

## Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico

Sample sizes using structural equation modeling with latent variables: A practical method

Volumen 17, Número 1

Enero-Abril

pp. 1-34

Este número se publicó el 1° de enero de 2017

DOI: <http://dx.doi.org/10.15517/aie.v17i1.27294>

Tomas Vargas Halabí

Ronald Mora-Esquivel

*Revista indizada en* [REDALYC](#), [SCIELO](#)

*Revista distribuida en las bases de datos:*

[LATINDEX](#), [DOAJ](#), [REDIB](#), [IRESIE](#), [CLASE](#), [DIALNET](#), [SHERPA/ROMEO](#),  
[QUALIS-CAPES](#), [MIAR](#)

*Revista registrada en los directorios:*

[ULRICH'S](#), [REDIE](#), [RINACE](#), [OEI](#), [MAESTROTECA](#), [PREAL](#), [CLACSO](#)

---

Los contenidos de este artículo están bajo una licencia [Creative Commons](#)



## Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico

Sample sizes using structural equation modeling with latent variables: A practical method

Tomas Vargas Halabí<sup>1</sup>  
Ronald Mora-Esquivel<sup>2</sup>

**Resumen:** La técnica de modelaje mediante ecuaciones estructurales resulta muy útil para investigadores educativos para trabajar diversos constructos latentes, en forma simultánea, y someter a prueba modelos que clarifiquen diferentes tipos de efectos. Empero, la literatura aún no ha brindado una respuesta adecuada en lo referente al tamaño de muestra procedente para utilizar dicha técnica. Asimismo, a pesar de que existe una diversidad de criterios para su estimación, aún no ha proporcionado un método práctico que permita compendiarlos y facilitar la argumentación del tamaño apropiado de esta. Por ello, el presente artículo tiene como objetivo ofrecer una guía práctica a investigadores educativos para utilizar la diversidad de criterios y justificar el cálculo apropiado del tamaño de muestra en ecuaciones estructurales, mediante la planificación a priori de modelos de medida y modelos estructurales. Se recurre a un método cuantitativo para determinar los insumos con base en diagramas de sendero de modelos a priori de una investigación, así como interfaces disponibles en internet para calcular los tamaños de muestra. Los datos se resumen en forma tabulada para efectos comparativos. La aplicación del método constata su utilidad para que el investigador defina un umbral de casos que permita satisfacer distintos criterios, con respecto a depender de una única regla o criterio para justificar el tamaño de la muestra apropiado. Asimismo, revela la ventaja que podría representar el uso de parcelas en la planificación a priori de modelos con esta técnica.

**Palabras clave:** análisis estadístico, modelos, ecuaciones estructurales, tamaño de muestra.

**Abstract:** Structural equation modeling is a useful technique for educational researchers to testing hypotheses about relations among latent constructs and to clarify different types of effects. However, the estimation of sample size to run properly this technique has not yet provided an adequate response in the literature. Although a variety of criteria for its estimation is available, the literature has not yet provided a practical method to summarize them and to provide an argument of the appropriate size. Accordingly, this article aims to provide a practical guidance to educational researchers to use different criteria and justify the appropriate sample size calculation in structural equations modeling by planning, a priori, measurement and structural models. To calculate sample sizes a quantitative method is utilized to determine inputs based on path diagrams of a priori models of research, as well as interfaces available online. Data are summarized in tabular form for comparative purposes. The method has been useful to define a minimum number of cases to satisfy different criteria and has shown its advantage than relying on a single rule or criterion to justify the appropriate sample size. It also reveals the advantage using parcels in planning prior models with this technique.

**Keywords:** statistical analysis, models, structural equations, sample size.

---

<sup>1</sup> Profesor de la Escuela de Psicología, Universidad de Costa Rica.  
Dirección electrónica: [tomas.vargas.ucr@gmail.com](mailto:tomas.vargas.ucr@gmail.com)

<sup>2</sup> Docente e Investigador del Centro de Investigaciones en Administración, Economía y Gestión Tecnológica (CIADEG-TEC) adscrito a la Escuela de Administración de Empresas del Instituto Tecnológico de Costa Rica. Dirección-electrónica: [rmora@itcr.ac.cr](mailto:rmora@itcr.ac.cr)

**Artículo recibido:** 24 de mayo, 2016

**Enviado a corrección:** 26 de julio, 2016

**Aprobado:** 21 de noviembre, 2016

## 1. Introducción

La investigación educativa puede beneficiarse de manera importante de los modelos de ecuaciones estructurales (SEM por sus siglas en inglés), ya que le permite abordar preguntas de investigación más complejas y someter a prueba modelos con múltiples variables en una sola investigación (Teo, Tsai y Yang, 2013). Como muestra de ello, Khine (2013) ofrece ejemplos del uso de esta técnica en la investigación educativa; en concreto, el efecto en el aprendizaje de los ambientes de clases, la motivación, el involucramiento, la autorregulación, la autoeficacia, la actitud hacia los contenidos, las destrezas de pensamiento, la relación entre maestros y estudiantes e incluso el clima organizativo en las escuelas. Sin duda SEM es una técnica que permite trabajar con una gran diversidad de constructos relevantes para el campo educativo de forma simultánea, controlando el error de medida, de forma que se someten a prueba modelos que clarifiquen los diversos tipos de efectos (directos, mediadores, de correlación e interacción) de los distintos factores sobre el aprendizaje.

De igual forma, se ha señalado en la literatura, de forma reiterada, que SEM requiere muestras grandes (Kline, 2011; Schumacker y Lomax, 2010; Ullman, 2013) con el fin de evitar estimaciones imprecisas en los errores estándar (Kline, 2011) e índices de ajuste (Cea, 2002). Además, se ha señalado que diferentes factores afectan los requerimientos del tamaño de la muestra ( $N$ ), por ejemplo: a) la complejidad del modelo, b) la ausencia de normalidad, c) la cantidad de valores perdidos, d) el tipo de estimación (Hair, Black, Babin y Anderson, 2014). Pese a esto, determinar el tamaño de la  $N$  en SEM sigue siendo un aspecto sin una respuesta adecuada en la literatura (Teo, Tsai y Yang, 2013; Westland, 2015).

Si bien es importante reconocer que los criterios que se citan en este trabajo para determinar el tamaño de la muestra en SEM pueden encontrarse dispersos en diversos textos y revistas académicas, una de las contribuciones de este consiste en desarrollar un procedimiento práctico y fundamentado que reúna dichos criterios en un solo procedimiento. Asimismo, el artículo constituye un esfuerzo por divulgar de forma didáctica, sin tecnicismos de más, el procedimiento desarrollado para ofrecer una herramienta a los y las investigadores en educación que les permita abordar uno de los aspectos más importantes y ampliamente discutidos de SEM.

De acuerdo con todo lo anterior, el objetivo principal de este artículo es ofrecer a los investigadores en el campo educativo una guía práctica para utilizar la diversidad de criterios y justificar el cálculo apropiado del tamaño de muestra en SEM, mediante la planificación a priori de modelos de medida y modelos estructurales. Con ello, se espera contribuir a su

actualización en el campo educativo con el fin de impulsar el uso de SEM, una técnica que les permite contrastar empíricamente múltiples hipótesis y diseñar los instrumentos necesarios para implementar los modelos desarrollados. Es pues un esfuerzo por hacer un aporte a la investigación educativa y a la formación de quienes tienen en sus manos esta desafiante tarea. En este sentido, los autores esperan realizar un modesto aporte al enorme y fundamental desafío de encontrar formas de mejorar la educación de las presentes y futuras generaciones de nuestro país al incorporar en la discusión de la investigación educativa reflexiones recientes en torno al tamaño de la muestra en SEM.

## **2. Revisión de la literatura**

La propuesta presentada en este artículo de estimación del tamaño de muestra en SEM requiere planificar, a priori, posibles modelos de análisis que el investigador utilizará en su estudio; estos le permitirán precisar el conjunto de parámetros necesarios para aplicar los diferentes criterios que demanda el método sugerido. Es por ello que este apartado está dedicado a conceptualizar brevemente SEM, los tipos de modelos que el investigador puede plantear en este proceso previo de planificación. Asimismo, se presentan algunas variaciones que pueden ser de utilidad; tales como, modelos anidados y parcelas. De esta forma, para concluir, se describen diferentes enfoques propuestos en la literatura, relativos a la determinación de la muestra en SEM.

### **2.1 Cuestiones conceptuales referentes a SEM**

Disponer de una adecuada fundamentación teórica o conceptual, que permita sustentar e interpretar los resultados empíricos, es el punto de partida para especificar un modelo SEM (Catena, Ramos y Trujillo, 2003; Hair et al., 2014; Lévy, Martín y Román, 2006; Westland, 2015); esto por cuanto SEM es un método estadístico multivariado que permite modelar relaciones de correlación y causalidad entre constructos latentes<sup>3</sup>. Es por ello que, para trabajar con este tipo de constructos (factores latentes), es necesario conocer o al menos vislumbrar de antemano las relaciones entre estos y entre los constructos y sus indicadores.

---

<sup>3</sup> Constructos latentes también llamados reflectivos; en tanto, el factor explica la varianza común de los indicadores. En este sentido, el factor es latente y no se puede homologar al total de indicadores. Por su parte, en los formativos el factor está formado por todos los indicadores, por lo que el análisis se realiza sobre la varianza total (Kline, 2011). En este trabajo se aborda el tamaño de la muestra en el primer tipo, por su amplio uso en Ciencias Sociales.

Ahora bien, por modelo se entenderá la representación de la estructura hipotetizada que vincula los indicadores con los factores latentes y estos últimos entre sí (Byrne, 2006). Asimismo, hace referencia a una red de factores no observables relacionados de forma lineal, en la cual las relaciones son correlaciones canónicas entre dichos factores, entendiéndose por canónica una relación entre factores que no son observados y medidos directamente (Westland, 2015).

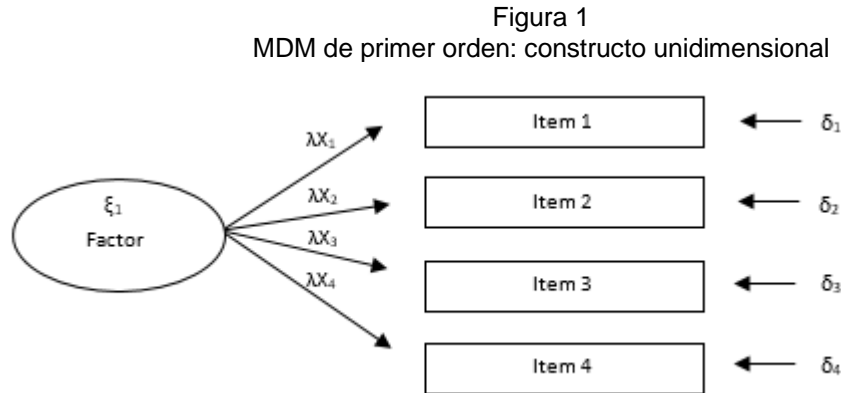
En concordancia con todo lo anterior, es pertinente señalar que SEM es una técnica principalmente confirmatoria (Ullman, 2013) e idónea para el análisis de una teoría estructural relativa a un fenómeno (Byrne, 2006). No obstante, cuando los resultados requieren la modificación de un modelo SEM, previamente establecido, se considera que el análisis toma un carácter exploratorio (Hair et al., 2014). Igualmente, en este caso, también se requiere un andamiaje empírico o conceptual que permita variar el modelo (Schumacker y Lomax, 2010). Este último uso de SEM es llamado, por Kline (2011), generación de modelos, en contraste con la especificación inicial, la cual puede considerarse como la representación de las hipótesis de un estudio mediante SEM (Varela, Abalo, Rial y Braña, 2006).

## **2.2 Tipología de modelos en SEM: modelos de medida y modelos estructurales**

Es esencial comprender la relación entre indicadores y factores latentes, dado que, en ciencias sociales y educación, las percepciones, los juicios y las impresiones han llevado al planteamiento de factores inherentemente no observables que necesitan, para su estudio, ser modelados como variables subyacentes (Westland, 2015, p. 83). De lo anterior se desprenden dos grandes componentes de modelos SEM que puede abordar un investigador. El primero constituido por los modelos de medida (MDM) que evalúan la relación entre los indicadores y sus correspondientes factores latentes. Un MDM operacionaliza un constructo en tanto describe cómo este explica la varianza de los indicadores (Catena et al., 2003). La técnica utilizada para los MDM es el Análisis Factorial Confirmatorio (AFC) (Brown y Moore, 2012). Hay que mencionar, además, que los MDM pueden ser de primer nivel (primer orden) o segundo nivel (segundo orden o jerárquico).

Por una parte, entre los primeros se encuentran aquellos que pretenden confirmar la relación de los indicadores con el factor latente ( $\xi$ ), tal como se aprecia en la Figura 1. Por otra parte, están aquellos que permiten someter a prueba la multidimensionalidad de un constructo, concretamente, los subfactores latentes que lo conforman. En ambos casos, se

someten a prueba las hipótesis de la existencia de saturaciones o cargas factoriales ( $\lambda$ ), definidas a priori, entre un constructo y sus indicadores; asimismo, la existencia de correlaciones o no ( $\Phi$ ) entre los factores. Para efecto de los MDM, es importante explicar el término de modelo congénico. Este se caracteriza por no tener cargas cruzadas entre los factores ( $\xi$ ) ni errores ( $\delta$ ) correlacionados entre los ítems (ver Figura 2) (Hair et al., 2014). Por esta razón, permiten mantener "íntacta" la dimensionalidad de los constructos, por lo que son muy útiles en la evaluación de escalas de medición.

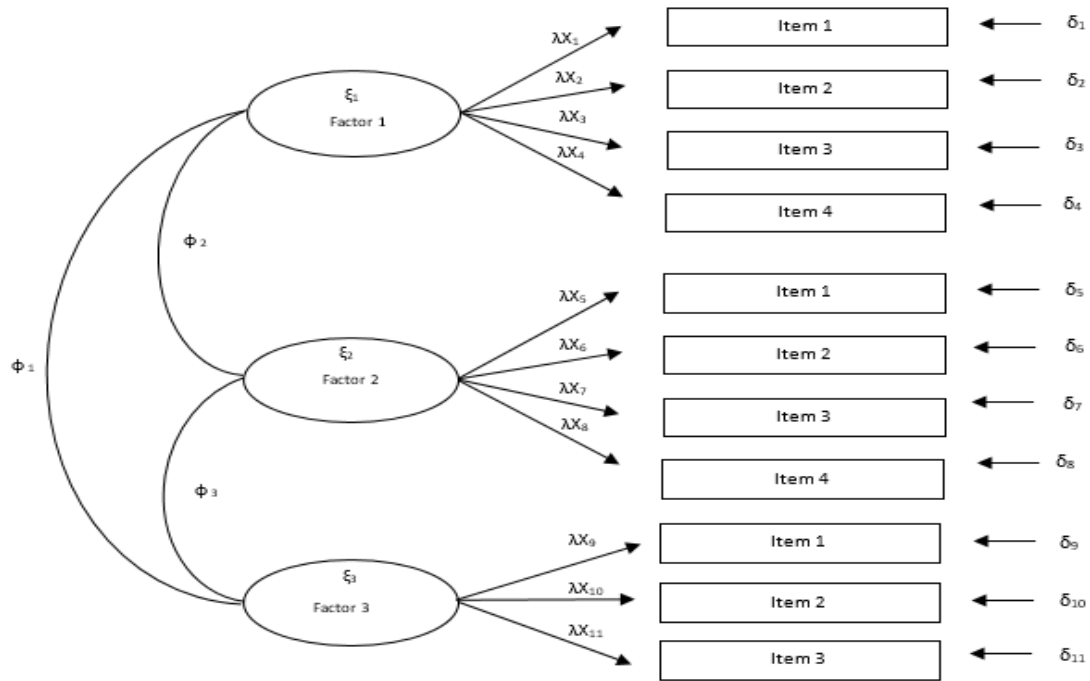


**Fuente:** Elaboración propia con base en Catena et al. (2004); Lévy et al. (2006).

En cuanto a los MDM de segundo nivel o jerárquicos están aquellos diseñados para corroborar la existencia de un factor latente de segundo orden (no conformado por indicadores,  $\xi$ ) que explica la varianza de los constructos de primer orden (o subfactores,  $\xi$ ). Por lo tanto, estos últimos estarán sujetos a un error de predicción denominado término de perturbación ( $\zeta$ ). Evidentemente, los factores de primer orden estarían constituidos por sus respectivos indicadores (ver Figura 3). Los MDM descritos son de utilidad para establecer la calidad de la medición de cada constructo latente (MacKinnon, 2008; Williams, Vandenberg y Edwards, 2009), en especial cuando se desarrollan mediante conceptos derivados de la teoría clásica de los test (Martínez, Hernández y Hernández, 2010).

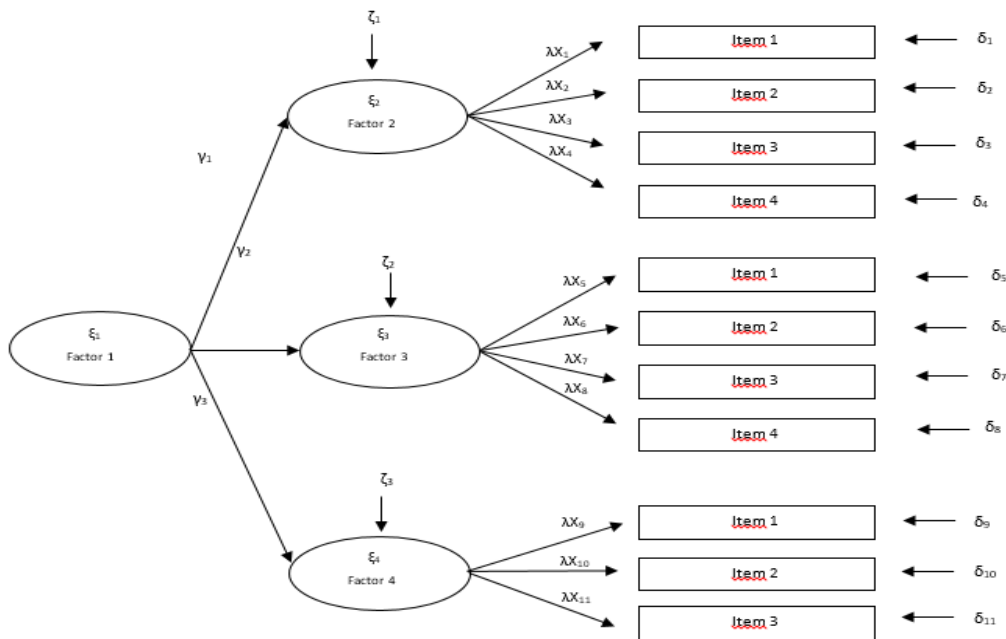
Un segundo componente de los modelos SEM se denomina modelo estructural (ME), el cual estima, de forma simultánea, diferentes efectos causales (Dattalo, 2008). Los ME están constituidos por relaciones causales entre factores (exógenos –  $\xi$  – y endógenos –  $\eta$  –). Estas relaciones pueden implicar covarianzas, así como efectos directos e indirectos entre los constructos (Lévy y González, 2006, p. 172). Estos últimos también se denominan de mediación (Williams et al., 2009).

Figura 2  
MDM de primer orden: multidimensionalidad de un constructo



Fuente: Elaboración propia con base en Catena et al. (2004); Lévy et al. (2006).

Figura 3  
MDM de segundo orden: grado de jerarquía de un constructo de mayor nivel



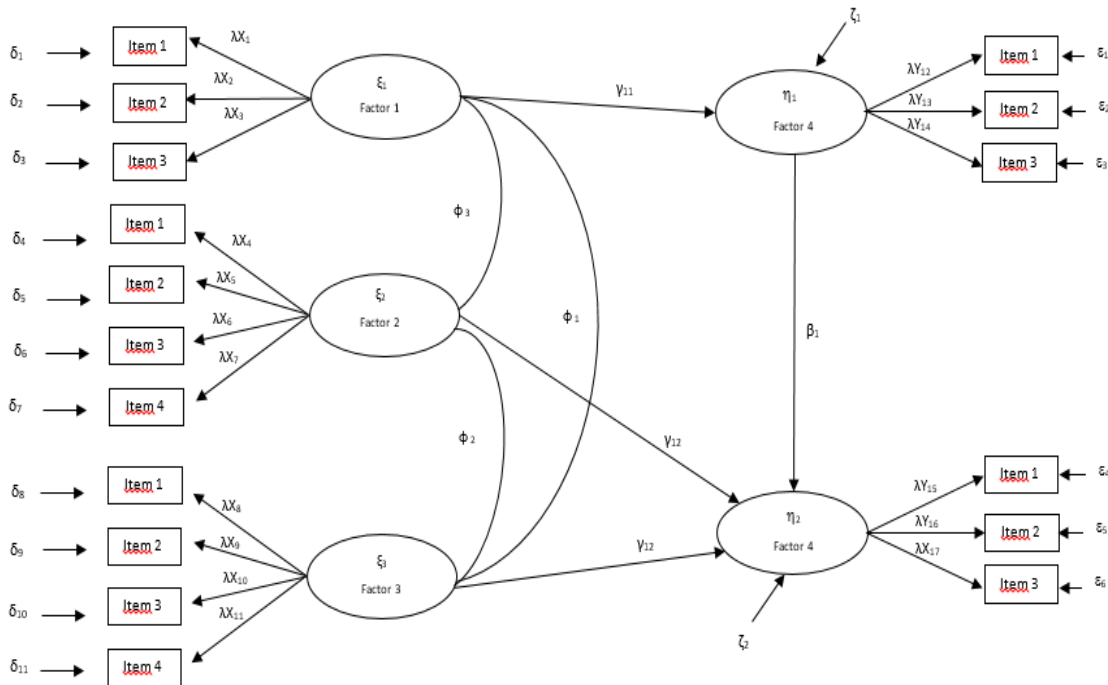
Fuente: Elaboración propia con base en Catena et al. (2004); Lévy et al. (2006).

Se debe agregar que los ME pueden ser recursivos (ver Figura 4), es decir, aquellos en los que las relaciones de causalidad son establecidas como unidireccionales (Lévy et al., 2006). Además, en los no recursivos existen factores latentes con causalidad recíproca (ver Figura 5).

### 2.3 Variaciones en el diseño de modelos SEM: modelos anidados y parcelas

Este tipo de modelos se generan cuando, a partir de un modelo base, se cambia en orden, uno a uno, sus parámetros (Hoyle, 2012). Esto puede ser muy útil cuando se requiere obtener evidencias de la validez convergente y discriminante de un instrumento o, en el caso de los modelos estructurales, cuando se requiere evaluar el efecto en el modelo de un nuevo parámetro. La variación entre el modelo base y el anidado se evalúa mediante la diferencia en  $\chi^2$ . Esto es similar a la regresión jerárquica en la que se procede por pasos con el fin de valorar el cambio en F y  $R^2$ . Valga hacer notar que la metodología SEM, en especial el AFC, se ha constituido en una técnica cada vez más utilizada para obtener este tipo de evidencias de validez. (Eid, Lischetzke y Nussbeck, 2006)

Figura 4  
ME recursivo



Fuente: Elaboración propia con base en Lévy et al. (2006).



Finalmente, hay que mencionar la utilización de parcelas. Al respecto, se han señalado dos enfoques. En el primero, denominado de desagregación parcial, el investigador combina indicadores de los factores en parcelas con el fin de simplificar la estructura del modelo y disminuir el número de variables observadas. El segundo enfoque se ha llamado de desagregación total, en el que cada indicador se mantiene sin combinar, lo cual brinda información acerca del funcionamiento psicométrico de cada ítem (Williams et al., 2009, p. 548). Sin embargo, aumenta de forma importante la cantidad de indicadores del modelo, lo que repercute, como se verá en el tercer apartado, en un aumento considerable en los requerimientos del tamaño de la  $N$ .

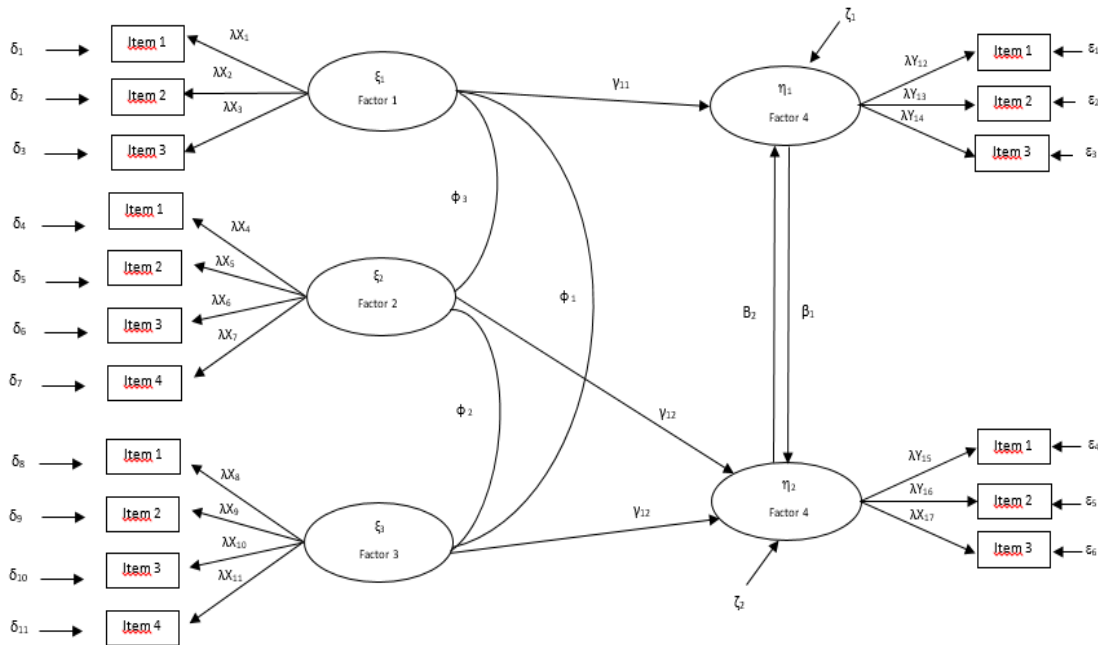
Little, Cunningham, Shahar y Widaman (2002) señalan que el uso de parcelas posee las siguientes ventajas: a) se obtiene mayor confiabilidad, b) comunalidades más altas, c) mejor relación número de casos por parámetro, d) menores posibilidades de violar los supuestos de la técnica, e) mejor comprensión de la relación entre constructos, f) modelos más parsimoniosos, g) reducción del riesgo de obtener residuos correlacionados y cargas factoriales cruzadas, es decir, mayores posibilidades de lograr un modelo congenérico, h) disminución de las fuentes de error de muestreo, i) mayor proximidad a una medición por intervalo, j) reducción de la probabilidad de correlaciones espurias, k) soluciones más estables y con mayores posibilidades de generalización.

Pese a lo anterior, Worthington y Whittaker (2006) se manifiestan en contra del uso de parcelas para el desarrollo de escalas. Igualmente, Little et al. (2002) advierten que este enfoque no debe utilizarse con factores multidimensionales, ante la dificultad de interpretar el constructo subyacente a cada parcela. Sin embargo, este inconveniente se atenúa cuando existen altas correlaciones entre los subfactores a los que corresponden las parcelas, elemento que debe ser tomado en cuenta por el investigador al utilizarlo en SEM.

#### **2.4 Enfoques referentes a la determinación de tamaño de $N$ en SEM**

Tal como se ha mencionado, el tamaño de la muestra es un aspecto esencial en SEM, sin embargo, la literatura no ha ofrecido una respuesta concluyente para determinar la cantidad de casos requeridos para un adecuado análisis SEM (Kline, 2011). Por el contrario, se ha brindado una gran diversidad de criterios que constituyen una masa desarticulada de literatura que dificulta el trabajo del investigador. Con el fin de darle cierto orden, para efectos de este trabajo, se han agrupado en 4 categorías: a) cantidad absoluta de casos, b) casos por parámetro, c) casos por variable observada, d) potencia estadística (ver Tabla 1).

Figura 5  
ME no recursivo



Fuente: Elaboración propia con base en Lévy et al. (2006).

Tabla 1  
Criterios para el cálculo de  $N_{enSEM}$

Criterio	Autores
<b>1. Cantidad absoluta de casos</b>	Catena et al. (2003); Hair et al. (2014); Stevens (2009).
200	
<b>2. Casos por parámetro</b>	
20	Kline (2011).
10	Kline (2011).
5	Worthington y Whittaker (2006).
<b>3. Casos variable observada</b>	
8 (incluyendo latentes)	Catena et al. (2003).
15	Catena et al. (2003).
<b>4. Potencia</b>	
$N$ que permita $1-\beta \geq .80$ (Ellis, 2010; Dattalo (2008) <sup>a</sup> Cumming, 2012).	

Fuente: Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup>Se puede calcular con el método para un parámetro de Saris y Satorra (1993), con base en el ajuste global del modelo ( $RMSEA$ ) mediante el procedimiento de MacCallum, Browne y Sugawara (1996) y con base en la estimación del límite inferior de casos necesarios para la detección del menor tamaño del efecto implicado en los coeficientes del modelo estructural acorde con Westland (2010).

## 2.5 Cantidad absoluta de casos

En la primera categoría se sugiere un valor mínimo de 200 casos (Catena et al., 2003; Hair et al., 2014; Stevens, 2009). Sin embargo, autores como Kline (2011) advierten que incluso una  $N$  de este tamaño podría ser pequeña cuando el modelo es complejo, no existe normalidad multivariada o se utilizan ciertos tipos de estimación. Hair et al. (2014) agregan que muestras muy grandes, muy superiores a 200, tienen el inconveniente de generar modelos muy sensibles; por ejemplo, que parámetros para una magnitud muy pequeña resulten estadísticamente significativos.

## 2.6 Casos por parámetro

Por una parte, la segunda categoría considera el tamaño de la muestra en función de aspectos como, el número de constructos, ítems y comunalidades. De manera concreta, se recomienda un mínimo de 100 casos para modelos con 5 o menos constructos, cada uno de los factores con más de 3 indicadores y con comunalidades por ítems superiores a .60 (Hair et al., 2014). Por otro parte, también se ha recomendado utilizar la cantidad de casos por parámetro; por ejemplo, Kline (2011) sugiere: a) muestra "ideal": 20 casos por parámetro, b) muestra "menos ideal": 10 casos por parámetro, c) muestra "inapropiada": menos de 10 casos por parámetro. De acuerdo con dicho autor, la opción c) reduciría la confiabilidad de los resultados. Pese a esto, Worthington y Whittaker (2006) señalan que en un AFC, 5 casos por parámetro es una muestra adecuada.

## 2.7 Casos por variable observada

Autores como Catena et al. (2003) han propuesto sumar el número de variables observadas o indicadores y el número de constructos latentes. A partir de esto, proponen que una muestra adecuada sería de 8 casos por el total de variables observadas y latentes. También, sugieren, al igual que Hair et al. (2014), que 15 casos por indicador o variable observada da como resultado una muestra apropiada que incluso permite minimizar problemas de normalidad multivariable.

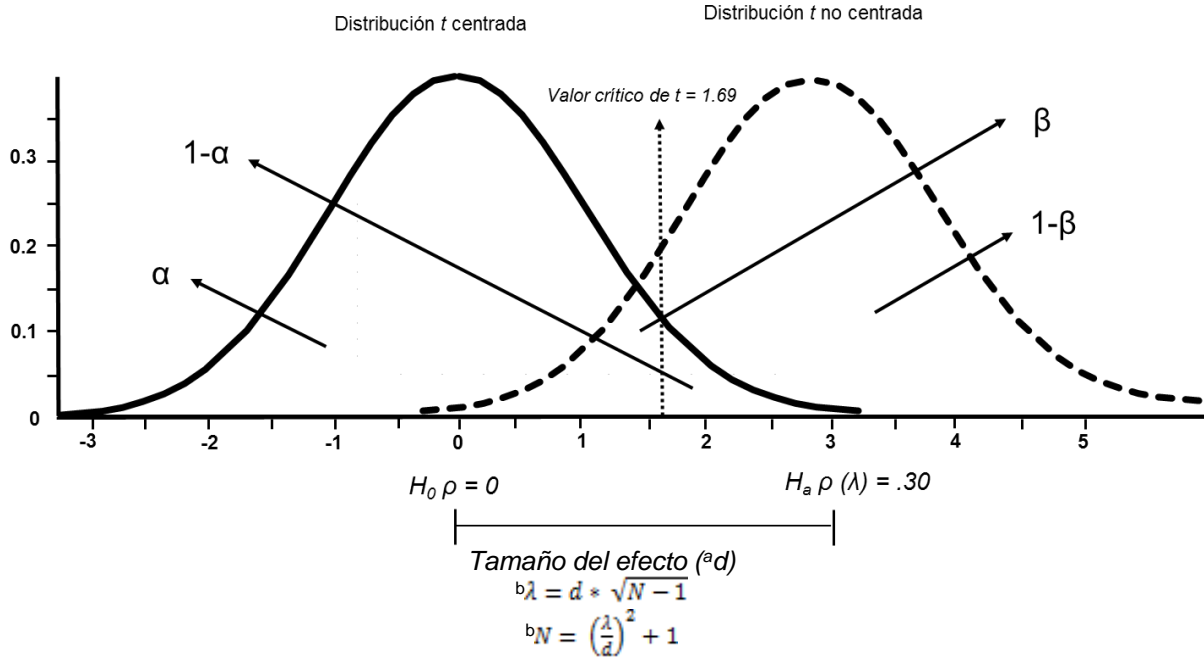
## 2.8 Potencia estadística

Cuando se utiliza una prueba de hipótesis, existe la posibilidad de cometer dos tipos de errores. El primero, denominado tipo I, consiste en rechazar la hipótesis nula ( $H_0$ ) cuando es cierta. El segundo, conocido como tipo II, ocurre al aceptar la  $H_0$  cuando es falsa (Cumming,

2012, Ellis, 2010). Se puede cometer uno de los dos errores, pero no ambos simultáneamente. Como es bien conocido, las pruebas de hipótesis trabajan con distribuciones muestrales cuya variabilidad, denominado error estándar, disminuye al aumentar el tamaño de la  $N$ . El error tipo II ( $\beta$ ) se puede manipular cambiando la tasa de error tipo I ( $\alpha$ ), el tamaño del efecto ( $d$ ) o variando el tamaño de la  $N$  (o lo que es lo mismo, modificando el error estándar de la distribución muestral correspondiente). Lo anterior, dependiendo de la prueba correspondiente, se expresa mediante una fórmula matemática que nos permite despejar el tamaño de la  $N$  para un determinado  $\alpha$  y  $d$ .

Con el fin de aclarar el concepto de  $1-\beta$  se abordará su cálculo para el caso de un coeficiente de correlación de Pearson ( $\rho$ ). Como se observa en la Figura 6, alrededor de  $H_0$  se construye una distribución muestral que se denomina centralizada, ya que  $H_0 \rho=0$ . La ventaja de las distribuciones centralizadas es que su valor esperado es fijo, en este caso, indica que la correlación es igual a cero. Por el contrario, la distribución no centralizada se establece alrededor de la hipótesis alternativa ( $H_a$ ), cuyo valor varía de acuerdo con la magnitud de la correlación que se considera importante detectar. Teóricamente, puede asumir una cantidad infinita de valores diferentes de cero. Este parámetro se denomina  $\lambda$ , siendo la diferencia entre 0 y  $\lambda$  lo que se denomina tamaño del efecto.

Figura 6  
Distribución centralizada y no centralizada para el cálculo de potencia para el coeficiente de correlación de Pearson



**Fuente:** Elaboración propia a partir de 121 casos, las probabilidades para t y z coinciden (Ritchey, 2008).

<sup>a</sup> En este caso, *d* coincide con la correlación de Pearson esperada ( $\rho_A$ ).

<sup>b</sup> Con base en Cohen (2008) y Howell (2010).

En el ejemplo propuesto,  $\lambda$  se definió como .30. Como es evidente a partir de la Figura 6, valores pequeños de  $\lambda$  tienden a superponer en gran medida las distribuciones, lo cual incrementa tanto  $\alpha$  como  $\beta$ . Por el contrario, correlaciones altas tienden a alejar las distribuciones aumentando el nivel de confianza ( $1-\alpha$ ) y la potencia ( $1-\beta$ ). Lo que es importante hacer notar es que el investigador puede definir  $\alpha$  (usualmente, un 5%) y si puede definir el tamaño del efecto que es relevante para la investigación, cuenta con una herramienta para determinar el tamaño de la *N*. Como se verá más adelante, el cálculo de  $1-\beta$  en SEM es mucho más complejo, sin embargo, se han brindado diferentes criterios que facilitan determinar *d*.

Para calcular  $1-\beta$  en SEM, se han propuesto 3 enfoques. El primero fue desarrollado por Saris y Satorra (1993). El paso inicial consiste en definir el modelo que constituye  $H_0$ . En el segundo, se le añaden al modelo anterior nuevos parámetros con el fin de plantear  $H_a$ . Dicho modelo, con los parámetros adicionales, es el que se considera verdadero en la población. El

tercer paso requiere especificar los valores de los parámetros. Por ejemplo, suponga que el modelo para  $H_0$  no incluye una correlación entre 2 variables latentes, mientras que el modelo alternativo ( $H_a$ ) la incluye con un valor de .40. De ahí que  $1-\beta$  consiste en la probabilidad de detectar la falsedad del modelo propuesto cuando el alternativo es cierto. Es decir, incluye una correlación de .40. En este caso, el tamaño del efecto es la diferencia en la estimación de los parámetros bajo la  $H_0$  y la  $H_a$  (Lee, Cai y MacCallum, 2012), por lo que el investigador requiere definir todos los valores de los parámetros del modelo, lo cual constituye una considerable cantidad de información que dificulta el uso de este método para el cálculo de  $1-\beta$  a priori. Sin embargo, es un enfoque robusto y útil cuando el investigador necesita enfocarse en un solo parámetro.

El segundo enfoque, propuesto por Westland (2010), también se basa en los parámetros del modelo. Sin embargo, no es necesario definirlos todos *a priori* debido a que lo que se establece es el valor más pequeño que se desea detectar de todo el conjunto de parámetros que puede tener un modelo. De esta forma, se obtiene un tamaño de la  $N$  para detectar como estadísticamente significativo el efecto mínimo o, en otras palabras, el parámetro más pequeño que teóricamente es relevante identificar. Por ello, se señala que este método está dirigido a establecer el límite inferior de casos para detectar el efecto más pequeño de un modelo SEM (Ryan, 2013). En términos matemáticos, para establecer el límite inferior de la muestra es necesario determinar todas las posibles combinaciones de factores latentes ( $i$ ) y definir el  $d$  mínimo para todas las  $i$ . Para esto, se selecciona la correlación más pequeña que se desea detectar entre las variables latentes del modelo. Para efectos prácticos, Westland (2015) sugiere, con base en Cohen (1988) para Ciencias Sociales, los siguientes criterios: a) .10 a .23 efecto pequeño, b) .24 a .36 efecto mediano, c) .37 o más, efecto grande.

Para establecer las hipótesis, se plantea el efecto de forma dicotómica: a)  $H_0: \rho < \lambda = 0$ , b.  $H_a: \rho \geq \lambda = 1$ . De lo anterior, se deriva que cada  $i$  con efecto 0 implica un nuevo modelo, por lo que la complejidad del cálculo obedece a la necesidad de determinar todas las " $i$ " para todos los posibles modelos  $\{0,1\}$ . El cálculo requiere definir, además, el número de variables latentes del modelo ( $k$ ), el número de indicadores ( $p$ ),  $\alpha$  y  $\beta$ . Con esto,  $n = f [k, p, \delta | \alpha, \beta]$ .

El tercer enfoque fue propuesto por MacCallum et al. (1996), con base en el ajuste global del modelo medido por el índice RMSEA. El límite inferior de este es 0, lo cual constituye un ajuste perfecto, mientras que niveles mayores a .08 evidencian una discrepancia inaceptable entre la matriz de varianzas y covarianzas estimada ( $\Sigma$ ) y la

observada ( $S$ ) (Bentler, 2006; Schäffer, 2007; Schumacker y Lomax, 2010). Es decir, incrementos en el índice se corresponden con aumentos en la discrepancia entre  $\Sigma$  y  $S$  (Byrne, 2010). Para una determinada  $N$  y un  $\alpha$  específico, se puede identificar el valor de  $\chi^2$  a partir de lo cual se ubica  $1 - \beta$ . En este enfoque,  $H_0$  se establece en función del ajuste óptimo deseado y  $H_a$  de acuerdo con el nivel desajuste máximo tolerado. Para esto, se ha definido la siguiente guía: a)  $RMSEA \leq .05$ , ajuste cercano, b)  $.05 \leq RMSEA \leq .08$ , ajuste aceptable, c)  $.08 < RMSEA \leq .10$ , ajuste mediocre, d)  $RMSEA > .10$ , ajuste pobre (Lee et al., 2012).

Como se indicó en este caso,  $1 - \beta$  indica la probabilidad de rechazar un ajuste deseado ( $H_0$ ) dada la existencia de cierto nivel de desajuste, por lo que se puede considerar un enfoque congruente con el principio de falsación. Este enfoque requiere mucho menos información que el de Saris y Satorra (1993), ya que no es necesario especificar los valores de todos los parámetros. Esto constituye una clara ventaja cuando el investigador desea un enfoque *a priori* de  $1 - \beta$ , tal y como lo recomienda Cumming (2012). Como se detallará más adelante, además de la definición de los parámetros que requiere el método de Westland (2010), es necesario definir los grados de libertad (gl) de cada modelo; con esto,  $N = f [k, p, gl, \delta | \alpha, \beta]$ . Ahora bien, cuando los gl son pocos, la  $N$  requerida mínima es grande, con el fin de poder obtener un determinado nivel de potencia (por ejemplo: .80). Esto ocurre cuando el número de variables observadas es pequeño o cuando el modelo especificado incluye una cantidad grande de parámetros por estimar libremente o ambas condiciones. Valga aclarar que el cálculo de la  $N$ , por este método, no asegura la estimación adecuada de los parámetros individuales, ya que se enfoca en el ajuste global del modelo.

De acuerdo con lo descrito previamente, los métodos de Westland (2015) y de MacCallum et al. (1996) resultan ideales para el cálculo *a priori* del tamaño de la  $N$  en SEM debido a que, entre otras cosas, requieren información que el investigador puede definir con relativa facilidad antes de la ejecución de cualquier trabajo de campo. Además, son complementarios ya que primero se puede definir la potencia relacionada con el ajuste global del modelo para, seguidamente, determinar los requerimientos para detectar los efectos considerados relevantes. En el siguiente apartado, se presentará la metodología diseñada para aplicar los criterios descritos para el cálculo del tamaño de la  $N$  en SEM.

### 3. Método para aplicar criterios de cálculo de tamaños de $N$ en SEM

Dattalo (2008) recomienda no basar en "reglas de oro" la determinación del tamaño de  $N$  en SEM. Para ello, a continuación, se le ofrece al lector una metodología práctica y sencilla

que reúne los criterios previamente presentados para estimar *a priori* el número de casos necesarios para usar apropiadamente SEM (ver Tabla 1).

Una vez desarrollado el andamiaje teórico previo y definidas las hipótesis de investigación, se deben plantear los posibles modelos de análisis (MDM y ME) que podrían generarse durante la investigación. Para cada uno de ellos es necesario especificar los siguientes aspectos generales: a) representación gráfica, b) número de factores y sus respectivos indicadores, c) elementos de información en  $S$ , c) número de parámetros por estimar, d) grados de libertad (gl), e) el ajuste global del modelo con base en el RMSEA, f) definir la correlación mínima que se desea detectar en cada modelo, g) definir  $\alpha$  y  $\beta$ . Es usual que se definan valores de  $\alpha=.05$  y  $\beta=.20$  (Cumming, 2012; Ellis, 2010; Howell, 2010; Murphy, Myers y Wolach, 2014) lo cual, aunque arbitrario, es ampliamente aceptado y puede ser racionalmente defendido. (Murphy et al., p. 22)

Cuando el investigador no tiene disponible un instrumento de medición previamente diseñado y validado, requiere definir los indicadores para evaluar los factores latentes de cada modelo de interés. Valga señalar que SEM requiere que  $S$  contenga más puntos de información que  $\Sigma$ , ya que carece de interés científico un modelo que plantee todas las relaciones posibles para estimar los datos observados. Por tal razón, se debe apuntar a las relaciones sustantivas, siendo la diferencia entre los puntos de información de  $S$  y  $\Sigma$  un indicador de ello, denominado gl o, de manera general, condición de orden. Para no incumplir lo anterior, es necesario definir, al menos, 3 o 4 indicadores para cada constructo. (Hair et al., 2014; Schumacker y Lomax, 2010)

Así, por ejemplo, un factor latente evaluado por 3 indicadores requiere el cálculo de 6 parámetros (3 errores y 3 cargas factoriales). La cantidad de puntos de información se calcula mediante la fórmula  $\frac{p*(p+1)}{2}$ , donde  $p$  representa el número de elementos de información (Cea, 2002; Ullman, 2013). En el caso del ejemplo planteado, el valor resultante es  $6 \left[ \frac{3*(3+1)}{2} \right]$ , con lo que el factor latente estaría exactamente identificado. Con 4 indicadores se obtendrían 10 puntos de información que, al restarle 8 parámetros, da como resultado 2 gl. En este caso, se considera que el factor latente está sobre-identificado. Autores como Hair et al. (2014) y Schumacker y Lomax (2010) señalan que el número óptimo de indicadores por constructo es 4, aunque con 3 se cumple con la condición de orden. Si bien, un número mayor de indicadores permite más gl, puede llevar a problemas con la unidimensionalidad de los constructos (Hair et al. 2014). Igualmente, se considera importante que, en la medida de lo posible, se plantee un



modelo congenérico, ya que permite identificar, sin ambigüedad, las fuentes de la varianza, lo cual es fundamental en el desarrollo de escalas de medición (DeVellis, 2012).

El siguiente paso es utilizar estos datos de acuerdo con cada categoría de criterios. Recuérdese que la primera establece un número fijo de 200 casos. El segundo conjunto de criterios establece como pauta 5, 10 o 20 casos por parámetro, por lo que, para obtener el tamaño de  $N$ , simplemente se multiplica la cantidad de parámetros de cada modelo por el criterio respectivo. La tercera categoría sugiere 8 casos por variables tanto observables como latentes o, alternativamente, 15 casos por variable observable (Catena et al., 2003). De igual modo que en el segundo enfoque, el tamaño de la  $N$  es el resultado del producto del número de variables por los criterios planteados.

Para la cuarta categoría, se sugiere la utilización de los procedimientos de MacCallum et al. (1996) y Westland (2010) para calcular la potencia en SEM. Es importante señalar que, en ambos casos, la información requerida puede ser razonablemente especificada *a priori* por el investigador. Además, para el tamaño del efecto se han preestablecido criterios muy sencillos de utilizar. Más aún, existen herramientas en línea que permiten realizar fácilmente los cálculos complejos implicados en ambos procedimientos. Concretamente, la aplicación desarrollada por Preacher y Coffman (2006) permite el cálculo de la  $N$  de acuerdo con el procedimiento sugerido por MacCallum et al. (1996), en tanto que la desarrollada por Soper (2015) lo hace de acuerdo con el algoritmo desarrollado por Westland (2010).

El programa diseñado por Preacher y Coffman (2006) requiere los  $gl$  de cada modelo, los valores de  $\alpha$  y  $1-\beta$ , así como el valor del RMSEA para  $H_0$  y  $H_a$ . Posteriormente, se genera el código en R (haciendo clic en la etiqueta "Generate RCode") y se ejecuta (haciendo clic en la etiqueta "Submit above to RWeb") para obtener el tamaño de la  $N$  requerido para cada modelo (ver interfaz en la Figura 7).

Figura 7  
Interfaz de la herramienta de Preacher y Coffman

**Compute Sample Size for RMSEA**

Alpha	<input type="text"/>
Degrees of Freedom	<input type="text"/>
Desired Power	<input type="text"/>
Null RMSEA	<input type="text"/>
Alt. RMSEA	<input type="text"/>

Generate R Code

R code will appear here

Submit above to Rweb      Erase R code

**Fuente:** Tomado de Preacher y Coffman (2006)

La herramienta de Soper (2015) requiere que se defina la correlación mínima que se desea detectar de acuerdo con las características del modelo. Aunque .37 se considera un efecto grande, es un nivel razonable para muchos fenómenos en dirección de empresas, ya que implica, apenas, un 13,69 % de la varianza. También es necesario definir  $\alpha$  y  $1-\beta$ , así como el número de variables latentes y observadas. Para obtener el tamaño de  $N$  para el efecto mínimo que se desea detectar, basta con dar clic en el botón "Calculate" (ver interfaz en la Figura 8). Valga señalar que efectos menores requieren muestras mayores debido al error estándar de la respectiva distribución muestral. De esta forma, cualquier efecto mayor, en igualdad de condiciones, arrojará un tamaño de muestra menor. En el siguiente apartado, se terminará de clarificar el procedimiento mediante la utilización de varios ejemplos entre los que se incluyen varios MDM y ME.

#### 4. Uso práctico del método sugerido

Con el fin de describir con detalle cada uno de los pasos para determinar el tamaño de una muestra en SEM con la metodología propuesta, en este apartado se desarrollará un ejemplo que toma como base la estructura de 7 modelos planteados por Vargas-Halabí, (2016) en una de sus investigaciones. El estudio tuvo como objetivo estudiar los efectos de la cultura organizativa (CO) en el desempeño innovador (DI) en las empresas. Una vez que se desarrolló un andamiaje conceptual adecuado y sus hipótesis de investigación, Vargas-Halabí (2016) identificó 5 MDM y 2 ME necesarios para validar los instrumentos que utilizó y someter a contrastación empírica sus hipótesis. Seguidamente, se describe la estructura de cada uno de los modelos.

Figura 8  
Interfaz del programa de Soper

This calculator will compute the sample size required for a study that uses a structural equation model (SEM), given the number of observed and latent variables in the model, the anticipated effect size, and the desired probability and statistical power levels. The calculator will return both the minimum sample size required to detect the specified effect, and the minimum sample size required given the structural complexity of the model.

Please supply the necessary parameter values, and then click 'Calculate'.

Anticipated effect size:  ?

Desired statistical power level:  ?

Number of latent variables:  ?

Number of observed variables:  ?

Probability level:  ?

Fuente: Tomado de Soper (2015)

El primer MDM corresponde a un AFC de la Denison Organizational Cultural Survey (DOCS) desarrollada por Denison y Mishra (1995) para evaluar la CO (ver Figura A del Apéndice 1). Este se planteó para evaluar la validez relacionada con la estructura interna de la DOCS, pero, principalmente, para determinar el funcionamiento de cada uno de los 60 ítems del instrumento. En este sentido, se utilizó un enfoque de desagregación total. Con una rápida inspección visual al diagrama de sendero (Figura A del Apéndice 1), se determina que el modelo contiene 4 factores de segundo nivel y 12 subfactores con 5 ítems cada uno (16 constructos y 60 ítems). Para determinar el número de puntos de información en **S**, se aplicó la respectiva fórmula  $\left(\frac{60 \cdot (60 + 1)}{2}\right)$  y se obtuvieron 1830 elementos de información. Seguidamente, con base en la Figura A, se determinó que el modelo requiere estimar 138 parámetros distribuidos de la siguiente forma: 60 errores de medida, 60 cargas factoriales, 12 coeficientes de regresión y 6 covarianzas. A continuación, se determinan los gl al restarle, al número de elementos información, la cantidad de parámetros por estimar libremente (1830-138) dando como resultado 1692 (MDM sobre-identificado). Además, cada factor latente de primer orden posee más de tres indicadores, con lo que se cumple con la condición de orden.

Ahora bien, el criterio absoluto establece una *N* de al menos 200 casos, lo cual plantea un punto de partida. Sin embargo, como resultado de la complejidad de este modelo, el enfoque de proporciones por parámetro (5, 10 y 20 sujetos) arroja un número mínimo de 690 y un máximo de 2.760 casos. Por su parte, el enfoque de casos por variables (8 sujetos por variable latente e ítem; 15 sujetos por variable observada) sugiere un mínimo de 608 y un máximo de 900 casos. En síntesis, tal y como se observa en la Tabla 2, los criterios de las

categorías dos (proporciones) y tres (casos por variable) arrojaron muestras muy elevadas. Precisamente, este es uno de los problemas más serios del enfoque de desagregación total, ya que reduce las posibilidades de ajuste del modelo y su estabilidad (Little et al., 2002; Williams et al., 2009). Además, estas  $N$  tan elevadas son consecuencia de la gran cantidad de indicadores, por lo que hay que tomarlas con cautela.

Tabla 2  
Estimación de tamaños de muestra mínimos de acuerdo con los 3 criterios aportados por la literatura

Criterios <sup>a</sup>	Resultado
<b>1. Absolutos</b>	
Mínimo absoluto	200
Número de constructos	16
<b>2. Proporciones</b>	
Número de parámetros del modelo	138
2.a 20 casos por parámetro	2760
2.b 10 casos por parámetro	1380
2.c 5 casos por parámetro	690
<b>3. Casos por variables</b>	
3.a 8 sujetos por ítem y variable latente	608
Número de ítems más variables latentes	76
3.b 15 sujetos por variable ítem	900
Número de ítems	60
<b>4. Potencia</b>	
Método del ajuste global <sup>b</sup>	37
Método del límite inferior <sup>c</sup>	116

**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Acorde con autores Tabla 1 y sección procedimiento. <sup>b</sup> Utilizando herramienta de Preacher y Coffman (2006) <sup>c</sup> Utilizando herramienta de Soper (2015).

La última categoría de criterios corresponde a los métodos de  $1-\beta$  seleccionados. Para ser congruentes con la secuencia de análisis de los modelos SEM, se inició calculado la  $N$  requerida por el procedimiento del ajuste global propuesto MacCallum et al. (1996). En consecuencia, es necesario definir un tamaño del efecto, para lo cual se seleccionó un ajuste aceptable, es decir, un RMSEA de .05 para la  $H_0$  y de .08 para la  $H_a$ . Los gl ya se habían obtenido (1362) y se definió el  $\alpha$  y  $1-\beta$  usuales (.05 y .80 respectivamente). El cálculo, utilizando la herramienta de Preacher y Coffman (2006), arrojó 36,52 casos que se redondea a la unidad siguiente (37) (ver Figura 9). La  $N$  tan baja obedece a la gran cantidad de gl.

A continuación, se procedió a calcular el tamaño de la  $N$  con base en el procedimiento desarrollado por Westland (2015) para detectar el parámetro de la magnitud mínima deseada. Como se puede observar en la Figura 7, ya han sido definidos la mayoría de los valores de las entradas que requiere la interfaz de Soper, con excepción del tamaño del efecto. En este

caso, se consideró que .30 era un nivel adecuado para el estudio, ya que en un instrumento no se recomiendan cargas factoriales menores (Hair et al., 2014). Se obtuvo una  $N$  de 116. Valga señalar que debe tomarse el dato de la etiqueta "*Minimum sample size for model structure*".

Figura 9  
Segmento de la salida del programa de Preacher y Coffman para el MDM1

```
Rweb:> #Computation of minimum sample size for test of fit
Rweb:>
Rweb:> rmsea0 <- 0 #null hypothesized RMSEA
Rweb:> rmseaa <- 0.05 #alternative hypothesized RMSEA
Rweb:> d <- 1692 #degrees of freedom
Rweb:> alpha <- 0.05 #alpha level
Rweb:> desired <- 0.8 #desired power
Rweb:>
Rweb:> #Code below need not be changed by user
Rweb:> #initialize values
Rweb:> pow <- 0.0
Rweb:> n <- 0
Rweb:> #begin loop for finding initial level of n
Rweb:> while (pow
Rweb:> #begin loop for interval halving
Rweb:> foo <- -1
Rweb:> newn <- n
Rweb:> interval <- 200
Rweb:> powdiff <- pow - desired
Rweb:> while (powdiff>.001) {
+   interval <- interval*.5
+   newn <- newn + foo*interval*.5
+   ncp0 <- (newn-1)*d*rmsea0^2
+   ncpa <- (newn-1)*d*rmseaa^2
+   #compute power
+   if(rmseaa>desired) {
+     foo <- -1
+   }
+ }
Rweb:>
Rweb:> minn <- newn
Rweb:> print(minn)
[1] 36.52344 ← / calculada de acuerdo con los parámetros
```

**Fuente:** Tomado de Preacher y Coffman (2006)

Ahora bien, tomando en cuenta el criterio fijo y el del método de potencia de Westland (2015), un tamaño razonable sería 200 casos. Sin embargo, queda muy por debajo de los criterios de proporciones y de casos por variable. Tomando en consideración estos resultados, así como las ventajas del enfoque parcelas, se planteó la utilización de estas para el MDM de la DOCS. Valga señalar que no se recomiendan las parcelas con constructos multidimensionales salvo que las subdimensiones estén altamente correlacionadas, lo cual es el caso de la DOCS (Vargas-Halabí, 2016).

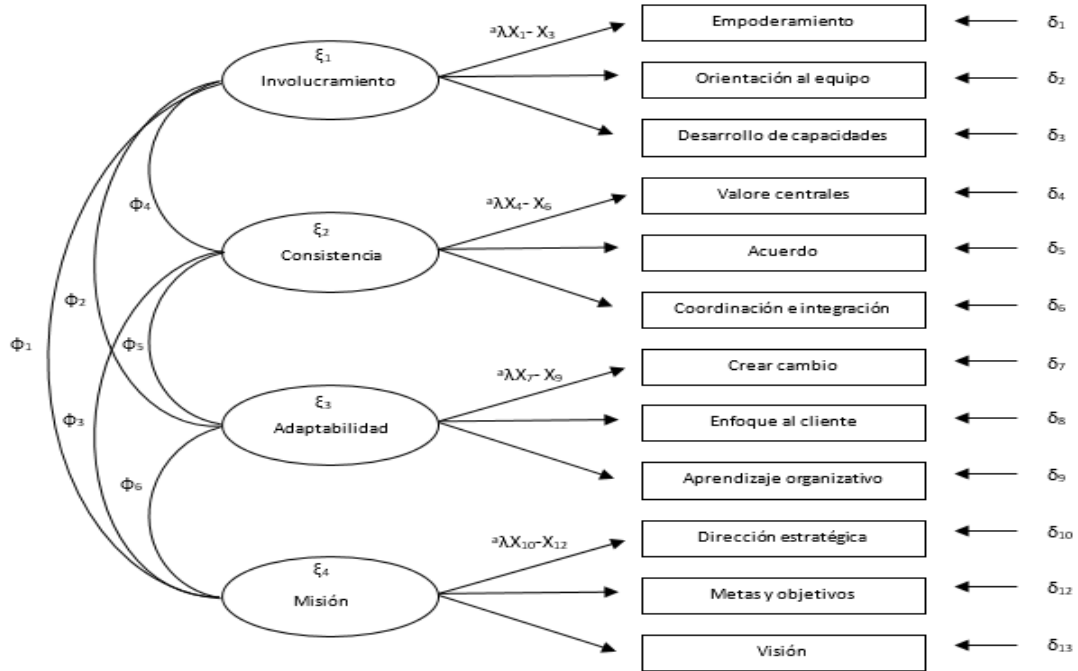
En virtud de los resultados obtenidos y de las limitaciones del enfoque de desagregación total, no se tomó en consideración el primer AFC para efectos de establecimiento del tamaño de la  $N$ . Ahora bien, la propuesta presentada en este artículo se orienta a determinar el tamaño de la  $N$  para una investigación, teniendo en consideración todos los modelos que esta requiere, por lo cual es necesario comparar los resultados obtenidos de los diferentes modelos. Con el MDM 1 se ilustraron los cálculos de estos y tamaños de  $N$  de acuerdo con las 4 categorías. Para los modelos restantes, se presenta la estimación de los insumos con base en los diagramas de sendero para, finalmente, comparar los tamaños de la  $N$  en forma tabulada.

El MDM 2 (Figura 10) consiste en un AFC de primer nivel con 4 variables latentes y 12 indicadores. El propósito es evaluar la validez relacionada con la estructura interna de la DOCS con un enfoque "pragmático liberal" o de desagregación parcial", como también se le denomina al uso de parcelas (Little et al., 2002; Williams et ál., 2009). En este caso,  $S$  posee 78 elementos de información; el modelo requiere estimar libremente 30 parámetros, lo cual nos arroja 48 gl.

El MDM 3 se planteó, con parcelas, para evaluar la validez convergente y discriminante de la DOCS. Para ello, fue necesario agregar 2 factores latentes al MDM 2 (ver Figura 11). El primero, la escala de deseabilidad social (DS), de Marlowe-Crowne adaptada a Costa Rica por Smith-Castro, Molina y Castelain (2014), con sus 11 ítems agrupados aleatoriamente en 3 parcelas (Vargas-Halabí, 2016). El segundo, se incluyó para incorporar el modelo el Desempeño Organizativo (DO) que se evaluaría mediante una escala, aun no diseñada en la etapa de la estimación de  $N$ , conformada por 4 ítems acorde con lo recomendado por Hair et al. (2014) y Schumacker y Lomax (2010). El primer factor permite evaluar la validez discriminante y, el segundo, la convergente, en tanto la DOCS ha mostrado una relación importante con el DO (Sackman, 2011). Con base en estas proyecciones, se esbozó un MDM con 6 factores y un total de 19 indicadores. En consecuencia, la  $S$  contendría 190 elementos, requeriría estimar 53 parámetros y poseería 137 gl.

El MDM 4 fue postulado para evaluar la validez convergente y discriminante del constructo DI, con la escala de DO para el primer tipo de validez y la descrita previamente para DS (ver Figura 12). De la misma forma que DO, el constructo latente DI sería evaluado por una escala por diseñar de 4 ítems. Este modelo estaría conformado por 3 factores, 11 indicadores, 66 elementos en  $S$ , 25 parámetros libremente estimados y 41 gl.

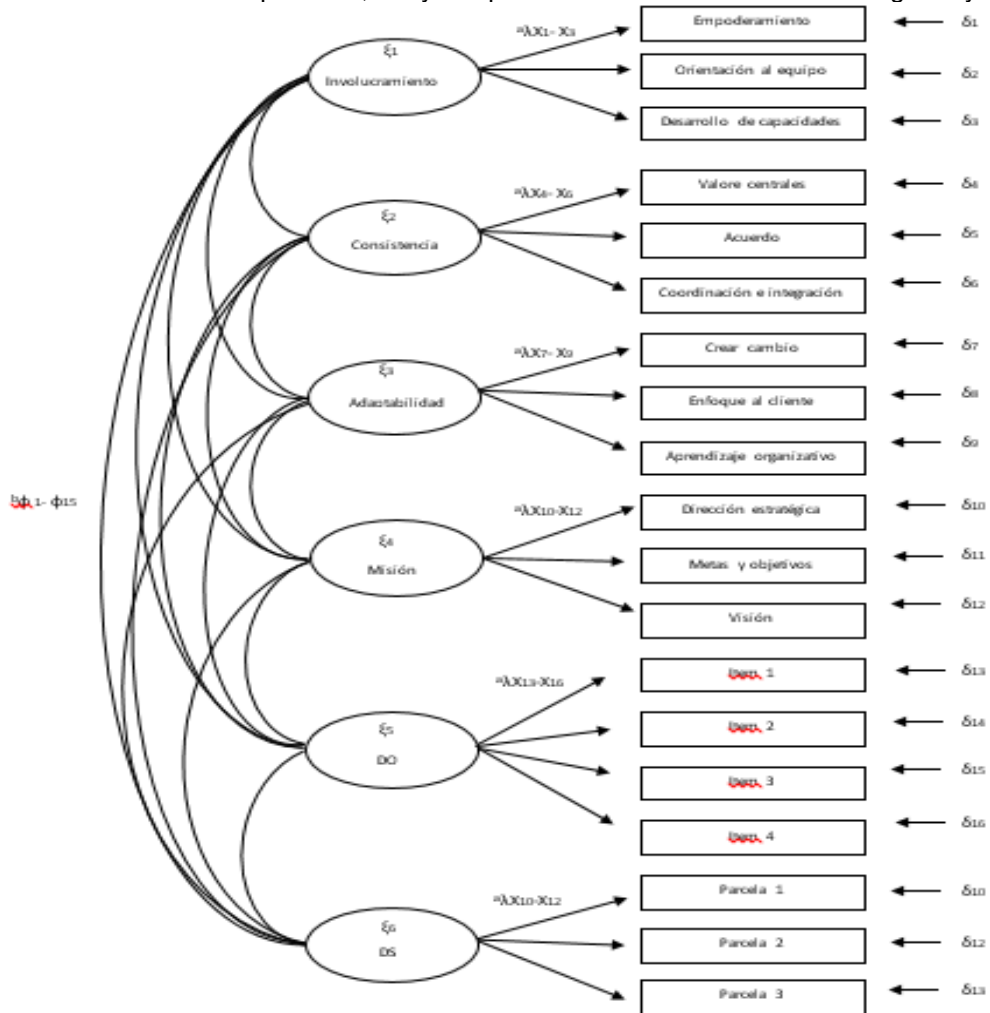
Figura 10  
MDM 2: DOCS con parcelas en un enfoque de desagregación parcial



**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Por razones de espacio solo se indica el rango de las cargas factoriales ( $\lambda X$ ) de cada factor, pero el modelo contempla la libre estimación de cada una.

Figura 11  
MDM 3: DOCS con parcelas, DS y DO para evidencias de validez convergente y discriminante

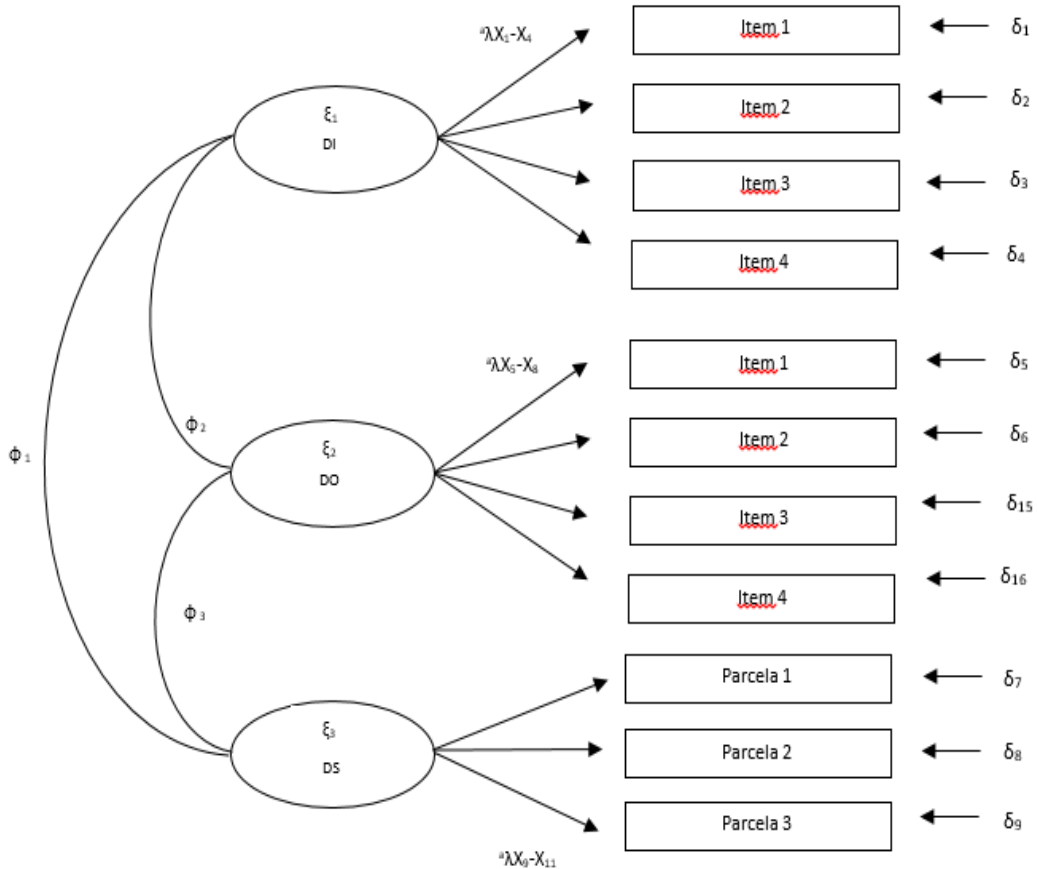


**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Por razones de espacio solo se indica el rango de las cargas factoriales ( $\lambda$ ) de cada factor latente endógeno, pero el modelo contempla la libre estimación de cada uno. <sup>b</sup> Por razones de espacio solo se indica el rango de las covarianzas ( $\phi$ ) entre los factores latentes exógenos, pero el modelo contempla la libre estimación de cada uno.



Figura 12  
MDM 4: DI, DO y DS para evidencias de validez convergente y discriminante

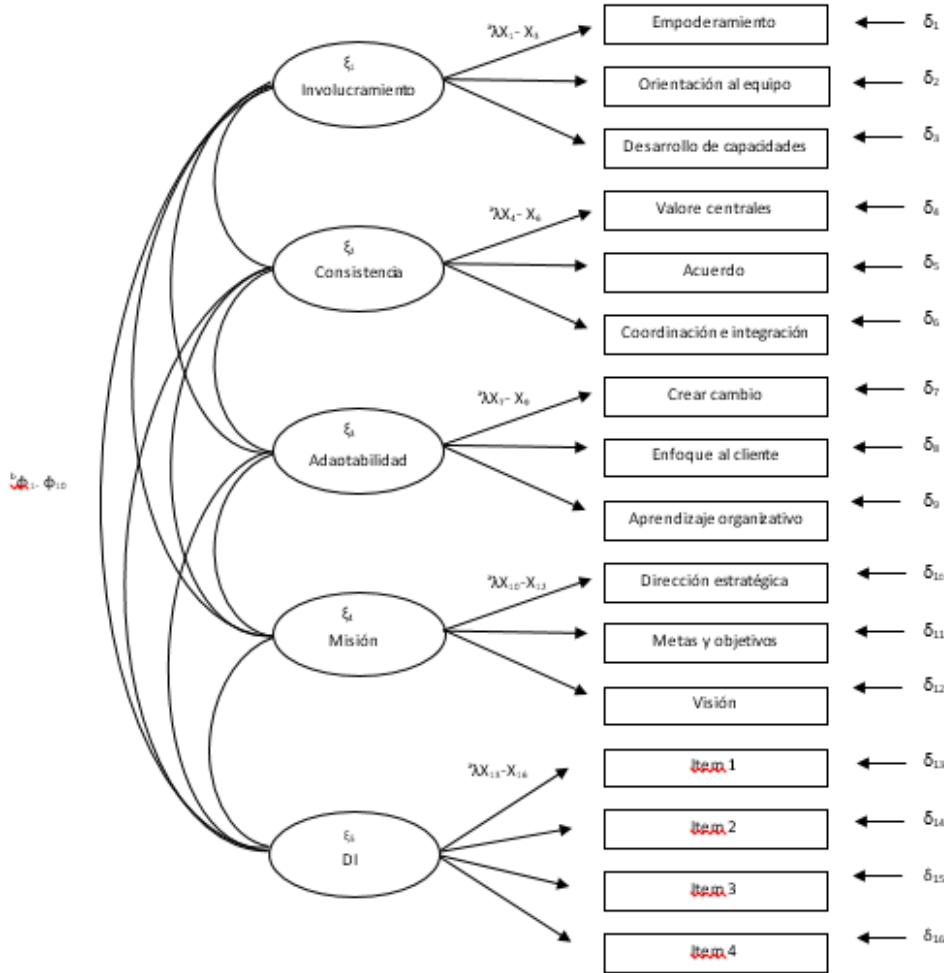


**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Por razones de espacio solo se indica el rango de las cargas factoriales ( $\lambda X$ ) de cada factor latente endógeno, pero el modelo contempla la libre estimación de cada uno. <sup>b</sup> Por razones de espacio solo se indica el rango de las covarianzas ( $\phi_x$ ) entre los factores latentes exógenos, pero el modelo contempla la libre estimación de cada uno.

Por su parte, el objetivo del MDM 5 sería evaluar el ajuste del MDM con los constructos DOCS y DI (ver Figura 13) como recomiendan Hair et al. (2014). Este modelo lo conformarían 5 factores, 16 indicadores y 163 elementos en **S**. La solución requeriría estimar 42 parámetros con 94 gl.

Figura 13  
MDM 5: DOCS con parcelas y DI para el ajuste de estas medidas

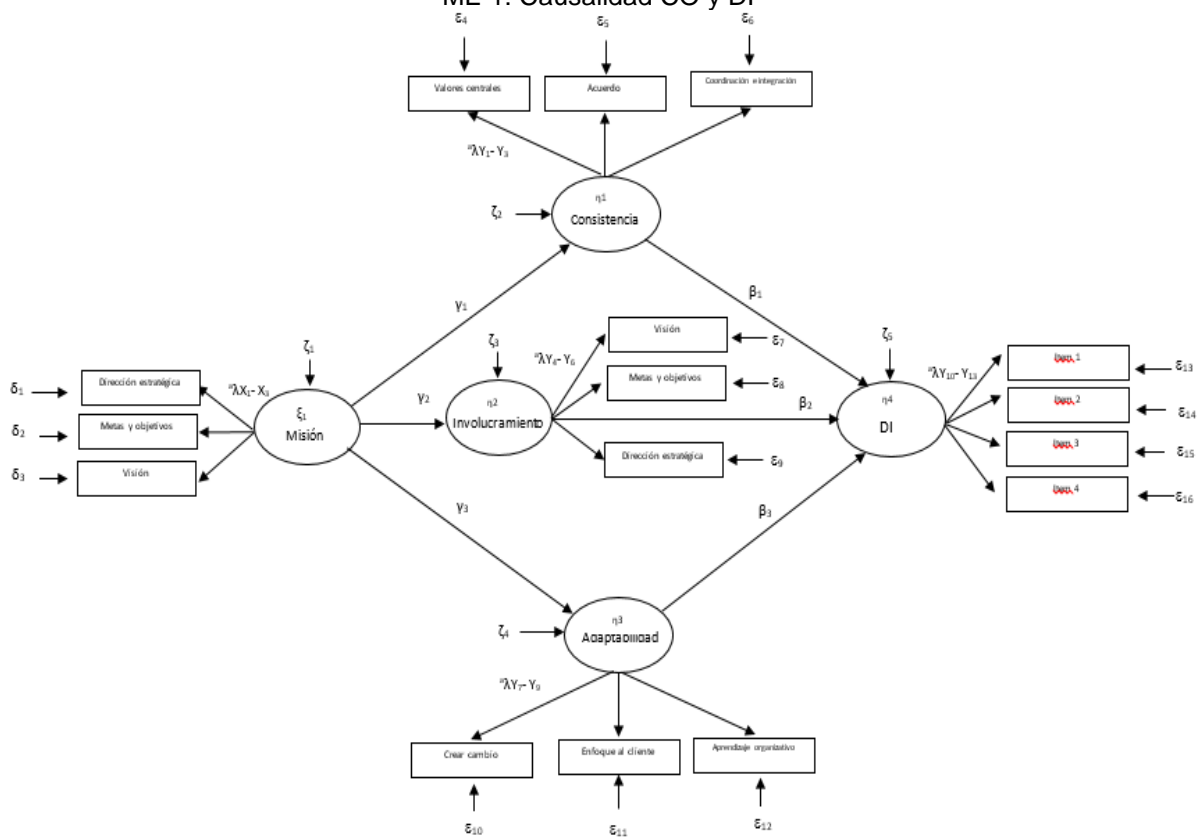


**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Por razones de espacio solo se indica el rango de las cargas factoriales ( $\lambda X$ ) de cada factor latente endógeno, pero el modelo contempla la libre estimación de cada uno. <sup>b</sup> Por razones de espacio solo se indica el rango de las covarianzas ( $\varphi x$ ) entre los factores latentes exógenos, pero el modelo contempla la libre estimación de cada uno.

El sexto modelo corresponde al primer ME que permite someter a prueba las 6 hipótesis propuestas por Vargas-Halabí (2016). En la Figura 14 se puede observar que se trata de un modelo conformado por 5 factores, 4 de ellos serían las dimensiones de la DOCS con parcelas y el constructo DI con 4 indicadores. Este ME 1 plantea la Misión como variable latente exógena y las 3 dimensiones restantes de CO (Adaptabilidad, Consistencia e Involucramiento) como mediadores, en paralelo, sin covarianzas entre ellas. Finalmente, DI como variable latente endógena o dependiente. Con ello, el modelo incluiría 16 indicadores, 136 elementos en **S** y requeriría estimar 38 parámetros con 98 gl (ver Figura 14).

Figura 14  
ME 1: Causalidad CO y DI



**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Por razones de espacio solo se indica el rango de las cargas factoriales ( $\lambda X$ ) de cada factor latente endógeno, pero el modelo contempla la libre estimación de cada uno.

El séptimo modelo o ME 2 contempla la posibilidad de llevar a cabo un análisis *post hoc* o exploratorio, previendo no dejar de lado la posibilidad de realizar ajustes al ME anterior. Se previó incluir nuevos parámetros, tales como: a) errores correlacionados, b) covarianzas entre factores, c) otros efectos directos e indirectos. Por ello, el modelo mantendría las mismas variables latentes e indicadores, con los mismos 136 elementos en **S** pero con posibilidad de incluir 20 parámetros adicionales a los 38. Esta posible solución requeriría estimar libremente 58 parámetros y, por ello, 78 gl.

Seguidamente, se procedió a realizar los cálculos correspondientes al tamaño de *N* para cada modelo, de la misma manera que con el MDM 1. Los resultados fueron compilados en la Tabla 3 para efectos comparativos. Como se puede observar, las primeras tres filas contienen los datos necesarios para los criterios de proporciones y de casos por parámetros. Para efectos de guiar al lector, considérese el ejemplo del MDM 2 con el criterio 2a

(proporción de 20 casos por parámetro) que arroja una  $N$  de 600 casos como resultado de multiplicar 30 parámetros (según la fila 3) por 20. Este procedimiento se repitió en los otros modelos y criterios (2b y 2c) modificando el valor por multiplicar.

Tabla 3  
Estimación de tamaños de muestra de acuerdo con los enfoques y los criterios aportados por la literatura

Criterios <sup>a</sup>	Modelos de medida			Modelos estructurales			Tamaño <sup>b</sup>	
	2	3	4	5	6	7	Mínimo	Máximo
	DOCS parcelas	DOCS convergente discriminante	DI convergente discriminante	DOCS y DI	ME 1 6 hipótesis	ME 2 = ME1 más 20 parámetros		
Número de factores	4	6	3	5	5	5		
Número de indicadores	12	19	11	16	16	16		
Número de parámetros	30	53	25	42	38	58		
Grados de libertad	48	137	41	94	98	78		
<b>1. Absoluto</b>								
1.a mínimo absoluto	200	200	200	200	200	200	200	200
<b>2. Proporciones</b>								
2.a 20 casos por parámetro	600	1060	500	840	760	1160	500	1160
2.b 10 casos por parámetro	300	530	250	420	380	580	250	580
2.c 5 casos por parámetro	150	265	125	210	190	290	125	290
<b>3. Casos por variables</b>								
3.a 8 sujetos por indicador y factor	128	200	112	168	168	168	112	200
3.b 15 sujetos por indicador	180	285	165	240	240	240	165	285
<b>4. Potencia<sup>i</sup></b>								
4.1 Método ajuste global	208	116	229	116	165	188	116	229
4.2.b Método límite inferior	200	177	123	173	173	173	123	200

**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> Según autores Tabla 1 y sección metodológica. <sup>b</sup> Con excepción del MDM 1.

Con el mismo MDM 2, el criterio 3a resultaría en 128 casos, como resultado de multiplicar por 8 la suma de indicadores y factores propuestos en este modelo ( $4 + 12 = 16$ ). En lo referente al criterio 3b, solo se multiplicó el número de indicadores por 15, que en el MDM 2 resulta en 180 ( $15 \times 12 = 180$ ) y así sucesivamente para los otros modelos. En lo relativo al cálculo del tamaño de la  $N$  con el criterio de potencia, recuérdese que utilizaron los mismos criterios que en el MDM 1.

En los enfoques 2 y 3, la diversa complejidad de los modelos planificados arroja tamaños de la  $N$  que varían desde un mínimo 112 casos (criterio 3a para el MDM 4) hasta un

máximo de 1160 casos (criterio 2a para el ME 2). Los criterios que exigen mayores tamaños de la  $N$  son 2a y 2b, especialmente a los MDM 3 y 5, al igual que el posible ME 2 por tener mayor número de parámetros por estimar libremente. La cuarta categoría apunta a un mínimo de 116 casos y un máximo de 229 para detectar un ajuste medio y, entre 123 y 200 casos para detectar el efecto mínimo. Se puede observar que, con un tamaño de  $N \geq 290$  casos, se puede justificar la ejecución de los 7 modelos al cumplir la mayoría de criterios señalados en la literatura, a saber: a) mínimo absoluto, b) el criterio 2c, de 5 casos por parámetro, c) los criterios 3a y 3b de casos por indicadores en los modelos, y d) el criterio de potencia estadística.

La Tabla 4 resulta útil para finalizar con el procedimiento al demostrar cómo el tamaño de  $N$  sugerida cumplió la mayoría de criterios, manteniendo una relación de casos por parámetros superior a 5:1 según Worthington y Whittaker (2006). Ahora bien, con el fin de prever posibles valores perdidos o incrementar la relación de casos por parámetros, el investigador podría elevar el tamaño de la  $N$ , tal como Vargas-Halabí (2016) lo planteó en su investigación al utilizar 314 casos.

Tabla 4  
Casos por parámetros de los 7 posibles modelos planteados a priori

Modelo	Tipo	gl	N.P. <sup>a</sup>	N según criterios de Potencia				N sugerida	C/P <sup>c</sup>	Criterios satisfechos
				MAG <sup>b</sup>	C/P <sup>c</sup>	MLI <sup>d</sup>	C/P <sup>c</sup>			
DOCS parcelas	MDM2	48	30	208	6.9	200	6,8	290	6,9	1a, 2b, 2c, 3a, 3b, 4
DOCS parcelas, convergente-discriminante	MDM3	137	53	116	2.2	177	3,3	290	5,8	1a, 2c, 3a, 3b, 4
DI, convergente-discriminante	MDM4	41	25	229	9.2	123	4,9	290	11,6	1a, 2b, 2c, 3a, 3b, 4
DOCS y DI	MDM5	94	42	116	2.8	173	4,1	290	6,9	1a, 2c, 3a, 3b, 4
ME1: 6 hipótesis	ME1	98	38	165	4.3	173	4,6	290	7,6	1a, 2c, 3a, 3b, 4
ME2 = ME1 reespecificado	ME2	78	58	188	3.2	173	3,0	290	5,0	1a, 2c, 3a, 3b, 4

**Fuente:** Elaboración propia con base en Vargas-Halabí (2016).

<sup>a</sup> N.P.= Número de parámetros <sup>b</sup> MAG= Método ajuste global <sup>c</sup> C/P= Casos por parámetro <sup>d</sup> MLI= Método del límite inferior.

## 5. Reflexiones finales

Este artículo ha reunido diversos criterios sugeridos en la literatura para determinar el tamaño de  $N$  para SEM, los cuales fueron agrupados en cuatro categorías. Asimismo, ha ejemplificado una metodología práctica para sintetizar sus cálculos mediante la formulación, *a priori*, de MDM y ME. Esta requiere de modelos fundamentados teóricamente y acordes con el

diseño de investigación, así como su presentación gráfica para facilitar la identificación de los insumos previos, específicamente, el posible número de: a) indicadores, b) factores, c) parámetros por estimar libremente, d) grados de libertad. Asimismo, es necesario definir el tamaño del efecto para el ajuste global del modelo y los parámetros del modelo.

Con la planificación previa de los modelos, la metodología propuesta permite reunir los tamaños de  $N$  e identificar rangos mínimos y máximos exigidos por los diferentes criterios para cada uno de los modelos *a priori*. Con ello, se obtendrá un umbral de casos que permita satisfacer la mayoría de criterios y, a su vez, corroborar, al menos, el cumplimiento mínimo de relación de casos por parámetros y no exceder límites de modelos extremadamente sensibles. Este proceder resultaría más útil que depender de una única regla o criterio para justificar el tamaño de  $N$  al utilizar SEM. En este sentido, no siempre será suficiente basar esta decisión con el criterio fijo de 200, mencionado en la literatura por diversos autores. La planificación previa ha mostrado que modelos de mayor complejidad, esto es, con mayor número de factores, indicadores y parámetros por estimar, podrían exigir tamaños de la  $N$  mayores.

De igual forma, el uso de parcelas podría representar una ventaja en la planificación de modelos en SEM (MDM y ME). Su uso coadyuvaría, por un lado, a reducir el tamaño de  $S$  respecto a modelos bajo el enfoque de desagregación total y, por otro lado, permitiría lograr soluciones por estimar más parsimoniosas. Asimismo, conduciría a una reducción en el tamaño de  $N$ , especialmente en los criterios más exigentes como aquellos basados en el número de parámetros por estimar. Sin embargo, el investigador debe recordar que, a pesar del inconveniente para interpretar constructos multidimensionales con el uso de parcelas, en su uso subyace el supuesto de correlaciones altas entre las subdimensiones que conformarían el constructo multidimensional. Lo anterior demanda, en el investigador, una fundamentación de estas correlaciones con evidencia teórica y de estudios empíricos previos.

El método propuesto está basado en una revisión rigurosa de la literatura. Sin embargo, hay aspectos que no se han incluido, por ejemplo, el caso de las simulaciones Monte Carlo. Esto por cuanto el objetivo del presente artículo ha sido realizar una contribución metodológica práctica y rigurosa para las y los investigadores, lo cual no significa que se hayan incluido todos los criterios mencionados en la literatura ni que la discusión en torno al tamaño de la  $N$  en SEM se haya resuelto. En este sentido, es importante advertir a las y los investigadores que utilizan SEM que en la literatura especializada pueden encontrar otros criterios que no se han incluido en esta metodología y que constituyen valiosas que pueden

valorar para sus trabajos con SEM. Lo que se ha presentado es una metodología que combina algunos de los criterios más importantes en torno al tema, pero no agota la discusión en torno al tema del tamaño en SEM. Por tal razón, los autores recomiendan continuar los esfuerzos dirigidos a mejorar propuestas metodológicas como la presentada en este artículo.

## Referencias

- Bentler, Peter. (2006). *EQS 6 Structural Equations Program Manual*. California: Multivariate Software, Inc.
- Brown, Timothy y Moore, Michael. (2012). Confirmatory factor analysis. En Rick Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-378). New York: The Guilford Press.
- Byrne, Barbara. (2006). *Structural equations modeling with EQS: Basics concepts, applications and programming*. New York: Routledge.
- Byrne, Barbara. (2010). *Structural equations modeling with AMOS: Basic concepts, applications and programming*. New York: Routledge.
- Catena, Andrés, Ramos, Manuel y Trujillo, Humberto. (2003). *Análisis multivariado: un manual para investigadores*. España: Biblioteca Nueva, S.L.
- Cea, María Ángeles. (2002). *Análisis mutivariable. Teoría y práctica en la investigación social*. Madrid: Editorial Síntesis.
- Cohen, Barry. (2008). *Explaining Psychological Statistics*. (Tercera edición). New Jersey: John Wiley & Sons, Inc.
- Cohen, Jacob. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2<sup>da</sup> ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cumming, Geoff. (2012). *Understanding the new statistics: Effect size, confidence intervals and meta-analysis*. New York: Taylor & Francis.
- Dattalo, Patrick. (2008). *Determining sample size: Balancing power, precision and practicality*. New York: Oxford University Press.
- Denison, Daniel y Mishra, Aneil. (1995). Toward a theory of organizational culture and effectiveness. *Organization Science*, 6(2), 201-223.
- DeVellis, Robert. (2012). *Scale development: Theory and applications* (3<sup>a</sup> ed.). California: SAGE.
- Eid, Michale, Lischetzke, Tanja y Nussbeck, Fridtjof. (2006). Structural equation models for multitrait- multimethod data. En Michael. Eid, y Ed Diener (Eds.), *Handbook of multimethod measurement in Psychology* (pp. 283-299). United States: APA.

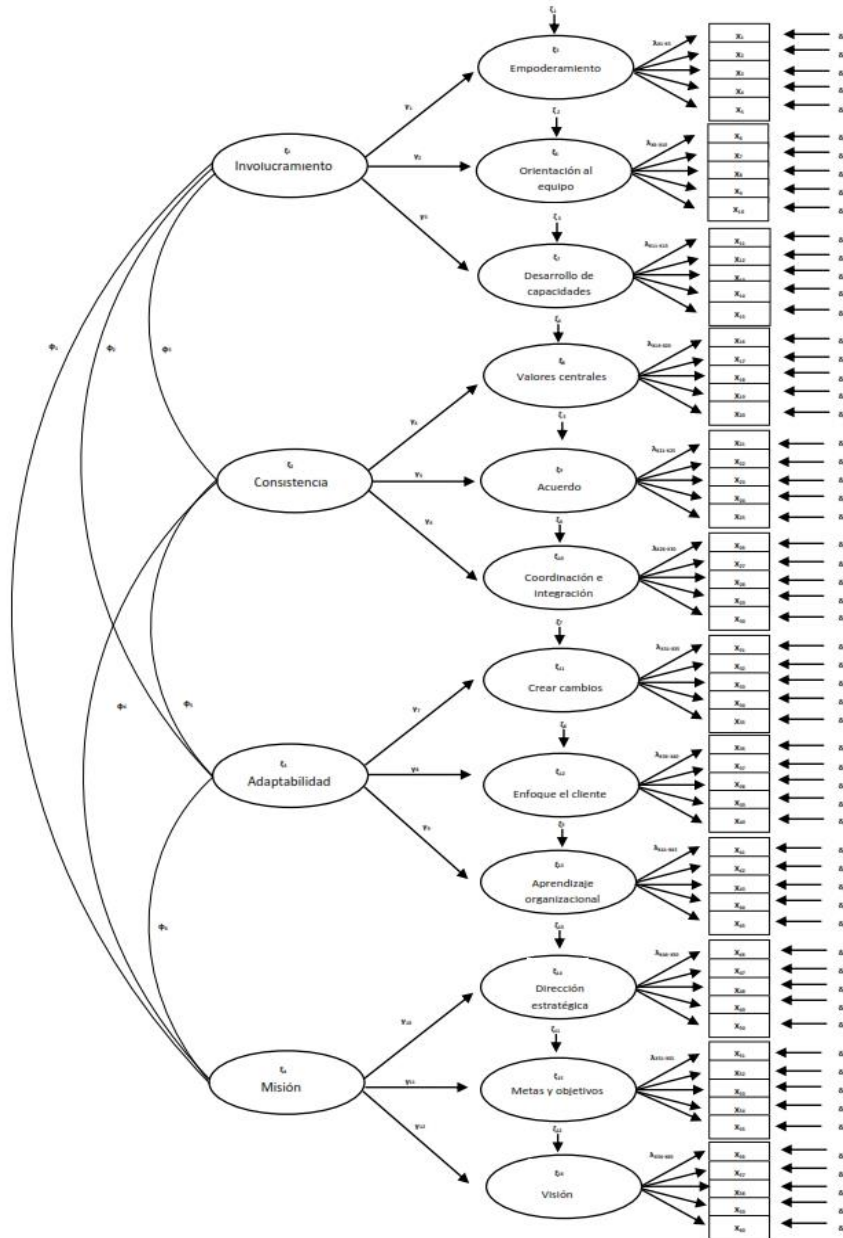
- Ellis, Paul. (2010). *The Essential guide to effect sizes: Statistical power, meta-analysis and the interpretation of research results*. New York: Cambridge University Press.
- Hair Jr., Joseph, Black, William, Babin, Barry y Anderson, Rolph. (2014). *Multivariate Data Analysis: Pearson new international edition (7ª ed.)*. New Jersey: Essex: Pearson.
- Howell, David. (2010). *Statistical Methods for Psychology (7ª ed.)* California: Cengage Wadsworth.
- Hoyle, Rick. (2012). Introduction and Overview. En Rick Hoyle. *Handbook of Structural Equation Modeling* (pp. 3-16). New York: The Guilford Press.
- Khine, Myint Swe. (2013). *Structural Equations Modeling in Educational Research and Practice*. En Myint Swe, Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 279-283). Boston: Sense Publishers.
- Kline, Rex. (2011). *Principles and Practice of Structural Equation Modeling (3ª ed.)*. New York: The Guilford Press.
- Lee, Taehun, Cai, Lli y MacCallum, Robert. (2012). Power analysis for tests of Structural Equation Models. En Rick Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 181-194). New York: The Guilford Press.
- Lévy, Jean-Pierre y González, Nuria. (2006). Modelización y causalidad. En Jean-Pierre Lévy (Dir.) y Jesús, Varela (Coord.), *Modelización con estructuras de covarianza en Ciencias Sociales: temas esenciales, avanzados y aportaciones especiales* (pp. 155-173). España: Netbiblo.
- Lévy, Jean-Pierre, Martín, María y Román, María. (2006). Optimización según estructuras de covarianzas. En Jean-Pierre Lévy (Dir.) y Jesús Varela (Coord.), *Modelización con estructuras de covarianza en Ciencias Sociales* (pp. 11-30). España: Netbiblo.
- Little, Todd, Cunningham, William, Shahar, Golan y Widaman, Keith. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151-173.
- MacCallum, Robert, Browne, Michael y Sugawara, Hazuki. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2) 130-149.
- MacKinnon, David. (2008). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.
- Martínez María Rosario, Hernández María Victoria y Hernández, María José. (2006). *Psicometría*. Madrid: Alianza Editorial, S.A.
- Murphy Kevin, Myers, Brett y Wolach, Allen. (2014). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Routledge: New York.



- Preacher, Kristopher y Coffman, Donna. (2006). *Computing power and minimum sample size for RMSEA* [Computer software]. Recuperado de <http://www.quantpsy.org/rmsear/rmsear.htm>
- Ritchey, Ferris. (2008). *Estadística para las Ciencias Sociales*. México: McGraw-Hill Interamericana.
- Ryan, Thomas. (2013). *Sample Size Determination and Power*. New Jersey: Wiley.
- Sackmann, Sonja. (2011). Culture and performance. En Neal Ashkanasy, Celeste Wilderon y Mark Peterson (Eds.), *The Handbook of organizational culture and climate* (pp. 188-224). California: SAGE.
- Saris, Willen y Satorra, Albert. (1993). Power evaluation in structural equation models. En Kenneth Bollen y Scott Long (Eds.), *Testing structural equation models* (pp. 163-180). California: SAGE.
- Schäffer, Utz. (2007). *Management Accounting & Control Scales Handbook*. Germany: Gabler Edition Wissenschaft.
- Schumacker, Randall y Lomax, Richard. (2010). *A beginner's guide to structural equation modeling*. New York: Routledge Taylor & Francis Group.
- Smith-Castro, Vanessa, Molina, Mauricio y Castelain, Thomas. (2014). Escala de Deseabilidad Social de Crowne y Marlowe. En Vanessa Smith-Castro (Comp.), *Compendio de Instrumentos de Medición IIP-2014* (pp. 143-146). San José: Instituto de Investigaciones Psicológicas.
- Soper, Daniel. (2015). *A-priori Sample Size Calculator for Structural Equation Models*. [Programa de Cómputo]. Recuperado de [www.danielsoper.com/statcalc](http://www.danielsoper.com/statcalc)
- Stevens, James. (2009). *Applied multivariate statistics for the social sciences* (5a. ed.). New Jersey: Routledge.
- Teo, Timothy, Tsai, Liang T., y Yang, Chih-Chien. (2013). Applying Structural Equation Modeling (SEM) in Educational Research: An introduction. En Myint Swe, Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 3-21). Boston: Sense Publishers.
- Ullman, Jodie. (2013). Structural equation modeling. In Barbara Tabachnick y Linda Fidell (Eds.), *Using Multivariate Statistics* (pp. 681-785). Boston: Pearson.
- Varela, Jesús, Abalo, Javier, Rial, Antonio y Braña, Teresa. (2006). Análisis factorial confirmatorio de segundo nivel. En Jean-Pierre Lévy (Dir.) y Jesús, Varela (Coord.), *Modelización con estructuras de covarianza en Ciencias Sociales: Temas esenciales, avanzados y aportaciones especiales* (pp. 239-258). España: Netbiblo.
- Vargas-Halabí, Tomás. (2016). *Cultura organizativa e innovación: un modelo explicativo* (Tesis doctoral). Universidad de Valencia, España.

- Westland, Christopher. (2010). Lower bounds on simple size in structural equation modeling. *Electronic Commerce Research and Applications*, 9(6), 476-487.
- Westland, Christopher. (2015). *Structural Equation Models: From paths to networks*. New York: Springer.
- Williams, Larry, Vandenberg, Robert y Edwards, Jeffrey. (2009). Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals*, 3(1), 543-604. Recuperado de <http://www.tandfonline.com/doi/abs/10.1080/19416520903065683>
- Worthington, Roger y Whittaker, Tiffany. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838.

Figura A  
MDM 1: DOCS con un enfoque de desagregación total



Fuente: Elaboración propia con base en Vargas-Halabí (2016)