

ANÁLISIS FACTORIAL CONFIRMATORIO. RECOMENDACIONES SOBRE MÍNIMOS CUADRADOS NO PONDERADOS EN FUNCIÓN DEL ERROR TIPO I DE JI-CUADRADO Y RMSEA

CONFIRMATORY FACTOR ANALYSIS. RECOMMENDATIONS FOR UNWEIGHTED LEAST SQUARES METHOD RELATED TO CHI- SQUARE AND RMSEA TYPE I ERROR

M^a. A. MORATA-RAMÍREZ¹, FRANCISCO P. HOLGADO-TELLO¹,
ISABEL BARBERO-GARCÍA¹ Y GONZALO MENDEZ²

Cómo referenciar este artículo/How to reference this article:

Morata-Ramírez, M. A., Holgado-Tello, F.P., Barbero-García, I. y Mendez, G. (2015). Análisis Factorial Confirmatorio. Recomendaciones sobre Mínimos Cuadrados no Ponderados en función del error Tipo I de Ji-cuadrado y RMSEA [Confirmatory Factory Analysis. Recommendations for Underweigted Least Squares method related to Chi-square And RMSEA Type I error]. *Acción Psicológica*, 12(1), 79-90. doi: <http://dx.doi.org/10.5944/ap.12.1.14362>

Resumen

En Psicología, para obtener evidencias sobre validez de constructo mediante Análisis Factorial Confirmatorio es habitual trabajar con variables ordinales que presentan asimetría. En este estudio de simulación se analiza el comportamiento del método de Mínimos Cuadrados no Ponderados (ULS) en escalas tipo Likert con base en los índices χ^2 de razón de verosimilitud (C2) y RMSEA. Para ello, se han manipulado cuatro factores experimentales: el número de factores o dimensiones (2, 3, 4, 5, 6), número de puntos de respuesta (3, 4, 5, 6),

grado de asimetría de la distribución de respuestas (simétrica, asimétrica moderada y severa) y tamaño muestral (100, 150, 250, 450, 650, 850) de los modelos simulados. Según los principales resultados, el índice C2 muestra siempre un error Tipo I mayor que RMSEA, con independencia de los factores experimentales analizados. Finalmente, se discuten diferentes alternativas de acción y se presentan futuras líneas de investigación.

Palabras clave: Análisis Factorial Confirmatorio; método ULS; escalas Likert; error Tipo I; índice χ^2 de razón de verosimilitud; RMSEA.

Correspondencia: Francisco Pablo Holgado Tello. Universidad Nacional de Educación a distancia.

Email: pfolgado@psi.uned.es

¹ Universidad Nacional de Educación a Distancia (UNED)

² Universidad Complutense de Madrid

Recibido: 09 de abril de 2015

Aceptado: 24 de mayo de 2015

Abstract

In order to obtain evidences about construct validity through Confirmatory Factor Analysis in Social Sciences, working with skewed ordinal variables has been usual. In this simulation study the performance of Unweighted Least Squares (ULS) method in Likert scales according to Likelihood Ratio Test (C2) and RMSEA indices is analysed through Type I error. For this purpose, four experimental factors have been manipulated: the number of factors or dimensions (2, 3, 4, 5, 6), the number of response points (3, 4, 5, 6), the degree of skewness of the responses distribution (symmetric, moderately and severely asymmetric) and the sample size (100, 150, 250, 450, 650, 850) of the simulated models. According to the main results, C2 index always shows a bigger Type I error than RMSEA, regardless of the experimental factors analysed. Finally, different action alternatives are discussed and future research lines are presented.

Keywords: Confirmatory Factor Analysis; ULS method; Likert scales; Type I error; Likelihood Ratio Test; RMSEA.

Introducción

El análisis factorial es el método más aplicado para obtener evidencias sobre validez de constructo, dado que informa sobre la estructura interna de los instrumentos de medida (Zumbo, 2007).

Grosso modo, en el Análisis Factorial Confirmatorio (AFC) inicialmente se especifica e identifica el modelo. Tras la recogida de datos, se estiman los parámetros y, a continuación, se evalúa el ajuste del modelo. Si el modelo presenta un ajuste adecuado, se podrá utilizar para evaluar e interpretar los parámetros. En caso contrario, se modificará el modelo y de nuevo se iniciará el proceso (Batista y Coenders, 2000; Barbero, Vila y Holgado, 2011; Catena, Ramos y Trujillo, 2003; Fadlelmula, 2011).

El AFC exige el cumplimiento de ciertos supuestos que, en su mayoría, entran en contradicción con la naturaleza de los datos que habitualmente se obtienen en Psicología mediante el uso de escalas tipo Likert (DiStefano, 2002). La falta de coherencia entre las características del instrumento de recogida de datos -en este caso, los cuestionarios con escalas tipo Likert- y los requisitos de uso del AFC cobra importancia en la fase del AFC de estimación de parámetros al elegir el método adecuado, pues de él dependerá en gran medida que al evaluar el

ajuste del modelo especificado tenga lugar una mayor o menor aproximación a la realidad empírica medida mediante los índices globales de bondad de ajuste.

La normalidad multivariable de las variables observadas es un requisito para el uso de AFC mencionado por Mulaik (1972) que se puede incumplir en escalas tipo Likert (Flora y Curran, 2004). En este sentido, la asimetría en la distribución de las variables observadas es uno de los principales aspectos que distorsionan este supuesto (Coenders y Saris, 1995). Precisamente, el método de estimación de Máxima Verosimilitud (ML), que precisa de normalidad multivariada, es el que se utiliza en la mayoría de Análisis Factoriales Confirmatorios (Brown, 2006).

Otro de los supuestos para el uso del AFC, mencionado por Mulaik (1972), establece que tanto las variables latentes como observadas han de ser continuas. Sin embargo, las escalas tipo Likert asumen que el constructo latente es de naturaleza continua pero las variables observadas que lo representan siguen una escala de medida ordinal (Coenders y Saris, 1995; DiStefano 2002; Flora y Curran, 2004). A pesar de ello, lo habitual ha sido tratar las variables observadas como si también fueran continuas (Cea, 2004). Así, el método de estimación de Máxima Verosimilitud (ML) ha tenido un uso muy extendido (Ryu, 2011) pero, al ser necesario que los datos se midan de acuerdo con una escala de medida de intervalo, quedaría desaconsejado su uso cuando se utilizan escalas tipo Likert (Brown, 2006; Flora y Curran, 2004). Además, los métodos de estimación de parámetros que, como ML, asumen que las variables observadas están medidas de acuerdo con una escala de medida de intervalo, basan sus cálculos en la matriz de correlación de Pearson (Holgado, Chacón, Barbero y Vila, 2010). A este respecto, Jöreskog y Sörbom (1996a) proponen, como alternativa al uso de correlaciones de Pearson, recurrir a correlaciones policóricas. Estas permiten superar los problemas que conllevan su uso, ya analizados por Johnson y Creech (1983) y O'Brien (1985) y tratados posteriormente en diferentes estudios (e.g., Coenders y Saris, 1995; Holgado et al., 2010).

Elegir el método adecuado en la fase del AFC de estimación de parámetros influirá en gran medida en la fase posterior de evaluación el ajuste del modelo teórico especificado. A este respecto, ante las dificultades que plantea el uso de escalas tipo Likert, se puede recurrir al método de Mínimos Cuadrados no Ponderados (en inglés, Unweighted Least Squares o ULS). Se trata de un método de estimación de parámetros para el que no está establecido que las variables observadas deban seguir una distribución determinada, que está recomendado para variables categóricas y que se basa en la matriz de correlaciones

policóricas (Batista y Coenders, 2000; Bollen, 1989; Brown, 2006; Schumacker y Lomax, 1996). Además del método ULS, cuando se manejan variables ordinales también se tienen en cuenta otros métodos alternativos. Entre ellos, se puede destacar el método RULS que, de acuerdo con Yang-Wallentin, Jöreskog y Luo (2010), es una variante robusta del método ULS. Al igual que este último, RULS también trabaja con una matriz de correlaciones policóricas, si bien estas correlaciones son el punto de partida para obtener posteriormente la matriz de covarianzas asintóticas AC que interviene en su matriz W de distribución libre. El programa LISREL proporciona una serie de índices χ^2 de razón de verosimilitud, entre los cuales aparece C3 (Jöreskog, 2004), conocido como “estadístico χ^2 escalado de Satorra-Bentler” (Batista y Coenders, 2000) y cuyos últimos desarrollos corresponden a trabajos como los de Satorra y Bentler (2010). Este índice C3 es uno de los recursos disponibles cuando se manejan variables categóricas o bien las variables dependientes no siguen la distribución normal (Finney y DiStefano, 2013). A su vez, esta matriz AC también es el foco de atención de estudios como el de Jennrick y Satorra (2014).

Además de las aportaciones de Yang-Wallentin et al. (2010), cabe destacar también estudios acerca de RULS como el de Savalei y Rhemtulla (2013). Entre los métodos de estimación de información limitada se cuenta con una variante de RULS: el 3S-RULS (Katsikatsou, Moustaki, Yang-Wallentin y Jöreskog, 2012), así como también con un método que frente a ML muestra buen comportamiento ante errores de especificación de los modelos: el 2SLS (Jung, 2013).

En cuanto a la fase de evaluación del ajuste del modelo del AFC, para conocer en qué grado hay una mayor o menor aproximación del modelo teórico a la realidad empírica se dispone de una amplia variedad de índices. Entre ellos, cabe destacar el índice de razón de verosimilitud χ^2 , procedente del método ML, que es el único que aporta una prueba de significatividad estadística (Cea, 2004; Ryu, 2011). El rechazo de la hipótesis nula a partir de un valor χ^2 significativo ($p < .05$) implica que el modelo teórico propuesto es inadecuado, por lo que es necesario especificarlo de nuevo (Batista y Coenders, 2000; Cea, 2004).

Dadas las limitaciones del índice de razón de verosimilitud por su sensibilidad al tamaño muestral y por fundamentarse en la distribución central de χ^2 (Bollen, 1989; Byrne, 1998), se recomienda complementar sus resultados con otros índices de bondad de ajuste. En este contexto, el índice RMSEA o Error cuadrático medio de aproximación por grado de libertad está reconocido como uno de los más informativos de los modelos en

ecuaciones estructurales. A la hora de determinar la bondad de ajuste tiene en cuenta los grados de libertad, por lo que este índice es sensible al número de parámetros que estima el modelo (Barbero et al., 2011; Byrne, 1998; Cea, 2004). En este sentido, los valores de RMSEA decrecen conforme aumenta el número de grados de libertad o el tamaño muestral (McCallum, Browne y Sugawara, 1996; Kline, 2011). En general, valores en RMSEA menores de .05 indican un buen ajuste y los valores comprendidos entre .05 y .08 un ajuste razonable (Browne y Cudeck, 1993).

En resumen, al llevar a cabo un AFC es preciso que haya consistencia entre los métodos de estimación y el instrumento de medida para que los modelos teóricos propuestos reproduzcan las relaciones entre las variables de un constructo con la mayor fidelidad posible. En relación con ello y dado los escasos estudios en los que se ha analizado el error Tipo I y comportamiento de RMSEA cuando se usa ULS, este estudio de simulación tiene como objetivo analizar cómo afectan una serie de factores experimentales que caracterizan los modelos teóricos (número de factores, número de puntos de respuesta, grado de asimetría de la distribución de respuestas a los ítems y tamaño muestral) sobre los valores del índice de razón de verosimilitud χ^2 y de RMSEA.

Método

Se manipularon 4 factores experimentales: (a) número de factores o variables latentes, (b) número de puntos de respuesta, (c) grado de asimetría de la distribución de las respuestas a los ítems, y (d) tamaño muestral.

El número de factores presentaba cinco niveles experimentales (2, 3, 4, 5 y 6). Los factores guardaban una relación ortogonal entre ellos, es decir, no estaban correlacionados. Para cada factor se simularon tres ítems, con el propósito de que pudieran ser identificados estadísticamente. En este contexto, hay escalas de regulación emocional como el CERQ (Garnefski y Kraaij, 2007) que presentan un reducido número de ítems (4) para medir cada una de sus nueve dimensiones. También se encuentran escalas con sólo tres ítems en psicología de las organizaciones (Holgado, Chacón, Barbero y Sanduverte, 2006).

Las saturaciones factoriales de los ítems siempre eran las mismas en todos los factores, es decir, .9, .8 y .7 para el primer, segundo y tercer ítem, respectivamente. Los ítems se generaron según una distribución normal $N(0, 1)$. A continuación, estas respuestas se categorizaron de acuerdo con una escala tipo Likert de 3, 4, 5 y 6 puntos, es decir, el número de puntos de respuesta quedó configu-

rado con cuatro niveles experimentales. Las escalas tipo Likert se categorizaron de tal modo que: (a) se mantuviera simétrica la distribución de las respuestas a todos los ítems, (b) la distribución de respuestas presentara asimetría moderada en todos los ítems o (c) la distribución de respuestas presentara asimetría severa en todos los ítems. De esta forma el grado de asimetría presentaba tres niveles experimentales: asimetría = 0, asimetría = 1 (o moderada) y asimetría = 2 (o severa).

Para categorizar las escalas Likert, siguiendo a Bollen y Barb (1981), el continuum se dividió en intervalos iguales desde $z = -3$ a $z = 3$ con el fin de calcular los umbrales de la condición en la que la distribución de respuestas a todos los ítems fuera simétrica (asimetría = 0). Para distribuciones asimétricas los umbrales se calcularon, de acuerdo con Muthén y Kaplan (1985), de manera que las observaciones se acumularan en uno de los extremos de los puntos de respuesta conforme el grado de asimetría se incrementara. La mitad de las variables de cada factor se categorizaron con la misma asimetría positiva y el resto de variables con la misma asimetría negativa con el propósito de simular factores de dificultad. Finalmente, el tamaño muestral presentaba seis valores experimentales: 100, 150, 250, 450, 650 y 850 sujetos.

La combinación de los cuatro factores experimentales (número de factores, número de puntos de respuesta, asimetría de los ítems y tamaño muestral) produjo 360 condiciones experimentales ($5 \times 4 \times 3 \times 6$) que se replicaron en 500 ocasiones. Para ello, estas repeticiones se realizaron mediante la versión 2.12.0 del programa R (R Development Core Team, 2010), que invocaba sucesivamente a PRELIS 2.0 (Jöreskog y Sörbom, 1996b) para la generación de las correspondientes matrices de datos de acuerdo con la especificaciones resultantes de la combinación de las condiciones experimentales. Para cada matriz generada se obtenía la matriz de correlaciones policóricas.

Una vez obtenidas las matrices de correlaciones para cada matriz de datos generada bajo las especificaciones concretas de los factores experimentales, se ejecutó el correspondiente Análisis Factorial Confirmatorio sucesivamente, es decir, hasta 500 veces (uno por cada replicación). Al igual que en el caso anterior, se utilizó la versión 2.12.0 del programa R (R Development Core Team, 2010), que invocaba sucesivamente a LISREL 8.8 (Jöreskog y Sörbom, 1996a).

Para facilitar la gestión y compilación de ficheros de índices de ajuste se generó un programa específico en lenguaje JAVA.

Análisis de datos

El error Tipo I se obtuvo calculando el porcentaje de veces que se rechaza la hipótesis nula del índice de razón de verosimilitud χ^2 en modelos especificados correctamente. A este respecto se ha considerado el valor nominal del 5%, que en la práctica se traduce en una probabilidad inferior a .05. Cabe recordar que al utilizar el método ULS el programa LISREL denomina a este valor χ^2 como C2, también conocido como “Normal Theory Weighted Least Squares Chi-Square” (Jöreskog, 2004). Por otra parte, para plantear la hipótesis nula respecto al índice RMSEA se ha determinado que se acepten los modelos teóricos cuando su valor es inferior a .08. Así, dentro de los modelos aceptados se incluyen también aquellos que presentan, siguiendo a Browne y Cudeck (1993), un ajuste razonable.

Resultados

Influencia del número de factores

En la Tabla 1 se presenta un análisis del error Tipo I en C2 y en RMSEA según el número de factores de los modelos teóricos a prueba. Además, en la tabla se indica la media y desviación típica de la probabilidad asociada a C2 y RMSEA para cada uno de los niveles del factor.

Como se puede ver en la Tabla 1, el valor porcentual del error Tipo I aumenta a medida que lo hace el número de factores. Concretamente, la probabilidad de rechazar un modelo correcto se eleva del 58.9% para modelos con dos factores al 93.9% para modelos con seis factores. Por tanto, se deduce que a medida que crece el número de factores se observa una tendencia decreciente en la aceptación de modelos mediante el índice χ^2 de razón de verosimilitud. Así, de aceptar el 41.1% de modelos con 2 factores se desciende al 6.1% con seis factores. Esta tendencia es coherente con las medias obtenidas respecto al número de factores. De esta forma, cuando el número de factores es dos, el valor medio de la probabilidad asociada a C2 es .145 mientras que, a medida que aumentan los factores del modelo teórico, la probabilidad asociada a C2 va disminuyendo.

En cuanto al índice RMSEA, el valor porcentual del error Tipo I se mantiene en valores en torno al 35-37% aproximadamente. Consecuentemente, se constata que el porcentaje de aceptación de los modelos se sitúa hacia el 62-64%, independientemente del número de factores que presenten. Por su parte, las medias de RMSEA se mantienen estables independientemente del número de factores.

Tabla 1

Influencia del número de factores

	Factores	Media	Desv. típ.	Error Tipo I (%)
C2	2	.145	.233	58.9
	3	.069	.164	78.0
	4	.039	.122	86.4
	5	.023	.090	90.9
	6	.014	.067	93.9
	Total	.058	.155	
RMSEA	2	.078	.071	37.6
	3	.075	.058	35.5
	4	.077	.056	36.2
	5	.079	.056	36.5
	6	.078	.056	35.9
	Total	.077	.060	

Nota. **Media:** Valor medio de la probabilidad asociada a cada índice de bondad de ajuste. **Desv. Típica:** Desviación típica de la probabilidad asociada a cada índice de bondad de ajuste. **Error Tipo I (%):** Porcentaje de modelos con buen ajuste que son rechazados. **C2:** Índice de razón de verosimilitud χ^2 obtenido para el método ULS. **RMSEA:** Error cuadrático medio de aproximación por grado de libertad.

Influencia del número de puntos de respuesta

En la Tabla 2 se presentan los resultados relativos al número de puntos de respuesta de los modelos.

Al utilizar el índice χ^2 de razón de verosimilitud, el error Tipo I desciende a medida que los modelos teóricos presentan un mayor número de puntos de respuesta. Así, mientras el porcentaje de aceptación de los modelos con tres puntos de respuesta es 92.6%, el correspondien-

te a modelos con seis puntos baja al 70.5%. En otras palabras, con un mayor número de puntos de respuesta disminuye la probabilidad de rechazar un modelo. Así, el porcentaje de modelos aceptados va ascendiendo conforme lo hace el número de puntos de respuesta, siendo dicho porcentaje del 7.4% para tales modelos con tres puntos de respuesta y del 29.5% para modelos con seis puntos. En este contexto, conforme crece el número de puntos de respuesta los valores medios de la probabilidad asociada a C2 también experimentan un incremento. Así, dichos valores pasan de .021 para modelos con tres puntos de respuesta a .099 para modelos con seis puntos.

Tabla 2

Influencia del número de puntos de respuesta

	Puntos de respuesta	Media	Desv. típ.	Error Tipo I (%)
C2	3	.021	.091	92.6
	4	.041	.128	86.2
	5	.072	.169	76.9
	6	.099	.198	70.5
	Total	.058	.155	
RMSEA	3	.095	.069	48.3
	4	.084	.062	39.6
	5	.069	.052	31.1
	6	.061	.048	26.3
	Total	.077	.060	

Nota. **Media:** Valor medio de la probabilidad asociada a cada índice de bondad de ajuste. **Desv. Típica:** Desviación típica de la probabilidad asociada a cada índice de bondad de ajuste. **Error Tipo I (%):** Porcentaje de modelos con buen ajuste que son rechazados. **C2:** Índice de razón de verosimilitud χ^2 obtenido para el método ULS. **RMSEA:** Error cuadrático medio de aproximación por grado de libertad.

Tabla 3

Influencia del grado de asimetría

	Asimetría	Media	Desv. típ.	Error Tipo I (%)
C2	0	.140	.220	57.3
	1	.028	.105	90
	2	.006	.046	97.5
	Total	.058	.155	
RMSEA	0	.041	.029	10.7
	1	.074	.042	36.3
	2	.117	.072	62.1
	Total	.077	.060	

Nota. Media: Valor medio de la probabilidad asociada a cada índice de bondad de ajuste. **Desv. Típica:** Desviación típica de la probabilidad asociada a cada índice de bondad de ajuste. **Error Tipo I (%):** Porcentaje de modelos con buen ajuste que son rechazados. **C2:** Índice de razón de verosimilitud χ^2 obtenido para el método ULS. **RMSEA:** Error cuadrático medio de aproximación por grado de libertad.

Con RMSEA, cuanto mayor es el número de puntos de respuesta, menor será la probabilidad de cometer el error Tipo I. Concretamente, el porcentaje de modelos rechazados desciende del 48.3% para modelos con tres puntos de respuesta al 26.3% de modelos que tienen seis puntos. Por esta razón hay una tendencia creciente en la cantidad de modelos aceptados conforme aumenta el número de puntos de respuesta. Así, se puede deducir que se acepta el 51.7% de estos modelos con 3 puntos de respuesta y el 73.7% de modelos con 6 puntos. Por su parte, la tendencia mostrada por las medias de RMSEA es descendente a medida que se cuenta con más puntos de respuesta.

Influencia del grado de asimetría

La Tabla 3 muestra los resultados obtenidos en relación con el grado de asimetría de la distribución de respuestas a todos los ítems que presentan los modelos teóricos.

En relación con el grado de asimetría de la distribución de respuestas, el índice χ^2 de razón de verosimilitud muestra que el error Tipo I aumenta a medida que también se incrementa la asimetría de la distribución de las respuestas, especialmente cuando la distribución de los modelos pasa de ser simétrica (siendo su porcentaje de rechazo del 57.3%) a tener asimetría moderada (con un porcentaje de rechazo del 90%). En otras palabras, a medida que aumenta el grado de asimetría, es más probable rechazar un modelo teórico. En este sentido, el índice χ^2 tiene mal comportamiento, disminuyendo el porcentaje de aceptación de tales modelos a medida que el grado de asimetría es mayor. Puede deducirse así un descenso muy acentuado al pasar del 42.7% de modelos

aceptados con distribuciones simétricas al 10 % cuando la distribución de respuestas de los modelos teóricos presenta asimetría moderada. Asimismo, cabe señalar que los valores medios de la probabilidad asociada a C2 disminuyen conforme aumenta la asimetría de la distribución de respuestas. Concretamente, de .140 para distribuciones simétricas se pasa a .006 para distribuciones con asimetría severa.

Por lo que respecta al índice RMSEA, el error Tipo I va aumentando con el grado de asimetría, siendo del 10.7% y del 62.1% para modelos teóricos con distribución simétrica y asimétrica severa, respectivamente. Por tanto, se deduce un descenso en el porcentaje de aceptación de los modelos, que desciende del 89.3% para modelos con distribución simétrica al 37.9% para modelos con asimetría severa en su distribución de respuestas. Igualmente, se observan incrementos en el valor medio de RMSEA, a medida que aumenta la asimetría de la distribución de respuestas.

Influencia del tamaño muestral

La Tabla 4 muestra los resultados relacionados con el tamaño muestral de los modelos teóricos.

En cuanto al índice χ^2 de razón de verosimilitud la elevada probabilidad de cometer el error Tipo I permanece estable hacia el 80-82%, con independencia del tamaño muestral de los modelos teóricos a prueba. En relación con ello, no se observa consecuentemente relación entre el tamaño muestral y el porcentaje de aceptación de los modelos, que oscila entre el 17% y el 19% aproximadamente. La estabilidad en los resultados también queda

Tabla 4

Influencia del tamaño muestral

	Muestra	Media	Desv. típ.	Error Tipo I (%)
C2	100	.056	.151	81.7
	150	.056	.152	82.3
	250	.057	.155	82.1
	450	.058	.156	81.7
	650	.061	.156	80.6
	850	.060	.158	81.1
	Total	.058	.155	
RMSEA	100	.133	.087	74.5
	150	.106	.058	65.7
	250	.079	.042	46.4
	450	.059	.035	23.3
	650	.046	.023	6.6
	850	.043	.029	4.4
	Total	.077	.060	

Nota. **Media:** Valor medio de la probabilidad asociada a cada índice de bondad de ajuste. **Desv. Típica:** Desviación típica de la probabilidad asociada a cada índice de bondad de ajuste. **Error Tipo I (%):** Porcentaje de modelos con buen ajuste que son rechazados. **C2:** Índice de razón de verosimilitud χ^2 obtenido para el método ULS. **RMSEA:** Error cuadrático medio de aproximación por grado de libertad.

reflejada en los valores medios de probabilidad asociada a C2, que oscilan entre .056 y .061.

Por otra parte, para RMSEA se observa una disminución en el error Tipo I a medida que el número de sujetos de los modelos es mayor. De hecho, mientras que para muestras con 100 sujetos se rechaza el 74.5% de modelos, para 450 sujetos este porcentaje disminuye al 23.3%. Para muestras con 650 y 850 sujetos el porcentaje de rechazo de modelos teóricos continúa disminuyendo, pues sus valores son del 6.6% y 4.4%, respectivamente. De ahí puede deducirse que el porcentaje de aceptación de los modelos mediante RMSEA va incrementándose conforme aumenta el tamaño de la muestra, de manera que se acepta más de la mitad de modelos teóricos a partir de 250 sujetos. En cuanto a los valores medios de la probabilidad asociada al índice RMSEA, puede observarse que se da una disminución en tales valores a medida que se incrementa el tamaño muestral. De esta manera, mientras que para muestras de 100 sujetos se obtiene un valor medio de .133, para muestras de 850 sujetos tal valor ha disminuido a .043.

Anova Multivariante de los factores manipulados. Efectos principales y de interacción.

Finalmente, con el objetivo de determinar la influencia de los factores manipulados sobre los índices de ajuste, se realizó un Anova Multivariante en el que las variables independientes fueron el número de factores, el número de categorías, la asimetría y el tamaño muestral y las variables dependientes fueron las probabilidades del valor Chi-Cuadrado C2 y el índice de ajuste RMSEA (ver Tabla 5). Dado el elevado número de replicaciones realizado, es de esperar que todos los efectos fueran significativos. Por ello, se usó la eta cuadrado parcial como medida del tamaño del efecto. Cohen (1988) establece que un tamaño del efecto igual o superior a 0.14 es suficientemente grande como para ser tenido en cuenta; por el contrario, tamaños iguales o inferiores a 0.01 son pequeños y valores en torno a 0.06 son medianos. Así, nos centraremos en aquellos efectos cuyo tamaño sea igual o superior a 0.14.

Tal y como se observa en la Tabla 5, no encontramos un efecto relevante para el número de factores y sí para el número de categorías en RMSEA ($\eta^2 = .144$), el grado de asimetría para C2 ($\eta^2 = .186$) y RMSEA ($\eta^2 = .516$) y, tal y como cabría esperar, el tamaño muestral afecta significativamente a RMSEA ($\eta^2 = .536$).

Tabla 5

Anova Multivariante. Efectos principales y de interacción

Fuente	Variable dependiente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Factores	C2	396.656	4	99.164	6616.689	.000	.129
	RMSEA	.536	4	.134	144.500	.000	.003
Categorías	C2	152.541	3	50.847	3392.751	.000	.054
	RMSEA	27.846	3	9.282	10016.355	.000	.144
Simetría	C2	611.444	2	305.722	20399.212	.000	.186*
	RMSEA	177.206	2	88.603	95611.980	.000	.516*
Muestra	C2	1.438	5	.288	19.195	.000	.001
	RMSEA	191.385	5	38.277	41305.083	.000	.536*
Factores * categorías	C2	25.110	12	2.092	139.620	.000	.009
	RMSEA	5.242	12	.437	471.391	.000	.031
Factores * simetría	C2	199.355	8	24.919	1662.735	.000	.069
	RMSEA	.479	8	.060	64.654	.000	.003
Categorías * simetría	C2	172.814	6	28.802	1921.821	.000	.060
	RMSEA	1.082	6	.180	194.614	.000	.006
Factores * categorías * simetría	C2	29.206	24	1.217	81.197	.000	.011
	RMSEA	3.489	24	.145	156.888	.000	.021
Factores * muestra	C2	1.790	20	.089	5.971	.000	.001
	RMSEA	2.829	20	.141	152.653	.000	.017
Categorías * muestra	C2	1.271	15	.085	5.652	.000	.000
	RMSEA	9.055	15	.604	651.395	.000	.052
Factores * categorías * muestra	C2	3.872	60	.065	4.306	.000	.001
	RMSEA	6.278	60	.105	112.905	.000	.036
Simetría * muestra	C2	1.572	10	.157	10.491	.000	.001
	RMSEA	40.657	10	4.066	4387.305	.000	.197*
Factores * simetría * muestra	C2	1.915	40	.048	3.195	.000	.001
	RMSEA	1.372	40	.034	37.010	.000	.008
Categorías * simetría * muestra	C2	1.275	30	.043	2.836	.000	.000
	RMSEA	2.116	30	.071	76.117	.000	.013
Factores * categorías * simetría * muestra	C2	5.111	117	.044	2.915	.000	.002
	RMSEA	4.487	117	.038	41.380	.000	.026
Error	C2	2684.584	179128	.015			
	RMSEA	165.996	179128	.001			
Total	C2	4907.381	179485				
	RMSEA	1713.805	179485				

Nota. Según Cohen (1988) un tamaño del efecto igual o superior a 0.14 es suficientemente grande como para ser tenido en cuenta, tamaños iguales o inferiores a 0.01 son pequeños y valores en torno a 0.06 son medianos.

Descriptivamente, tal y como se ha visto anteriormente, el índice RMSEA aumenta conforme crece el grado de asimetría, de tal manera que solamente son rechazados los modelos con asimetría severa. Por otro lado, a medida que aumenta el tamaño muestral RMSEA disminuye significativamente.

Si bien los efectos principales son interesantes, y al menos a nivel descriptivo se han visto en las tablas anteriores, los de interacción no son menos. En este sentido, sólo hay un efecto de interacción relevante para la combinación de los niveles de simetría en función del tamaño muestral que afecta a RMSEA ($\eta^2 = .197$).

En este sentido, cuanto menor es el tamaño muestral del modelo mayor es el valor de RMSEA. El valor de RMSEA aumenta en un nivel parecido entre los diferentes tamaños muestrales conforme aumenta el grado de asimetría, excepto para tamaños muestrales pequeños que aumenta en mayor medida que el resto cuando la asimetría es mayor. Esta relación se describe en la Figura 1. Como se puede observar, según este índice de ajuste se aceptarían los modelos con distribución de respuestas simétrica, los modelos de 450 a 850 sujetos cuya distribución tiene asimetría moderada (los modelos de 250 sujetos se encuentran próximos al punto de corte de .08) y los modelos de 450 a 850 sujetos con asimetría severa en la distribución de respuestas.

Discusión

De acuerdo con los resultados, el error Tipo I que muestra el índice C2 es siempre mayor que el de RMSEA para cualquiera de los factores experimentales analizados. En el caso concreto del tamaño muestral, se hace patente la sensibilidad del estadístico de ajuste χ^2 hacia dicho factor experimental, la cual da lugar a que cuando se trabaja con muestras grandes se tienda a rechazar los modelos teóricos propuestos (Bollen, 1989; Hu y Bentler, 1995). De hecho, para cualquiera de los tamaños muestrales analizados, con el índice χ^2 de razón de verosimilitud se rechaza aproximadamente el 80% de modelos a prueba. En cambio, con el índice RMSEA el porcentaje de rechazo de los modelos disminuye a medida que se incrementa el número de sujetos. Por tanto, de acuerdo con RMSEA, para cometer el menor error Tipo I parece recomendable que el tamaño

muestral del modelo teórico sea el mayor posible.

No obstante, debemos ser cautelosos con la conclusión anterior, pues en el cálculo del índice RMSEA intervienen los grados de libertad y el tamaño muestral. En este sentido, cuanto mayor sea el tamaño muestral y los grados de libertad del modelo teórico a prueba, menor será el valor de RMSEA y, por consiguiente, el ajuste será mayor (Kline, 2011; MacCallum, et al., 1996). Por tanto, para interpretar los resultados obtenidos mediante RMSEA habría que atender a la complejidad del modelo, indicada por los grados de libertad (Byrne, 1998).

Respecto a lo anterior, una de las limitaciones del presente estudio es el no haber tenido en cuenta la influencia de la interacción entre diferentes factores experimentales sobre el error Tipo I cometido. Así, en la Tabla 3 se observa que cuando la distribución de respuestas de las variables observadas pasa de ser simétrica a tener asimetría moderada, la probabilidad de rechazar los modelos teóricos a prueba sufre un incremento, mucho más marcado para C2. En este sentido, se constata que el test del índice χ^2 de razón de verosimilitud, dado que procede del método ML, es también sensible al hecho de que la distribución de las variables observadas se encuentre alejadas de la distribución normal multivariable (Schumacker y Lomax, 1996; Ryu, 2011) aunque también sería interesante conocer la influencia mutua entre, por ejemplo, grado de asimetría y tamaño muestral. En relación con ello, otra de las limitaciones del presente estudio relacionadas con el análisis del grado de asimetría radica en que debería haberse tenido en cuenta también la curtosis, como señalan Wright y Herrington (2011).

Otro aspecto que debe tenerse presente acerca del índice χ^2 de razón de verosimilitud es el hecho de que dicho índice se basa en la distribución central de χ^2 . Esto

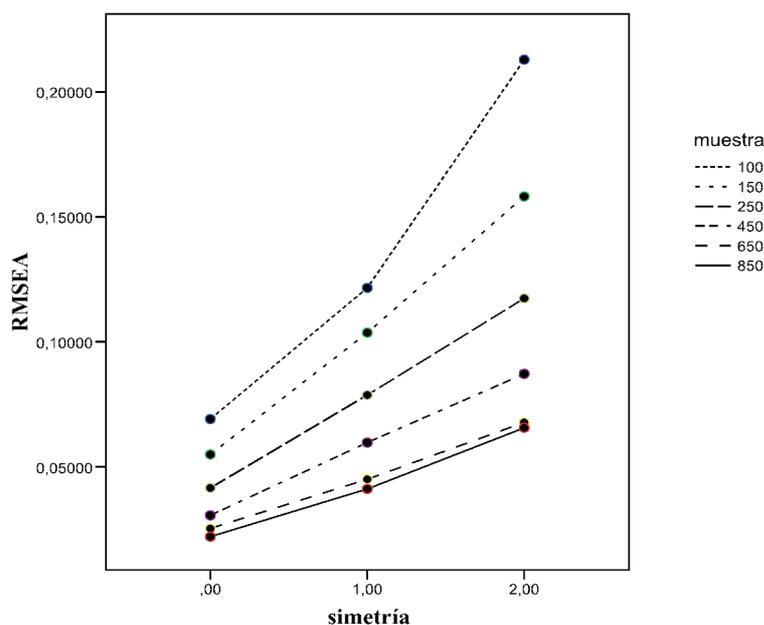


Figura 1. Efecto de la simetría por tamaño muestral sobre RMSEA

significa que se asume un ajuste perfecto del modelo teórico propuesto, es decir, que se mantiene la hipótesis nula de la prueba estadística vinculada a dicho índice de ajuste (Byrne, 1998). En la práctica, cualquier pequeño desajuste del modelo propuesto unido a un tamaño muestral elevado va a incrementar la probabilidad de que tal modelo sea rechazado (Barbero et al., 2011). En este punto cabe recordar que la potencia de la prueba χ^2 se interpreta como la probabilidad de descubrir restricciones incorrectas o parámetros que deban ser añadidos en el modelo (Batista y Coenders, 2000). En relación con ello, Fadlelmula (2011) indica que la aceptación o el rechazo de la hipótesis nula para determinar si un modelo tiene buen ajuste se interpreta de manera diferente de acuerdo con el nivel de potencia. Consecuentemente, sería interesante incorporar el estudio de la potencia en futuros trabajos de simulación donde se comparen modelos mal especificados con modelos bien especificados. En esta línea, en el estudio de Nestler (2013) se comparan entre sí diferentes métodos de estimación aplicados a modelos cuyos ítems son de respuesta dicotómica.

En cuanto al método de estimación utilizado en el presente estudio, ULS contribuye a que los resultados del índice RMSEA no estén distorsionados aunque el tamaño muestral sea elevado (Simsek y Noyan, 2012).

En resumen, cuando se utiliza ULS hay que tener en cuenta que cuando los modelos presenten muchas dimensiones latentes, las variables de respuesta tengan pocos puntos de respuesta y sean asimétricas es probable que, aunque el modelo sea correcto, sea rechazado según C2. Por otro lado, cuando los ítems presenten pocos puntos de respuesta, sean asimétricos y el tamaño muestral sea reducido, es probable que RMSEA aconseje rechazar el modelo aun cuando sea correcto. Un aspecto a destacar es que, si bien las dimensiones simuladas estaban incorrelacionadas, en ámbitos como la Psicología es frecuente que los conceptos medidos guarden una relación oblicua, esto es, estén relacionados entre sí, compartiendo algún porcentaje de variabilidad.

Por dar algunas recomendaciones prácticas que deberían ser tenidas en cuenta a la hora de interpretar RMSEA y C2 en función del modelo que se esté evaluando, podemos indicar que:

1. RMSEA es un índice de ajuste que funciona correctamente independientemente del número de factores y que mejora conforme aumentan el número de puntos de respuesta de la escala y el tamaño muestral. Sin embargo, a mayor asimetría, empeora su comportamiento.

2. C2 es relativamente independiente del número de sujetos y mejora conforme se incrementa el número

de factores y el grado de asimetría. No obstante, empeora cuando crece el número de puntos de respuesta de la escala.

3. De acuerdo con RMSEA, se aceptan modelos teóricos cuyo tamaño muestral oscila entre 450 y 850 sujetos, con independencia del grado de asimetría que presente la distribución de respuestas. Sin embargo, es necesario que dicha distribución sea simétrica para que se acepten los modelos con muestras a partir de 100 sujetos.

En cualquier caso, es necesario obtener información complementaria acerca de la potencia del estudio para determinar si el modelo teórico propuesto encaja con los datos empíricos.

Referencias

- Barbero, M. I., Vila, E. y Holgado, F. P. (2011). *Introducción básica al análisis factorial [Basic introduction to factor analysis]*. Madrid, España: UNED.
- Batista, J. M. y Coenders, G. (2000). *Modelos de ecuaciones estructurales [Structural equation models]*. Madrid, España: La Muralla.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. y Barb, K. H. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, 46(2), 232-239. doi: 10.2307/2094981
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guildford Press.
- Browne, M. W. y Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Bollen y J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, applications and programming*. Londres, UK: Lawrence Erlbaum Associates.
- Catena, A., Ramos, M. M. y Trujillo, H. M. (2003). *Análisis multivariado. Un manual para investigadores [Multivariate analysis. A handbook for researchers]*. Madrid, España: Biblioteca Nueva.

- Cea, M. A. (2004). *Análisis multivariable. Teoría y práctica en la investigación social [Multivariate analysis. Theory and practice in social research]*. Madrid, España: Síntesis.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª Ed.). New Jersey: Lawrence Erlbaum Associates.
- Coenders, G., y Saris, W. E. (1995). Categorization and measurement quality. The choice between Pearson and Polychoric correlations. En W. E. Saris y Á. Münnich (Eds.), *The Multitrait-Multimethod approach to evaluate measurement instruments* (pp. 125-144). Budapest, Hungría: Eötvös University Press.
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling*, 9(3), 327-346. doi: 10.1207/S15328007SEM0903_2
- Fadlelmula, F. K. (2011). Assessing power of structural equation modeling studies: A meta-analysis. *Educational Research Journal*, 1(3), 37-42.
- Flora, D. B. y Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466-491. doi: 10.1037/1082-989X.9.4.466
- Garrido, L. E. (2012). *Dimensionality assessment of ordinal variables: An evaluation of classic and modern methods* (Tesis doctoral). Universidad Autónoma de Madrid, Madrid.
- Finney, S. J. y DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. En G. R. Hancock y R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2ª Ed.) (pp. 439-492). Greenwich, CT: Information Age Publishing.
- Garnefski, N. y Kraaij, V. (2007). The Cognitive Emotion Regulation Questionnaire: Psychometric features and prospective relationships with depression and anxiety in adults. *European Journal of Psychological Assessment*, 23(3), 141-149. doi: 10.1027/1015-5759.23.3.141
- Holgado, F. P., Moscoso, S., Barbero, M. I. y Sanduvete, S. (2006). Training Satisfaction Rating Scale: Development of a measurement model using polychoric correlations. *European Journal of Psychological Assessment*, 22(4), 268-279. doi: 10.1027/1015-5759.22.4.268
- Holgado, F. P., Chacón, S., Barbero, I. y Vila, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44(1), 153-166. doi: 10.1007/s11135-008-9190-y
- Hu, L. y Bentler, P. M. (1995). Evaluating model fit. En R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.
- Jennrich, R. y Satorra, A. (2014). The nonsingularity of Γ in covariance structure analysis of nonnormal data. *Psychometrika*, 79(1), 51-59. doi: 10.1007/s11336-013-9353-1
- Johnson, D. R. y Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48, 398-407. doi: 10.2307/2095231
- Jöreskog, K. G. y Sörbom, D. (1996a). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Jöreskog, K. G. y Sörbom, D. (1996b). *PRELIS 2: User's reference guide*. Chicago: Scientific Software International.
- Jöreskog, K. G. (2004). *On Chi-Squares for the independence model and fit measures in LISREL*. Recuperado de <http://www.ssicentral.com/lisrel/techdocs/ftb.pdf>
- Jung, S. (2013). Structural equation modeling with small sample sizes using two-stage ridge least-squares estimation. *Behavior Research Methods*, 45(1), 75-81. doi: 10.3758/s13428-012-0206-0
- Katsikatsou, M., Moustaki, I., Yang-Wallentin, F. y Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis*, 56(12), 4243-4258. doi: 10.1016/j.csda.2012.04.010
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3ª Ed.). New York: The Guilford Press.

- Liu, X. S. (2012). Implications of statistical power for confidence intervals. *British Journal of Mathematical and Statistical Psychology*, 65(3), 427-437. doi: 10.1111/j.2044-8317.2011.02035.x
- McCallum, R. C., Browne, M. W. y Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149. doi: 10.1037/1082-989X.1.2.130
- Mulaik, S. A. (1972). *The foundations of factor analysis* (Vol. 88). New York: McGraw-Hill.
- Muthén, B. y Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189. doi: 10.1111/j.2044-8317.1985.tb00832.x
- Nestler, S. (2013). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, 66(1), 127-143. doi: 10.1111/j.2044-8317.2012.02044.x
- O'Brien, R. M. (1985). The relationship between ordinal measures and their underlying values: Why all the disagreement? *Quality and Quantity*, 19, 265-277. doi: 10.1007/BF00170998
- R Development Core Team (2010). *R: A language and environment for statistical computing 2.12.0*. Viena, Austria: R Foundation for Statistical Computing.
- Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods*, 43(4), 1066-1074. doi: 10.3758/s13428-011-0115-7
- Satorra, A., y Bentler, P. M. (2010). Ensuring positiveness of the scaled difference Chi-square test statistic. *Psychometrika*, 75(2), 243-248. doi: 10.1007/s11336-009-9135-Y
- Savalei, V. y Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, 66(2), 201-223. doi: 10.1111/j.2044-8317.2012.02049.x
- Schumacker, R. E. y Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. New Jersey: Lawrence Erlbaum Associates.
- Simsek, G. G. y Noyan, F. (2012). Structural equation modeling with ordinal variables: a large sample case study. *Quality and Quantity*, 46 (5), 1571-1581. doi:10.1007/s11135-011-9467-4
- Wright, D. B. y Herrington (2011). Problematic standard errors and confidence intervals for skewness and kurtosis. *Behavior Research Methods*, 43, 8-17. doi: 10.3758/s13428-010-0044-x
- Yang-Wallentin, F., Jöreskog, K. G. y Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, 17 (3), 392-423. doi: 10.1080/10705511.2010.489003
- Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. En C.R. Rao y S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics*, (pp. 45-79). Amsterdam: Elsevier Science.