertemente correlacionados. Si analizamos de la misma figura 91, las secciones de reporte las ecuaciones de medida (*measurement equations*), se deberá observar que las cargas factoriales sean < 1 significativas a 5%.

De la figura 91, la sección de reporte de soluciones estandarizadas (*standardized solutions*), así como para el modelo de medida observamos que todas las cargas factoriales son significativas, como se desprende de sus respectivas *t* para todas las variables. Asimismo, si calculáramos el promedio de los mismos veríamos que superan claramente .7.

Y de la misma figura 91 en su sección de prueba del multiplicador de Lagrange (*Lagrange multiplier test*) el ajuste del modelo mejoraría significativamente con la consideración, de tener suficiente marco teórico y el EFA validado para considerar cambiar, por ejemplo la V11 a F4 ambas como intento de mejora del modelo.

Cálculo de validez discriminante

Es aquí donde continuamos el estudio de la validez discriminante a partir del método.

1. Test de las diferencias entre las Chi cuadrado, que como se había indicado, se realiza de la siguiente forma:

- ▶ Se estima el modelo de medida pero fijando a 1 las covarianzas entre los dos factores cuya validez discriminante suscita dudas (F3 y F4).
- ▶ Se restan las Chi cuadrado de ambos modelos y los grados de libertad.

Modelo de medida.

Chi-cuadrado de Satorra-Bentler= 2179.7309; $gl= 698$.

Modelo de medida con COV F3-F4=1.

Chi-cuadrado de Satorra-Bentler= 1842.0295; $gl= 591$.

- ▶ El valor del estadístico (la resta) debe ser superior al valor crítico de la Chi cuadrado para ese nivel de grados de libertad. Esto querrá decir que el

MPP1	=V1	=	.774 F1	+	.633 E1	.599
MPP2	=V2	=	.734 F1	+	.679 E2	.539
MPP3	=V3	=	.745 F1	+	.667 E3	.555
MPP4	=V4	=	.631 F1	+	.776 E4	.398
MPP5	=V5	=	.749 F1	+	.663 E5	.560
MPP6	=V6	=	.731 F1	+	.682 E6	.535
MPP7	=V7	=	.742 F1	+	.670 E7	.551
MPP8	=V8	=	.812 F1	+	.583 E8	.660
MPP9	=V9	=	.829 F1	+	.559 E9	.687
MPP10	=V10	=	.820 F1	+	.572 E10	.673
MPP11	=V11	=	.743 F1	+	.669 E11	.552
MPP12	=V12	=	.789 F1	+	.615 E12	.622
MPP13	=V13	=	.706 F1	+	.708 E13	.499
MPP14	=V14	=	.706 F1	+	.708 E13	.499
MPP15	=V15	=	.706 F1	+	.708 E13	.499
MPP16	=V16	=	.706 F1	+	.708 E13	.499
MPP17	=V17	=	.785 F2	+	.619 E17	.616
MPP18	=V18	=	.841 F2	+	.541 E18	.708
MPP19	=V19	=	.835 F2	+	.550 E19	.698
MPP20	=V20	=	.715 F2	+	.699 E20	.512
MPL1	=V21	=	.754 F3	+	.657 E21	.568
MPL2	=V22	=	.813 F3	+	.582 E22	.661
MPL3	=V23	=	.790 F3	+	.613 E23	.624
MPL4	=V24	=	.826 F3	+	.563 E24	.683
MPL5	=V25	=	.839 F3	+	.545 E25	.703
MPL6	=V26	=	.833 F3	+	.553 E26	.694
MPL7	=V27	=	.823 F3	+	.568 E27	.678
MPL8	=V28	=	.727 F3	+	.687 E28	.528
MPL9	=V29	=	.744 F3	+	.669 E29	.553
MPL10	=V30	=	.832 F3	+	.554 E30	.693
MPL11	=V31	=	.753 F3	+	.658 E31	.567
MPL12	=V32	=	.811 F4	+	.585 E32	.658
MPO1	=V33	=	.788 F4	+	.615 E33	.621
MPO2	=V34	=	.897 F4	+	.441 E34	.805
MPO3	=V35	=	.877 F4	+	.480 E35	.770
MPO4	=V36	=	.905 F4	+	.425 E36	.819
MPO5	=V37	=	.852 F4	+	.524 E37	.725
MPO6	=V38	=	.885 F4	+	.465 E38	.784
MPO7	=V39	=	.754 F4	+	.657 E39	.568

Fuente: EQS 6.2, con datos y adaptación propios.

modelo en el que los factores están relacionados es mejor que aquel en que esos factores son el mismo (ausencia de validez discriminante).

La diferencia de las Chi-cuadrado = 337.704; $gl=17$ (figura 96). De la que se desprende el razonamiento siguiente:

Valor crítico con 17 gl :

$p < .005 = 35.7184$.

$p < .0025 = 37.9462$.

$p < .001 = 40.7911$, por lo tanto, a $337.704 < .001$. Se constata así que se tiene validez discriminante a más de 99%.

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1

PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:

PFV PFV PFF PDD GVV GVF GFV GFF BVF BFF

CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT						
STEP	PARAMETER	CHI-SQUARE	D.F.	PROB.	CHI-SQUARE	PROB.	HANCOCK'S SEQUENTIAL		ROBUST PREDICTED	
							D.F.	PROB.	RMSEA	CFI
1	V11, F4	32.474	1	0.000	32.474	0.000	591	1.000	0.072	0.868
2	V9, F3	54.696	2	0.000	22.221	0.000	590	1.000	0.071	0.870
3	V24, F2	74.309	3	0.000	19.613	0.000	589	1.000	0.071	0.872
4	V10, F2	91.238	4	0.000	16.929	0.000	588	1.000	0.070	0.874
5	V36, F2	106.802	5	0.000	15.564	0.000	587	1.000	0.070	0.876
6	V26, F1	121.372	6	0.000	14.569	0.000	586	1.000	0.070	0.877
7	V8, F4	135.437	7	0.000	14.066	0.000	585	1.000	0.069	0.879
8	V39, F1	147.940	8	0.000	12.503	0.000	584	1.000	0.069	0.880
9	V37, F3	160.791	9	0.000	12.851	0.000	583	1.000	0.069	0.881
10	V34, F3	178.233	10	0.000	17.442	0.000	582	1.000	0.068	0.883
11	V6, F3	190.444	11	0.000	12.211	0.000	581	1.000	0.068	0.884

Fuente: EQS 6.2, con datos y adaptación propios.

FIGURA 96. Tabla de distribución de Chi-cuadrado.

TABLA 3-Distribución Chi Cuadrado χ^2

P = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, v = Grados de Libertad

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735	0,7083	0,5707	0,4549
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996	1,8326	1,5970	1,3863
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831	2,9462	2,6430	2,3660
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377	4,0446	3,6871	3,3567
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0795	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731	5,1319	4,7278	4,3515
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948	6,2108	5,7652	5,3481
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061	7,2832	6,8000	6,3458
8	26,1239	23,7742	21,9549	20,0992	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094	8,3505	7,8325	7,3441
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060	9,4136	8,8632	8,3428
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971	10,4732	9,8922	9,3418
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7097	12,8987	12,1836	11,5298	10,9199	10,3410
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661	12,5838	11,9463	11,3403
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451	13,6356	12,9717	12,3398
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209	14,6853	13,9961	13,3393
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940	15,7332	15,0197	14,3389
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646	16,7795	16,0425	15,3385
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330	17,8244	17,0646	16,3382
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993	18,8679	18,0860	17,3379

Fuente: Levin y Rubin (2004), con datos y adaptación propios.

2. Test del intervalo de confianza (Anderson y Gerbing, 1988).

Se calcula un intervalo de confianza para la covarianza de \pm dos errores estándar alrededor de la estimación resultante del CFA.

Si incluye el 1 no se puede confirmar la validez discriminante, en caso contrario sí. En nuestro caso, de la figura 91, de la sección de reporte covarianzas entre las variables independientes (*covariances among independent variances*), de la relación más alta, se tiene:

```

I F2 - F2 .379*I
I F1 - F1 .048 I
I 7.930@I
I (.070)I
I (5.432@I
I
I F3 - F3 .613*I
I F1 - F1 .070 I
I 8.750@I
I (.075)I
I (8.196@I
I
I F4 - F4 .669*I
I F1 - F1 .078 I
I 8.535@I
I (.082)I
I (8.179@I
I
I F3 - F3 .377*I
I F2 - F2 .044 I
I 8.484@I
I (.056)I
I (6.712@I
I
I F4 - F4 .343*I
I F2 - F2 .045 I
I 7.632@I
I (.058)I
I (5.966@I
I
I F4 - F4 .766*I
I F3 - F3 .076 I
I 10.060@I
I (.072)I
I (10.622@I

```

Fuente: EQS 6.2, con datos y adaptación propios.

Así, el razonamiento de cálculo por intervalo, es:

F3-F4

Covarianza al cuadrado $(.766)^2 = .586$

Inferior $= .766 - 2 \times .076 = .614$

Superior $= .766 + 2 \times .076 = .918$

F2-F4

Covarianza al cuadrado $(.343)^2 = .118$

Inferior = $.343 - 2 \times .045 = .253$

Superior = $.343 + 2 \times .045 = .433$

F2-F3

Covarianza al cuadrado $(.377)^2 = .142$

Inferior = $.377 - 2 \times .044 = .289$

Superior = $.377 + 2 \times .044 = .465$

F1-F4

Covarianza al cuadrado $(.669)^2 = .447$

Inferior = $.669 - 2 \times .078 = .513$

Superior = $.669 + 2 \times .078 = .825$

F1-F3

Covarianza al cuadrado $(.613)^2 = .376$

Inferior = $.613 - 2 \times .070 = .473$

Superior = $.613 + 2 \times .070 = .753$

F1-F2

Covarianza al cuadrado $(.379)^2 = .144$

Inferior = $.379 - 2 \times .048 = .283$

Superior = $.379 + 2 \times .048 = .475$

Se comprueba, que al ser la correlación más alta entre factores es < 1 , por lo tanto, existe validez discriminante.

3. Test de la varianza extraída (Fornell y Larcker, 1981).

Se procede de forma sistemática, de la siguiente manera:

- ▶ Se calcula el IVE para cada uno de los factores implicados como se explicó al analizar la confiabilidad.
- ▶ Se compara el IVE con el cuadrado de las covarianzas entre los dos factores.
- ▶ Hay validez discriminante si los IVE de los dos factores superan el cuadrado de la covarianza.
- ▶ Cuadrado de la covarianza: $0.766^2 = 0.587$ (figura 97).

Cálculo de validez nomológica

Se debe recordar que, el modelo teórico que se estaba contrastando planteaba relaciones teóricas entre variables que podrían constatar la validez nomológica de ser ciertas. La mejor forma de hacerlo es estimar el modelo, no el de medida, como hemos hecho con el CFA, sino el teórico. Si con las modificaciones pertinentes para

FIGURA 97. Reporte tabla de validez discriminante del modelo teórico.

Factor	Producto (F1)	Precio (F2)	Plaza (F3)	Promoción (F4)
Producto (F1)	.571	.118	.376	.447
Precio (F2)	.283-.475	.633	.142	.118
Plaza (F3)	.473-.753	.289-.465	.632	.586
Promoción (F4)	.513-.825	.253-.433	.614-.918	.718

Nota: La diagonal representa el índice de varianza extraída (IVE), mientras que arriba de la diagonal se presenta la varianza (la correlación al cuadrado), abajo de la diagonal, se muestra la estimación de la correlación de los factores con un índice de fiabilidad de 95%.

Fuente: EQS 6.2, con datos y adaptación propios.

lograr un buen ajuste, los constructos cuya validez se está evaluando mantienen relaciones significativas previstas por la teoría, habremos establecido la validez nomológica (concurrente). El modelo teórico ha de tener un ajuste significativamente mejor que el de medida (lo comprobaremos con las diferencias entre las Chi cuadradas). De nos ser así, no podremos afirmar la validez nomológica. Así, como se había visto el cálculo de validez discriminante por el test de diferencias de Chi-cuadrados, se tiene:

Se restan las Chi cuadrado de ambos modelos y los grados de libertad.

Modelo de medida.

Chi-cuadrado de Satorra-Bentler= 2179.7309; gl= 698.

Modelo de medida con COV F3-F4=1.

Chi-cuadrado de Satorra-Bentler= 1842.0295; gl= 591.

El valor del estadístico (la resta) debe ser superior al valor crítico de la Chi-cuadrado para ese nivel de grados de libertad. Esto querrá decir que el modelo en el que los factores están relacionados es mejor que aquel en que esos factores son el mismo (ausencia de validez discriminante).

La diferencia de las Chi-cuadrado= 337.704; gl=17 (figura 100), de la que se desprende el razonamiento siguiente:

Valor crítico con 17 gl:

$p < .005 = 35.7184$

$p < .0025 = 37.9462$

$p < .001 = 40.7911$, por lo tanto, a $337.704 < .001$. Se constata así que se tiene validez discriminante a más de 99%, y que también el modelo teórico, ofrece un ajuste significativamente mejor que el modelo de medida, lo cual permite constatar la validez nomológica de las escalas.

FIGURA 98. Reporte tabla de consistencia interna y validez convergente del modelo teórico sin V1 y V3.

Factor	Item V	Carga factorial >0.6 (a)	Valor robusto t	Factor promedio de carga factorial	Fiabilidad de la escala		
					Alfa de Cronbach >=0.7 (b)	IFC >0.7 (b)	IVE >0.5 (c)
Producto (F ₁)	V1	.774	1.000a	.754	.938	.945	.571
	V2	.734	30.543				
	V3	.745	23.859				
	V4	.631	18.145				
	V5	.749	20.440				
	V6	.731	12.083				
	V7	.742	12.857				
	V8	.812	17.794				
	V9	.829	20.182				
	V10	.820	20.196				
	V11	.743	22.787				
	V12	.789	20.807				
	V13	.706	12.382				
Precio (F ₂)	V17	.785	1.000a	.794	.862	.873	.633
	V18	.841	19.147				
	V19	.835	15.118				
	V20	.715	12.427				
Plaza (F ₃)	V21	.754	1.000a	.794	.940	.950	.632
	V22	.813	26.640				
	V23	.790	21.947				
	V24	.826	27.235				
	V25	.839	25.555				
	V26	.833	24.449				
	V27	.823	26.245				
	V28	.727	19.193				
Promoción (F ₄)	V29	.744	22.333	.846	.944	.953	.718
	V30	.832	30.486				
	V31	.753	28.996				
	V32	.811	1.000a				
	V33	.788	24.686				
	V34	.897	25.868				
	V35	.877	25.295				
	V36	.905	25.817				
	V37	.852	21.998				
	V38	.885	25.344				
	V39	.754	22.254				

Conclusión
 Resultados: Chi-cuadrado= 1842.0295; gl= 591 p < .01; NFI = .814; NNFI = .856; CFI = .865; RMSEA = .073
 Conclusion: las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos. Esto es, no hay error en el ajuste del modelo. Por lo tanto, existe suficiente evidencia de validación convergente y confiabilidad, que justifica la consistencia interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999)”

Nota: *** Parámetros constreñidos a ese valor en el proceso de identificación = $p < 0.001$.

(a) De acuerdo con Bagozzi y Yi (1988). (b) De acuerdo con Hair et al. (1999).

© Índice de varianza extraída (IVE), de acuerdo con Fornell y Larcker (1981).

Fuente: EQS 6.2, con datos y adaptación propios.

CFA, SEM y EQS.

Reportando resultados: consistencia interna y validez convergente

En este punto, es importante tener una referencia del acomodo de los datos, como se sugiere en la figura 98.

donde las columnas significan:

Factores, listado de los involucrados en el estudio. En nuestro caso:

Productos (F1); Precio (F2); Plaza (F3) y Promoción (F4)

Ítem, consecutivo de las variables (V) no suprimidas mostradas, como: V1; V2; V3, etc.

Carga factorial > .6 (a) (figura 91), donde sólo se reportan los factores no suprimidos y con carga factorial >.6. Es posible que el investigador utilice escalas previas que si a su juicio, por el marco teórico, deban asignarse más ítems, es válido realizarlo siempre y cuando, se mencione que la nueva escala es una adaptación de la previa más la añadida como propuesta del investigador dadas las condiciones de la realidad en la que se aplica contra las que originalmente la escala se creó. En esta etapa es donde se valida o no la pertinencia de los nuevos ítems basados en el marco teórico planteado.

Valor Robusto t (figura 40 para mayor detalle).

Con lo anterior, es posible afirmar que si los valores de t (sección de reporte ecuaciones de medida (*measurement equations*) y ecuaciones de constructo (*construct equations*) para los factores, figura 91) son mayores (>) 2.58 errores estándar de la media, entonces se anotarán tres asteriscos (***) por carga factorial, lo cual indica que $p < .0001$. Así se tienen una seguridad que las cargas factoriales aportan sus valores a un 99% de confiabilidad.

Alfa de Cronbach, calculada y reportada de acuerdo con el apartado CFA, SEM y EQS. Cálculo del alfa de Cronbach, sin V14, V15 y V16, mediante uso de base de datos SPSS marketing mix.sav.

IFC. Se toman los datos de la figura 94 y es muy similar al alfa de Cronbach.

IVE. Se toman los datos de la figura 95.

Conclusión

Así afirmamos que para maximizar la mercadotecnia mix deberá ser atendido en mayor medida, por cada una de las cargas factoriales con valor absoluto alto y hacer una explicación de la realidad, con base en el marco teórico del porqué debe ser así (o no), según los argumentos del investigador. Todo lo anterior basado en las cargas factoriales encontradas con una probabilidad de equivocación de 1/100 o 99% de que lo afirmado ocurra (ver las ***).

Donde los renglones por debajo de las cargas factoriales, significan:

S-B χ^2 Índice Satorra Bentler Chi-cuadrado, es el recomendado cuando se tienen más de dos factores para generar datos robustos. Así, se explica la lectura reportada en la base de la figura 39:

Conclusión:

Resultados: Chi-cuadrado= 1842.0295; gl= 591 $p < .01$; NFI = .814; NNFI = .856; CFI = .865; RMSEA = .073.

“Las relaciones entre las variables y los factores tienen un ajuste adecuado a los datos. Esto es, no hay error en el ajuste del modelo. Aunque existe suficiente evidencia de validación convergente y fiabilidad, que justifica la fiabilidad interna de la escala aplicada (Nunnally y Bernstein, 1994; Hair et al., 1999), sí se tiene validez discriminante), por lo que es posible determinar un Reporte tabla de validez discriminante del modelo teórico”.

Así, una posible redacción de los resultados, sería:

“...Por otro lado, para evaluar la fiabilidad y validez de las escalas de medida se realizó un análisis factorial confirmatorio (AFC) utilizando el método de máxima verosimilitud con el software EQS 6.2 (Bentler, 2005; Brown, 2006; Byrne, 2006). Asimismo, la fiabilidad de las escalas de medida se evaluó a partir del coeficiente alfa de Cronbach y el índice de fiabilidad compuesta (IFC) (Bagozzi y Yi, 1988). Todos los valores de la escala excedieron el nivel recomendado de .70 para el alfa de Cronbach y el IFC que proporciona una evidencia de fiabilidad y justifica la fiabilidad interna de las escalas (Nunally y Bernstein, 1994; Hair et al., 1995)”.

Los resultados de la aplicación del AFC se presentan en la tabla 1 y sugiere que el modelo de medición proporciona un buen ajuste de los datos (S-BX2 = XXX.YYY; $df = XX$; $p = 0.000$; NFI > .9XXX; NNFI >.9XXX; CFI>.9XXX; y RMSEA <.08XXX). Como evidencia de la validez convergente, los resultados del AFC indican que todos los ítems de los factores relacionados son significativos ($p < 0.001$) y el tamaño de todas las cargas factoriales estandarizadas son superiores a 0.60 (Bagozzi y Yi, 1988). La tabla xxx muestra una alta consistencia interna de los constructos, en cada caso, la α de Cronbach excede el valor de .70 recomendado por Nunnally y Bernstein (1994). El IFC representa la varianza extraída entre el grupo de variables observadas y el constructo fundamental (Fornell y Larcker, 1981). Generalmente un IFC superior a .60 es considerado como deseable (Bagozzi y Yi, 1988), en este estudio este valor es superior. El índice de la varianza extraída (IVE) fue calculado para cada par de constructos, resultando un IVE superior a .5 en todos los factores (Fornell y Larcker, 1981)...”

Con estos resultados, ya estamos, en posibilidades de aprobar o desaprobar, hipótesis en función de las cargas factoriales encontradas y realizar la discusión del porqué está sucediendo (o no). Con base a filtros que aplique a su base de datos de marketing, sav; algunas de ellas podrían ser, como ejemplo:

H1: “Ha modificado líneas de productos o servicios por emergencia (V4), es el indicador más representativo del factor Producto (F2)”. Se desaprueba, ya que es incluso menos representativa (.631). La más representativa es V9: “Se centra en la maximización de las necesidades de sus clientes en cuanto a los requerimientos de sus productos”.

H2: “Los factores más importantes que considera la PyME manufacturera de Aguascalientes, en orden decreciente son: Precio, Promoción, Producto, Plaza en sus planes de administración de mercadotecnia”. Se desaprueba, ya que son: Promoción (.846), Precio (.794), Plaza (.794) y finalmente, Producto (.754).

H3. “Con base al modelo de marketing mix, aplicado en las PyMES manufactureras de Aguascalientes, el cual muestran que los esfuerzos en Precio (F2) y Plaza (F3), son muy similares y que lo más representativo es el esfuerzo en en Promoción (F4) y el menos, Producto (F1). Por lo tanto, se esperan los mismos resultados en un rango de cargas factoriales de +- 5% al aplicarlo al mismo sujeto de estudio en la Cd. de Guadalajara”. El resultado sería a revisar una vez obtenidos los datos del caso Guadalajara.

De ser posible, el resultado final de las cargas factoriales (figura 98), la puede complementar con un diagrama de trayectorias que muestre los datos en el mismo.

Por último, anexamos en la figura 99, una tabla de datos técnicos (otra investigación).

FIGURA 99. Tabla de datos técnicos, pruebas y valores utilizados en la investigación.

DATOS TÉCNICOS DE LA INVESTIGACIÓN			
CARACTERÍSTICAS	ESTUDIO		
Universo	400 especialistas PyMES manufactureras de Aguascalientes		
Alcance	zona metropolitana ciudad de Aguascalientes, Aguascalientes, México		
Método de recolección de datos	E-mail y encuestas directas		
Escala	Likert 5		
Pruebas de validez y confiabilidad	4 profesores (<i>visión académica</i>); 1 CEO pyme; 1 gerente back office; 1 gerente front office; 1 diseñador de software; 1 consultor de tecnologías de información (<i>visión de expertos</i>)		
Fecha de levantamiento de datos	Enero 2017- abril-2017		
Total de entrevistas efectivas	400 que incluyen: 100 empresarios Pyme; 100 administradores de proceso interno (<i>back office</i>) Pyme/ 100 administradores de servicio Pyme; 100 consultores que conforman la <i>visión de expertos</i>		
Pruebas aplicadas en este estudio	Descripción/valor	Autor	
1	Radio NC/VoQ= Número de casos (NC) / Variables del Cuestionario (VoQ)	NC= 400 (>=100 y <=1000) especialistas PyME. VoQ = 39 Radio NC/VoQ= 400/39=10.25>5 (>=5 recomendación de Hair, 2014)	Hair <i>et al.</i> , 2014

2	Análisis Factorial Confirmatorio (CFA) por el método de Máxima Verosimilitud y Análisis de Covarianza con uso del software EQS 6.2	Para verificar la fiabilidad y validez de las escalas de medición	Bentler, 2006; Brown, 2006; Byrne, 2006
3	Alfa de Cronbach (CHA) e Índice de Fiabilidad Compuesta (IFC)	CHA (por factor vía SPSS) y $IFC \geq 0.7$ para la fiabilidad de la escala de medición	Bagozzi y Yi, 1988; Nunnally y Bernestain, 1994; Hair <i>et al.</i> , 2014
4	Estimado de Mardia Normalizado (M)	$m > 5.00$. Distribuido como una unidad de variación normal tal que los valores grandes reflejan una curtosis positiva significativa y un gran número de valores negativos reflejan una curtosis negativa significativa. Bentler (2006) sugiere en la práctica, valores > 5.00 que son indicativos de datos, que no son distribuidos normalmente	Mardia, 1970; Mardia, 1974; Bentler, 2006; Byrne, 2006
5	Estadística Satorra-Bentler (S-B χ^2)	SB χ^2 . Al especificarse ME=ML, ROBUST, la salida genera una estadística de chi cuadrada robusta, llamada χ^2 . Esto permite minimizar los valores atípicos y lograr la bondad de ajuste	Satorra y Bentler, 1988
6	Índice de ajuste normalizado (NFI)	$NFI \geq 0.8$ y ≤ 0.89 . El índice ha sido utilizado por más de dos décadas por Bentler y Bonett (1980), como criterio práctico de elección, como lo demuestra en gran parte el estatus "clásico" actual del trabajo original de Bentler (1992), y Bentler y Bonett (1987), (citado por Byrne, 2006). Sin embargo, el NFI ha mostrado una tendencia a subestimar el ajuste en muestras pequeñas.	Bentler y Bonnet, 1980; Byrne, 2006
7	Índice de ajuste comparativo (CFI)	$CFI \geq 0.8$ y ≤ 0.89 . Bentler (1990, citado por Byrne, 2006) revisó la NFI para considerar el tamaño de la muestra y el CFI propuesto. Los valores para el NFI y el CFI oscilan entre cero y 1.00 y se derivan de la comparación entre los modelos de hipótesis e independencia. Como tal, cada uno proporciona una medida de la covariación completa en los datos. Aunque se consideró inicialmente que un valor > 0.90 era representativo de un modelo bien ajustado (véase Bentler, 1992, citado por Byrne, 2006), se ha recomendado un valor de corte revisado cercano a 0.95 (Hu y Bentler, 1999, citado por Byrne, 2006). Aunque ambos índices de ajuste se informan en la producción de EQS, Bentler (1990, citado por Byrne, 2006) sugirió que el CFI debería ser el índice de elección	
8	Índice No-Normalizado de ajuste (NNFI)	$NNFI \geq 0.8$ y ≤ 0.89 . Es una variante del NFI que tiene en cuenta la complejidad del modelo. Los valores para el NNFI pueden exceder los reportados para el NFI y también pueden caer fuera del rango cero a 1.00 (Byrne, 2006).	
9	Raíz cuadrática media de error por aproximación (RMSEA)	$RMSEA \geq 0.05$ y ≤ 0.08 . El RMSEA considera el error de aproximación en la población haciendo la pregunta: "¿Qué tan bueno es el modelo, con valores de parámetros desconocidos pero óptimamente elegidos, de tal forma que se ajuste en la matriz de covarianza de población de estar disponible?" (Browne y Cudeck, 1993: 137-138, citado por Byrne, 2006). Esta discrepancia, medida por la RMSEA, se expresa con el grado de libertad, por lo que es sensible al número de parámetros estimados en el modelo (es decir, la complejidad del modelo). Los valores inferiores a 0.05 indican un buen ajuste, y valores tan altos como 0.08 representan errores razonables de aproximación en la población (Brown y Cudeck, 1993, citado por Byrne, 2006).	Hair <i>et al.</i> , 2014; Byrne, 2006; Chau, 1997; Heck, 1998.

9	Raíz cuadrática media de error por aproximación (RMSEA)	Al dirigirse a Steiger (1990) se pide el uso de intervalos de confianza para evaluar la precisión de las estimaciones de RMSEA. EQS informa un intervalo del 90% alrededor del valor RMSEA. A diferencia de las estimaciones puntuales del ajuste del modelo (que no reflejan la imprecisión de la estimación), los intervalos de confianza pueden proporcionar esta información, proporcionando así al investigador más ayuda en la evaluación del ajuste del modelo.	Hair <i>et al.</i> , 2014; Byrne, 2006; Chau, 1997; Heck, 1998.
10	Validez convergente (cv)	Todos los ítems de los factores relacionados son significativos ($p < 0.001$), el tamaño de todas las cargas factoriales son mayores a 0.60 (Bagozzi y Yi, 1988). Es el grado en el cual, las diferentes métricas de evaluación concurren en sus mediciones del mismo constructo. Idealmente, estos valores deben de ser moderadamente altos (Byrne, 2006)	Bagozzi y Yi, 1988; Byrne, 2006
11	Índice de varianza extraída (IVE)	$IVE > 0.50$ En todos los factores pareados como constructos a través de una matriz de relaciones, donde la diagonal representa al IVE, mientras arriba de la diagonal se representa la varianza (la correlación al cuadrado); debajo de la diagonal, se encuentra el estimado de la correlación de los factores con un intervalo de confianza de 95%. Vea la tabla de <i>validez discriminante del modelo teórico</i> .	Fornell y Larcker, 1981
12	Validez discriminante (DV)	DV, es la medida en que los métodos de evaluación independientes divergen en la medición con diferentes rasgos, idealmente estos valores deberían demostrar una convergencia mínima (Byrne, 2006). DV se proporciona en dos formas: en primer lugar, con un intervalo de 95% de fiabilidad, ninguno de los elementos individuales de la matriz de correlación de factores latentes contiene 1.0 (Anderson y Gerbing, 1988). En segundo lugar, el IVE entre cada par de factores es mayor que su IVE correspondiente (Fornell y Larcker, 1981). Por lo tanto, sobre la base de estos criterios, las diferentes mediciones efectuadas en la escala muestran suficiente evidencia de fiabilidad, CV y DV.	Byrne, 2006; Anderson y Gerbing, 1988; Fornell y Larcker, 1981
13	Validez nomológica (NV)	Se prueba con el chi cuadrado, a través del cual se comparó el modelo teórico con el modelo ajustado. Los resultados indican que si no hay diferencias significativas, son buenos modelos teórico en la explicación de las relaciones observadas entre las construcciones latentes	Anderson y Gerbing, 1988; Hatcher, 1994

Fuente: EQS 6.2, con datos y adaptación propios.

Bibliografía

- Anderson, J. C. y Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, 1 (3), 411-423.
- Allen, S. J. y Hubbard, R. (1986). Regression equations of the latent roots of random data correlation matrices with unities on the diagonal. *Multivariate Behavioral Research*, 21, 393-398.
- Anastasi, A. y Urbina, S. (1998). *Psychological testing*. Englewood Cliffs, NJ: Prentice-Hall.
- Arbuckle, J. L. (2003). *Amos 5.0 [Computer software]*. Chicago, IL: SPSS.

- Bagozzi, Richard P. y Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science Spring*, 16 (1), 74-94.
- Bagozzi, R. P. y Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling*, 1 (1), 35-67. .
- Bagozzi, R. P., Yi, Y. y Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36 (3), 421-458.
- Bearden, W. O., Hardesty, D. y Rose, R. (2001). Consumer self-confidence: refinements in conceptualization and measurement. *Journal of Consumer Research*, 28 (june), 121-134.
- Bearden, W. O. y Netemeyer, R. G. (1998). *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research*. Thousand Oaks, CA: Sage.
- Bearden, W. O., Netemeyer, R. G. y Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, 15 (march), 473-481.
- Bearden, W. O., Sharma, S. y Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19 (november), 425-430.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107 (2): 238-246
- (2006). EQS 6.1. Structural Equations Program Manual. CA: Multivariate Software Inc.
- y Bonnet, D. G. (1980). Significance tests and goodness of fit in analysis of covariance structures. *Psychological Bulletin*, 88 (3), sep.-dec., pp. 588-606.
- y Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods y Research*, 16 (1), 78-117.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley y Sons.
- y Lennox, R. (1991). Conventional wisdom on measurement: A structural equations perspective. *Psychological Bulletin*, 110, 305-314.
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 3, 291-294.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Browne, M. W. y Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A.

- Bollen y J. S. Long (Eds.). *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Bruner, G. y Hensel, P. (1997). *Marketing scales handbook: A compilation of multi-item measures*. 2nd ed. Chicago: American Marketing Association.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- (2005a). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*, 85, 17-30.
- (2005b). Factor analysis confirmatory. In B. S. Everitt y D. C. Howell (Eds.). *Encyclopedia of Statistics in Behavioural Science*, pp. 599-606. London, UK: Wiley.
- (2006). *Structural Equation Modeling with EQS, basic concepts, applications, and programming*. London: LEA Publishers.
- Calder, B. J., Phillips, L. W. y Tybout, A. M. (1982). The concept of external validity. *Journal of Consumer Research*, 9 (december), 240-244.
- Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15, 546-553.
- y Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carver, C. S. (1989). How should multi-faceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology*, 56 (4), 577-585.
- Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 174-243). Chicago: Rand McNally.
- Chau, P. Y. K. (1997). Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Sciences*, 28 (2), 309-334.
- Chou, C.-P., Bentler, P. M. y Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.
- Hu, L.-T., Bentler, P. M. y Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16 (february), 64-73.
- Churchill, G. A. y Iacobucci, D. (2002). *Marketing research methodological Foundations*. 8th ed. Fort Worth, TX: Harcourt College Publishers.

- Churchill, G. A. y Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21(november), 360-375.
- Clark, L. A. y Watson, D. (1995). Constructing validity: Basic issues in scale development. *Psychological Assessment*, 7(3), 309-319.
- Cliff, N. (1988). The eigenvalue-greater-than-one rules and reliability of components. *Psychological Bulletin*, 103(2), 276-279.
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56, 754-761.
- Cook, T. D. y Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and application. *Journal of Applied Psychology*, 78, 98-104.
- Crocker, L. y Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart, y Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 31, 93-96.
- y Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Diamantopoulos, A. y Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 36, 269-277.
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20 (september), 303-315.
- Floyd, F. J. y Widaman, K. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.
- Fornell, C. y Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (february), 39-50.
- Friestad, M. y Wright, P. (2001). *Pre-adult education on marketplace persuasion tactics: Integrating marketplace literacy and media literacy*. Unpublished faculty working paper, Eugene: University of Oregon.
- Ganster, D. C., Hennessey, H. W. y Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, 26 (june), 321-331.

- Gerbing, D. W. y Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, 11 (june), 572-580.
- (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 25 (may), 186-192.
- Hair, J. J., Anderson, R., Taham, R. y Black, W. (1999). *Análisis multivariante*. Madrid: Prentice Hall Iberia.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9 (june), 139-164.?
- Hayduk, L. A. (1996). LISREL: *Issues, debates, and strategies*. Baltimore, MD: Johns Hopkins University Press.?
- Heck, R. H. (1998). Factor analysis: exploratory and confirmatory approaches in Marcoulides, G. A. Mahwah. *Modern methods for business research*. NJ: Lawrence Erlbaum Associates Inc.
- Haynes, S., Nelson, N. K. y Blaine, D. (1999). Psychometric issues in assessment research. In P. C. Kendall, J. N. Butcher y G. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (pp. 125-154). New York: John Wiley y Sons.
- Haynes, S., Richard, D. C. y Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.
- Herche, J. y Engellend, B. (1996). Reversed polarity items and scale dimensionality. *Journal of the Academy of Marketing Science*, 24 (4), 366-374.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-186.
- Hoyle, R. (1995). *Structural equation modeling: Issues and applications*. Newbury Park, CA: Sage.
- Hu, L. y Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 (1), 1-55.
- Hull, J. G., Lehn, D. A. y Tedlie, J. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology*, 61 (6), 932-945.
- Iacobucci, D., Ostrom, A. y Grayson, K. (1995). Distinguishing service quality and customer satisfaction: The voice of the consumer. *Journal of Consumer Psychology*, 4 (3), 277-303.

- Jarvis, W. B. G. y Petty, R. E. (1996). The need to evaluate. *Journal of Personality and Social Psychology*, 70 (1), 172-194.
- Jöreskog, K. y Sörbom, D. (1989). LISREL7: *A guide to the program and applications*, 2nd ed. Chicago: SPSS.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen J. S. Long (Eds.). *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage
- Kaplan, R. M., y Saccuzzo, D. P. (1997). *Psychological testing: Principles applications, and issues*. 4th ed. Pacific Grove, CA: Brooks/Cole.
- Kelley, J. R. y McGrath, J. E. (1988). *On time and method*. Beverly Hills, CA: Sage.
- Kenny, D. A. y Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 122, 165-172.
- Kumar, A. y Dillon, W. R. (1987). Some further remarks on measurement-structure interaction and the unidimensionality of constructs. *Journal of Marketing Research*, 24 (november), 438-444.
- Lastovicka, J. L., Bettencourt, L. A., Hughner, R. S. y Kuntze, R. J. (1999). Lifestyle of the tight and frugal: Theory and measurement. *Journal of Consumer Research*, 26 (june), 85-98.?
- Levin, R. I. y Rubin, D. S. (2004) *Estadística para administración y economía*. 7a. México: Pearson Prentice Hall.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.?
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya*, B36, 115-128.
- Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data. In A. E. Raftery (Ed.). *Sociological Methodology 1997* (pp. 453-481). Washington, DC: American Sociological Association.
- MacCallum, R. C. y Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114, 533-541.?
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- y Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, 105, 430-445.

- Mejía-Trejo, J. (2017a). Banda ancha digital e innovación abierta: Primeras impresiones en el sector de tecnologías de la información. From *Revista del Centro de Investigación*. Universidad Lasalle: <http://ojs.dpi.ulsal.mx/index.php/rci/article/view/1146>
- (2017b). The determinant factors of open business model. From *Nova Scientia*: <http://novascientia.delasalle.edu.mx/ojs/index.php/Nova/article/view/710/334>
- (2017c). *Las ciencias de la administración y el análisis multivariante. Proyectos de investigación, análisis y discusión de resultados*. Tomo 2: Las técnicas interdependientes. Guadalajara: Universidad de Guadalajara, disponible en http://www.cucea.udg.mx/include/publicaciones/consultar/info_publicacion.php?id=237.
- (2017d). *Las ciencias de la administración y el análisis multivariante. Proyectos de investigación, análisis y discusión de resultados*. Tomo 1: Las técnicas dependientes. Guadalajara: Universidad de Guadalajara, disponible en http://www.cucea.udg.mx/include/publicaciones/consultar/info_publicacion.php?id=236.
- Mejía-Trejo, J., J., S.-G. y Vázquez-Ávila, G. (2015). *Empirical model for mobile learning and their factors. Case study: Universities located at Metropolitan Zone of Guadalajara, México*. From *Apertura. Revista de Innovación Educativa?*, disponible en <http://www.udgvirtual.udg.mx/apertura/index.php/apertura/index.php/apertura/article/view/725>
- Mejía-Trejo, J., Sánchez-Gutiérrez, J. y Maldonado-Guzmán, G. (2016). The customer knowledge management and innovation. *Contaduría y Administración* (61), 456-477.
- Mejía-Trejo, J., Sánchez-Gutiérrez, J. y Vázquez-Ávila, G. (2014). Culture of competitiveness in the new global context. In J. Sánchez-Gutiérrez, J. Vargas-Barraza, y J. Mejía-Trejo, *Modeling MLQ5X for Innovation and Value Creation* (vol. 1, pp. 1-18). Guadalajara: Universidad de Guadalajara.
- Messick, S. (1993). Validity. In R. L. Linn (Ed.). *Educational measurement*. 2nd ed., pp. 105-146). Phoenix, AZ: American Council on Education and the Oryx Press.
- Mick, D. G. (1996). Are studies of dark side variables confounded by socially desirable responding? The case of materialism. *Journal of Consumer Research*, 23 (september), 106-119.?
- Nederhof, A. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15 (july-september), 263-280.

- Netemeyer, R. G., Boles, J. S. y McMurrian, R. C. (1996). Development and validation of Work-Family Conflict and Family-Work Conflict scales. *Journal of Applied Psychology*, 81 (4), 400-410.
- Netemeyer, R. G., Durvasula, S. y Lichtenstein, D. R. (1991). A cross-national assessment of the reliability and validity of the CETSCALE. *Journal of Marketing Research*, 28 (august), 320-327.?
- Netemeyer, R. G., Pullig, C. y Bearden, W. O. (2002). Observations on some key psychometric properties of paper-and-pencil measures. In A. G. Woodside y E. M. Moore (Eds.), *Essays by distinguished marketing scholars of the Society for Marketing Advances* (pp. 115-138). Amsterdam: JAI.
- Netemeyer, R. G., Bearden, W. O, Sharma, S. (2003) *Scaling Procedures. Issues and applications*. USA: Sage Publications.
- Neuberg, S. L., West, S. G., Thompson, M. M. y Judice, T. N. (1997). On dimensionality, discriminant validity, and the role of psychometric analyses in personality theory and measurement: Reply to Kruglanski et al. (1997) defense of the Need for Closure Scale. *Journal of Personality and Social Psychology*, 73, 1017-1029.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22, 287-293.
- Nunnally, J. y Bernstein, I. H. (1994). *Psychometric theory*. 3rd ed. New York: McGraw-Hill.
- O'Rourke Norm y Hatcher Larry (2013). *Step by step approach to using the SAS System for factor analysis and structural equation modeling*. 2nd ed. Cary, NC, USA: SAS Institute Inc.
- OECD Organisation for Economic Co-operation and Development (2003). Knowledge Management. Measuring Knowledge Management in the Business Sector. First Steps. In O. f.-O. Development (Ed.).
- (2005). Oslo Manual. Guidelines for Collecting and Interpreting Innovation Data. In O. f.-O. Development (Ed.).
- Osgood, C. E. y Tannenbaum, P. H. (1955). The principle of congruence in the prediction of attitude change. *Psychological Bulletin*, 62, 42-55.
- Park, C. W., Mothersbaugh, D. L. y Feick, L. (1994). Consumer knowledge assessment. *Journal of Consumer Research*, 21 (june), 71-82.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46 (3), 598-609.
- (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver y L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York: Academic Press.

- (1993). *The Balanced Inventory of Desirable Responding: Reference manual for BIRD Version 6*. Toronto: Multi-Health Systems.
- Perreault, W. D., Jr. y Leigh, L. E. (1989). Reliability of nominal data based on qualitative judgments. *Journal of Marketing Research*, 26 (may), 135-148.
- Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, 18 (may), 133-145.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21 (september), 381-391.
- Pitonyak, D. (2010). *From The importance of belonging*, disponible en <http://www.hsri.org/files/NCRI/Belonging.pdf>. Recuperado el 2 de noviembre de 2017.
- PNC Premio Nacional de Calidad México (2017). *Modelo nacional para la competitividad*. Disponible en: www.pnc.org.mx/e-book-modelo-nacional-para-la-competitividad/. Recuperado el 2 de noviembre de 2017.
- Price, J. P. y Mueller, C. W. (1986). *Handbook of organizational measurement*. Marshfield, MA: Pittman.
- Richins, M. L. (1983). An analysis of consumer interaction styles in the marketplace. *Journal of Consumer Research*, 10 (june), 73-82.
- Richins, M. L. y Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, 19 (december), 303-316.
- Robinson, J. P., Shaver, P. R. y Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver y L. S. Wrightsman (Eds.). *Measures of personality and social psychological attitudes* (pp. 1-15). San Diego: Academic Press.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosser, J. R. (2001). *The C-OAR-SE procedure for scale development in marketing*. Faculty working paper. Australia: University of Wollongong, Australia.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54, 131-151.
- Satorra, A. y Bentler, P. M. (1988). Scaling corrections for chi square statistics in covariance structure analysis. *American Statistical Association*, pp. 308-313.
- (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye C. C. Clogg (Eds.). *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

- (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Saxe, R. y Weitz, B. A. (1982). The SOCO scale: A measure of the customer orientation of salespeople. *Journal of Marketing Research*, 19 (august), 343-351.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350-353.
- Schumacker, R. E. y Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.?
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley y Sons.
- Smith, G. T. y McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300-308.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.
- Steiger, J. H. y Lind, J. C. (1980, may). *Statistically-based tests for the number of common factors*. Paper presented at the meeting of the Psychometric Society, Iowa City.
- I. A. Sumner, G. A. (1906). *Folkways*. New York: Ginn Custom Publishing.
- Tian, K. T., Bearden, W. O. y Hunter, G. L. (2001). Consumers' need for uniqueness: Scale development and validation. *Journal of Consumer Research*, 28 (june), 50-66.
- Tian, K. T., Bearden, W. O. y Manning, K. C. (2002). *Agents' socially desirable responding: Scale development and validation*. Faculty working paper, University of South Carolina, Columbia.
- Trochim, A. (2002). *Construct validity*. Retrieved november 19, 2002, from <http://trochim.human.cornell.edu/kb/constval.htm>.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26.
- Wright, P. L. (1975). Factors affecting cognitive resistance to advertising. *Journal of Consumer Research*, 2 (june), 1-9.

*Creación de Escalas
en las Ciencias
de la Administración*

se terminó de imprimir
en marzo de 2018
en los Talleres Gráficos
de Imprelibros BM
Brillante No.913
Col. Alcalde Barranquitas
Guadalajara, Jal.
Tel.: 3613-8426

El tiraje fue de 100 ejemplares

imprelibrosbm@gmail.com





Obtener reconocimiento y credibilidad en resultados de investigación científica depende de qué estrategia esclarece y justifica mejor, la técnica de medición empleada. Así, es posible plantear diversas estrategias que permiten desarrollar y afinar las mediciones, sin embargo, su verdadero impacto dependerá del tipo de fenómeno científico que esté siendo medido, tanto con variables observables directamente como de sus relaciones subyacentes.

Es así que las ciencias de la administración, deberán centrar su interés en el planteamiento, diseño, desarrollo y validación de constructos tanto de relaciones directas como subyacentes. Particularmente, éste último tipo de relaciones implica constructos que al representar abstracciones sólo pueden evaluarse de manera indirecta.

La evaluación indirecta implica el diseño y uso de múltiples elementos (en nuestro caso, indicadores) que miden al constructo, es decir, plantean la escala para medir. Así, partiendo de un problema de investigación determinado y caracterizados los factores, variables e indicadores que mejor la describan en forma de constructos, el objetivo principal de este documento es el de plantear y diseñar las relaciones de dichos factores, variables e indicadores para discutir cómo medirlos y validarlos a través de la creación de escalas.



CUCEA

978-607-547-184-6



UNIVERSIDAD DE GUADALAJARA

Centro Universitario de Ciencias Económico Administrativas

Secretaría Académica

Unidad de producción editorial

A QUIEN CORRESPONDA

Por este conducto me permito informar que la obra titulada *Creación de escalas en las ciencias de la administración* fue evaluada favorablemente para su publicación por dictaminadores externos.

La citada obra se encuentra actualmente en proceso de edición; su aparición al público está programada para el próximo mes de abril del presente.

Se extiende la presente a solicitud del interesado para los fines legales y administrativos que a él convengan y quedo a sus órdenes para cualquier aclaración al respecto.

Sin otro particular, reciba un cordial y respetuoso saludo.

ATENTAMENTE
"PIENSA Y TRABAJA"

Zapopan, Jalisco, a 24 de noviembre del 2017

Miguel Ángel Serrano Núñez

UNIDAD DE PRODUCCION EDITORIAL