

Revista Latinoamericana de Psicología

www.elsevier.es/rlp



ORIGINAL

Análisis factorial de ítems de respuesta forzada: una revisión y un ejemplo

Carlos Calderón Carvajal^{a,*} y Carmen Ximénez Gómez^b

^a Universidad de Magallanes, Chile

^b Universidad Autónoma de Madrid, España

Recibido el 17 de octubre de 2012; aceptado el 4 de marzo de 2014

PALABRAS CLAVE

Test de respuesta forzada;
Datos ipsativos;
Juicio comparativo;
Análisis factorial

KEYWORDS

Forced-choice Tests;
Ipsative Data;
Comparative Judgment;
Factor Analysis

Resumen

Los test de respuesta forzada son ampliamente utilizados para reducir el efecto de los diferentes tipos de sesgo en la respuesta asociados a los test psicológicos (por ejemplo, la aquiescencia o la deseabilidad social). Sin embargo, este tipo de test genera los denominados datos ipsativos, los cuales poseen propiedades que hacen desaconsejable la aplicación de las técnicas clásicas de análisis factorial para su evaluación psicométrica. Pese a ello, en la práctica, muchos investigadores siguen empleando estos procedimientos para analizar los ítems de respuesta forzada; esto lleva necesariamente a conclusiones erróneas. El presente trabajo expone las propiedades analíticas de los ítems de respuesta forzada, así como un ejemplo que ilustra cómo afectan estas propiedades a la aplicación de las técnicas estadísticas clásicas y conducen a interpretaciones erróneas. Adicionalmente, se presenta una de las principales alternativas para analizar este tipo de datos basada en el modelo de juicio comparativo de Thurstone, así como los resultados de un estudio de simulación que ilustra su aplicación y efectividad en la recuperación de la estructura factorial original.

Copyright © 2012, Fundación Universitaria Konrad Lorenz. Publicado por Elsevier España, S.L.U. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Factor analysis of forced-choice items: A review and an example

Abstract

Forced-choice tests are widely used in order to reduce the impact of different response set biases typically associated to psychological tests (e.g. acquiescence or social desirability). However, these tests produce ipsative data which have undesirable properties, thereby making an inappropriate application of classical factor analysis techniques for psychometric evaluation commonly used by researchers. This paper

*Autor para correspondencia.

Correo electrónico: carlos.calderon@umag.cl (C. Calderón Carvajal).

explains the analytical properties of forced-choice tests, along with an example that illustrates how these properties have an impact on the application of conventional statistical techniques and produce improper results. Additionally, one of the current proposals is presented in order to analyze these data based on the comparative judgment model by Thurstone, along with the results of a simulation study which illustrates its implementation and effectiveness in recovering the original factor structure.

Copyright © 2012, Konrad Lorenz University Foundation. Published by Elsevier España, S.L.U. This is an open-access article distributed under the terms of the Creative Commons CC BY-NC ND Licence (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

En ciencias sociales suelen emplearse escalas de medida que proporcionan un tipo de datos denominados normativos. Por ejemplo, los formatos de respuesta tipo Likert, donde los ítems se evalúan en una escala de acuerdo-desacuerdo con diferentes puntos. Desde un enfoque analítico, los datos normativos se caracterizan porque la puntuación total en el test difiere de un individuo a otro, debido a que cada observación se obtiene de forma independiente. La independencia de las observaciones es una característica deseable desde el punto de vista estadístico, ya que es condición necesaria para el adecuado funcionamiento de las técnicas de análisis de datos más utilizadas por los investigadores (el análisis de la varianza, la regresión lineal o el análisis factorial). A pesar de esta ventaja analítica, los datos normativos presentan una importante limitación práctica. Existe un sinfín de contextos en los cuales se produce el falseamiento de respuesta (*faking*, en inglés) por parte de los individuos, debido a que muchos de los instrumentos se ven influidos por la deseabilidad social dadas las características del atributo o de la situación de evaluación (Stark, Chernyshenko, Chan, Lee y Drasgow, 2001).

Uno de los esfuerzos más prometedores para solventar el falseamiento de respuesta es mediante los ítems de respuesta forzada o medidas ipsativas (Chan y Bentler, 1993). En psicología existen test que presentan alternativas de respuesta con grados de deseabilidad similar. Uno de los más conocidos es el Inventario de Estilos de Aprendizaje de Kolb (LSI) (1985). Este test se utiliza en contextos educativos; los individuos deben ordenar de forma jerárquica las 4 alternativas de respuesta presentadas en cada ítem. Este test consta de 12 ítems de tipo autoinforme que deben ordenarse en 4 opciones de respuesta (4, 3, 2 y 1), desde *la que más me representa* (4) a *la que menos me representa* (1), respondiendo a la pregunta “¿Cómo aprendo mejor?”. Cada alternativa de respuesta corresponde a uno de los 4 estilos de aprendizaje: acomodador, divergente, asimilador y convergente. Como resultado, la puntuación total de cada ítem ($1 + 2 + 3 + 4 = 10$) y la puntuación total del test ($10 \times 12 = 120$) es la misma para todos los individuos.

El modo de puntuar ipsativamente a un individuo en distintas dimensiones provoca la interdependencia de las respuestas entre dichas dimensiones. Debido al formato de respuesta, una puntuación muy alta en una dimensión conlleva necesariamente a puntuaciones bajas en las restantes dimensiones. Este problema se incrementa en la medida que disminuye el número de dimensiones que mide el test (Cattell, 1944). Para el caso de que el test evalúe solo 2 dimensiones, la correlación entre ambas es necesariamente -1.

Una complicación asociada es que los datos ipsativos carecen de matrices de covarianzas interpretables debido a que generan matrices singulares y con términos negativos (Aitchinson, 1986), lo que compromete la posibilidad de estudiar la estructura de covarianzas subyacente a los datos, al igual que su fiabilidad y validez. Algunos estudios muestran que no se deben aplicar los métodos de factorización estándar, en particular el análisis de componentes principales (ACP), en datos ipsativos ya que las correlaciones tienden a producir factores bipolares artificiales, acarreado de este modo la obtención de estructuras factoriales erróneas (Closs, 1996; Dunlap y Cronwell, 1994).

El presente artículo tiene dos propósitos fundamentales. El primero es mostrar los errores que se cometen al aplicar técnicas convencionales de análisis factorial a este tipo de datos, y el segundo, presentar la principal alternativa para llevar a cabo un análisis factorial de ítems con formato de respuesta forzada. El trabajo se estructura en 4 secciones. La primera parte presenta las características de los ítems con formato de respuesta forzada y cómo generan datos con propiedades ipsativas, así como las propiedades analíticas de los datos ipsativos que hacen desaconsejable la aplicación de técnicas estándares de factorización. En la segunda sección se presenta un ejemplo con datos simulados que permite ilustrar cómo los datos ipsativos generan distorsiones en las soluciones factoriales que pueden llevar a conclusiones erróneas sobre la estructura interna de los datos. La tercera sección muestra la formulación analítica de uno de los principales esfuerzos para resolver el problema del análisis factorial de datos ipsativos producto del formato de respuesta forzada, así como un estudio de simulación que ilustra la aplicación del análisis factorial con datos de respuesta forzada. Por último, se discuten las conclusiones e implicaciones prácticas del trabajo.

El formato de respuesta forzada

A diferencia del formato Likert, en el que se responde a un enunciado seleccionando una alternativa de un conjunto de categorías ordenadas, en el formato de respuesta forzada el individuo debe ordenar distintos enunciados de acuerdo con cómo describen su comportamiento. La figura 1 muestra un ejemplo referido a un bloque de 4 alternativas de respuesta que se corresponden con las 4 dimensiones medidas por el test. El individuo debe ordenar las alternativas desde *la que más le representa* (4) hasta *la que menos le representa* (1). Cada bloque (p) está compuesto por $n = 4$

Yo aprendo mejor...			
<input type="checkbox"/>	...sintiendo	<input type="checkbox"/>	...observando
<input type="checkbox"/>	...pensando	<input type="checkbox"/>	...haciendo

Figura 1. Ejemplo de ítem de respuesta forzada.

ítems que miden $d = 4$ dimensiones. Por ejemplo, con $p = 12$ bloques de $n = 4$ alternativas por bloque, el test tendría $m = 48$ ítems en total ($n \times p = 48$).

En lugar de evaluar cada enunciado por separado, el individuo debe responder a cada una de las alternativas en relación con las demás alternativas del bloque, lo cual produce una interrelación de las puntuaciones entre los ítems que pertenecen al mismo bloque. Puesto que no es posible asignar a las alternativas del mismo bloque la misma puntuación, el formato de respuesta forzada contribuye a eliminar el sesgo de respuesta producto de la aquiescencia (Cheung y Chan, 2002) y a reducir el efecto halo (Bartram, 2007), ya que se debe priorizar una alternativa sobre las restantes. Sin embargo, los datos resultantes tienen propiedades ipsativas.

Datos ipsativos

El término ipsativo proviene del latín *ipse*: “sí mismo” y fue usado por primera vez por Cattell (1944) para designar uno de los tres tipos básicos de escalas en medición psicológica: a) las escalas interactivas, donde el atributo se mide desde otros atributos y otros individuos; b) las escalas normativas, donde el atributo de un individuo se mide en relación con otros individuos, y c) las escalas ipsativas, donde el atributo de un individuo se mide en relación con sus propias puntuaciones en otros atributos, lo que hace que la suma de las puntuaciones de las variables para cada individuo sea una constante (Chan y Bentler, 1993). Expresado formalmente (Clemans, 1966), sea $\mathbf{x} = [x_1, x_2, \dots, x_p]'$ un vector columna de orden $p \times 1$ tal que

$$\sum_{i=1}^p x_i = \mathbf{1}' \mathbf{x} = c, \quad (1)$$

donde $\mathbf{1}$ es un vector unitario de orden $p \times 1$ y c es una constante, entonces \mathbf{x} es un vector p - dimensional de datos con propiedades ipsativas.

El problema de la factorización de datos ipsativos

Debido a la restricción de suma constante, los datos ipsativos poseen algunas propiedades estadísticas no deseables que no se encuentran en un conjunto típico de datos normativos (Chan y Bentler, 1993). Por ejemplo, carecen de una matriz de covarianzas interpretable (Aitchinson, 1986). Dado que $x_1 + x_2 + \dots + x_p = c$, para todo $i = 1, 2, \dots, p$ se cumple:

$$\begin{aligned} cov(x_i, x_1 + x_2 + \dots + x_p) &= cov(x_i, c) = 0, \\ cov(x_i, x_1) + cov(x_i, x_2) + \dots + cov(x_i, x_i) + \dots + \\ cov(x_i, x_{p-1}) + cov(x_i, x_p) &= 0. \end{aligned} \quad (2)$$

De la ecuación (2) se deduce que al menos una de las $p - 1$ covarianzas es negativa, con independencia de si las variables correlacionan positiva o negativamente, dado que $var(x_i)$ es siempre positiva, y que esto ocurre en todas las filas y columnas de la matriz de covarianzas de \mathbf{x} , denominada Σ_x . Por tanto:

$$\Sigma_x \mathbf{1} = \mathbf{0}, \text{ o bien } \mathbf{1}' \Sigma_x = \mathbf{0}', \quad (3)$$

Según esto, toda estimación de Σ_x (es decir, S_x) basada en una muestra de n observaciones que satisface la ecuación (1) también estará sujeta a la restricción de suma 0 de la ecuación (3). De esta ecuación se deduce que S_x y Σ_x no son de rango completo. Esta característica es de elevada relevancia, ya que para este tipo de datos no es posible la aplicación de las técnicas multivariantes clásicas desarrolladas para vectores sin este tipo de restricción, o al menos no directamente, ya que puede llevar a interpretaciones erróneas (Aitchinson, 1986). Más específicamente, los procedimientos de estimación no pueden llevar a cabo los análisis si la matriz de covarianzas es singular. Por ejemplo, para el caso de la estimación de mínimos cuadrados generalizados, la función $F_{GLS} = 1 / 2tr[(\mathbf{S} - \Sigma)\mathbf{S}^{-1}]$ requiere que \mathbf{S} sea regular, mientras que para máxima verosimilitud, el cálculo del estadístico de ajuste χ^2 depende del determinante de \mathbf{S} (para un resumen más detallado sobre el análisis factorial de datos ipsativos, véase Ximénez y Calderón, 2012).

A pesar de estas limitaciones, son numerosos los estudios que han continuado usando los procedimientos convencionales para el estudio de la estructura interna de esta clase de test, especialmente a través del ACP, el cual, a diferencia de los métodos de estimación del análisis factorial clásico, no requiere la no-singularidad de la matriz de entrada. Es abundante la literatura que desaconseja la utilización del ACP, ya que lleva a la obtención de estructuras erróneas (Baron, 1996; Closs, 1996; Cornwell y Dunlap, 1994; Dunlap y Cornwell, 1994; Loo, 1999). Esto se debe a que el análisis de matrices de correlaciones de datos ipsativos tiende a producir factores bipolares artificiales, fallando en el reflejo de la estructura factorial poblacional verdadera debido a la presencia de términos negativos en las matrices de entrada y a que la suma de las correlaciones entre los factores es siempre 0 (Clemans, 1966).

Problemas de fiabilidad de los test de respuesta forzada

El tema de la fiabilidad de los test de respuesta forzada ha sido causa de un fuerte debate entre algunos investigadores (Baron, 1996; Cornwell y Dunlap, 1994; Hicks, 1970; Johnson, Wood y Blinkhorn, 1988; Meade, 2004; Saville y Willson, 1991; Tenopyr, 1988). En la actualidad se ha demostrado que la forma tradicional de puntuar los test con ítems de respuesta forzada genera importantes distorsiones en las estimaciones de la consistencia interna. Estas se ven incre-

mentadas en la medida en que disminuye el número de factores, lo cual se torna problemático, ya que precisamente uno de los propósitos de los investigadores es la búsqueda de la parsimonia de los modelos factoriales.

El principal problema de los ítems con formato de respuesta forzada es que violan una de las asunciones básicas de la Teoría Clásica de los Test: la independencia de los errores de medida. Esto se debe a que la respuesta de un individuo a una de las alternativas de un bloque depende de las respuestas que dé a las restantes alternativas del mismo bloque (Meade, 2004). Esta violación es crítica porque hace inapropiada la utilización del estadístico alfa de Cronbach y de otros tipos de medidas de fiabilidad tales como el test-retest.

En síntesis, no se recomienda el uso de los estadísticos de fiabilidad en el marco de la Teoría Clásica de los Test ya que pueden conducir a resultados dudosos y conclusiones erróneas. Este problema cobra mayor complejidad debido a que cada test de respuesta forzada implica constructos distintos y está diseñado en un formato específico, lo cual hace imposible establecer generalizaciones acerca de los posibles problemas de consistencia interna (Tenopir, 1988).

El modelo Thurstoniano

El modelo de Thurstone (1931) proporciona un marco conceptual de gran *utilidad* para comprender la naturaleza de los datos ipsativos de respuesta forzada. Según la ley del juicio comparativo de Thurstone, el individuo realiza un proceso de comparación frente a la tarea de ordenar distintos objetos. El modelo propone 3 supuestos básicos: a) cada objeto que se compara (ítem) provoca una variable latente (*utilidad*) como resultado del proceso de discriminación; b) al comparar 2 objetos, el individuo elige aquel con mayor nivel de *utilidad*, y c) la *utilidad* corresponde a una variable latente con distribución normal.

Sea t_i la *utilidad* latente asociada al ítem i , el individuo preferirá el ítem i sobre otro ítem k si la *utilidad* latente del primero es mayor a la del segundo. Según esto, la preferencia se puede codificar como:

$$y_l = \begin{cases} 1 & \text{si } t_i \geq t_k, \\ 0 & \text{si } t_i < t_k. \end{cases} \quad (4)$$

Este modelo puede extenderse fácilmente a la tarea de ordenamiento de más de 2 objetos. Frente a p alternativas de respuesta, el sujeto debe ordenar los ítems desde p (el que más le representa) hasta 1 (el que menos le representa). El modelo Thurstoniano asume la existencia de un vector aleatorio $Y = [Y_1, Y_2, \dots, Y_p]'$ de orden $p \times 1$ que posee una

distribución normal multivariante. De este modo, el individuo ordena los objetos según su nivel en la variable Y_i .

Este modelo permite asumir la existencia de una variable con propiedades normativas (*utilidades*) que subyace a cada objeto (ítem) frente a la tarea de ordenamiento. Por tanto, cada individuo utilizará el nivel de *utilidad* que subyace a cada objeto evaluado para realizar la tarea de ordenamiento en un bloque de respuesta forzada. De este modo, se asume un modelo factorial convencional de variables con propiedades normativas, relacionado con los datos ipsativos producto del formato de respuesta forzada.

Un ejemplo ilustrativo

Propondremos un ejemplo simple que proporcione un modelo sencillo que nos permita ilustrar cómo los datos de un cuestionario con ítems de respuesta forzada pueden afectar al estudio de la estructura interna de un conjunto de datos. Para ello, nos basaremos en el diseño propuesto en el LSI, de Kolb (1985), que es uno de los test de respuesta forzada más ampliamente utilizado tanto en contextos de investigación como aplicados.

Para la generación de los datos, utilizaremos un modelo de 3 bloques ($p = 3$), de 4 ítems (alternativas) por bloque ($n = 4$); los cuales miden 4 dimensiones independientes ($d = 4$). Cada bloque considera 4 variables continuas con distribución normal multivariante (*utilidades*) que están linealmente relacionadas con cada una de las 4 dimensiones. Cada puntuación normativa se asocia a una puntuación ipsativa, la cual corresponde al orden decreciente del valor de la *utilidad* con respecto a cada bloque. La muestra simulada se compone de 1.000 casos, con una estructura de 4 factores. El diseño del test se expone en la tabla 1. Cada factor se definió mediante 3 variables con distribución normal ($\mu = 0$ y $\Sigma = \Lambda\Phi\Lambda' + \Theta_g$). Las saturaciones factoriales se fijaron a ,707 y los errores de medida a ,50 ($\theta_g = 1 - \lambda^2$) para obtener una solución estandarizada. Cada una de las variables correspondientes a cada factor fue ubicada en distintos grupos para formar los 3 bloques de 4 ítems. Cada ítem del bloque se vinculó a 1 de las 4 dimensiones del modelo. Finalmente, a cada variable se le asignó el valor correspondiente al orden decreciente en cada uno de los bloques de pertenencia.

Como ha sido habitual en estudios anteriores que analizan esta clase de datos, hemos realizado un ACP tanto para los datos ipsativos como para los normativos, lo cual permitirá comparar ambas soluciones con el propósito de determinar el funcionamiento del ACP en la recuperación de la estructura interna de los datos ipsativos de respuesta forzada. Al igual que con los estudios del LSI de Kolb (1985), hemos obtenido las soluciones de 4 y 2 factores, que se rotaron con el método varimax. Los resultados se muestran en la tabla 2.

Tabla 1 Asignación de los ítems a cada uno de los bloques y factores del cuestionario

	Factor 1	Factor 2	Factor 3	Factor 4
Bloque 1	ítem 1	ítem 2	ítem 3	ítem 4
Bloque 2	ítem 5	ítem 6	ítem 7	ítem 8
Bloque 3	ítem 9	ítem 10	ítem 11	ítem 12

Tabla 2 Solución factorial con análisis de componentes principales sobre datos normativos e ipsativos de respuesta forzada

Valores poblacionales		Datos normativos						Datos ipsativos de respuesta forzada							
		4 factores				2 factores		<i>r</i> ítem/ test	4 factores				2 factores		<i>r</i> ítem/ test
λ_j	ψ_j	F1	F2	F3	F4	F1	F2		F1	F2	F3	F4	F1	F2	
$y_{1,1}$,707	,50	,827				,646	,572		,737			,350		,133
$y_{5,1}$,707	,50	,840				,674	,582	-,492	-,563	,542		-,806	,333	-,220
$y_{9,1}$,707	,50	,811				,668	,574		,648		-,364	,301		,216
$y_{2,2}$,707	,50		,829				,581	,788					-,768	,238
$y_{6,2}$,707	,50		,816				,590		,800			,352		-,133
$y_{10,2}$,707	,50		,825				,580	,717				,413	,338	,220
$y_{3,3}$,707	,50			,816		,303	,498			-,789			-,768	,219
$y_{7,3}$,707	,50			,823		,329	,512	,587	,789			,505	,316	-,182
$y_{11,3}$,707	,50			,816		,346	,467	,597		-,643	-,325		-,759	,148
$y_{4,4}$,707	,50				,814	,563		,602	-,522	-,533	,525	-,829	,324	,135
$y_{8,4}$,707	,50				,820	,588		,622		-,749			-,790	-,196
$y_{12,4}$,707	,50				,813	,547		,579			,905	-,611		,229
% var. exp.			17,11	16,99	16,82	16,66	18,26	17,12		21,01	19,97	19,77	9,75	22,69	20,32

*Se calcularon las correlaciones ítem-test para cada una de las dimensiones.

**Se suprimieron las saturaciones inferiores a ,30.

Ordene los 4 enunciados desde el 4 ("el que más me representa") hasta el 1 ("el que menos me representa")

Ítem A Ítem B Ítem C Ítem D

Figura 2. Formato de un bloque de respuesta forzada de 4 ítems ($n = 4$).

A partir de los resultados obtenidos, podemos realizar 2 observaciones relevantes. En primer lugar, y como era de esperar, el ACP sobre los datos normativos logra recuperar satisfactoriamente la estructura de los datos, aunque las saturaciones se ven sobreestimadas, lo cual es una desventaja bien conocida del ACP. Para el caso de la solución de 4 factores, podemos ver claramente la presencia de la estructura factorial verdadera, en donde todos los ítems alcanzan saturaciones superiores a ,80. En cuanto a la solución de 2 factores, podemos ver un claro patrón en el cual los ítems de 2 de las dimensiones se agrupan en el primer factor y los restantes ítems se agrupan en el segundo factor. En cuanto a la fiabilidad, las correlaciones ítems/test se encuentran en torno a ,50, lo cual es coherente con el error de medida introducido en el modelo teórico. En segundo lugar, podemos ver que el ACP no logra recuperar la estructura factorial original para los datos de respuesta forzada. Se observa tanto para la solución de 4 como para la de 2 factores que no solo es inadecuada la agrupación de variables, sino que existen ítems que comparten valores similares en más de un factor. Por otro lado, existe un grupo considerable de ítems que obtienen saturaciones factoriales negativas, lo cual sugiere la presencia de factores bipolares subyacentes, lo que no se corresponde con la estructura factorial original. Finalmente, ambas soluciones factoriales obtienen porcentajes de varianza explicada considerables (70 y 43%), lo cual sugiere la presencia de una estructura factorial correcta. De este modo, el ACP realizado sobre datos ipsativos producto de un formato de respuesta forzada arroja estructuras factoriales erróneas; esto a su vez implica inferencias equívocas debido a la aparición de factores bipolares artificiales producidos por el tipo de puntuación del formato de respuesta forzada.

Análisis factorial de ítems de respuesta forzada

Uno de los esfuerzos más prometedores por resolver los problemas analíticos en el análisis de datos de los test de respuesta forzada es el de Maydeu-Olivares y Böckenholt (2005). Los autores argumentan que el problema no radica en la utilización del formato de respuesta forzada, sino en la codificación de los datos obtenidos a través de este tipo de formato. La propuesta se basa en un procedimiento de codificación binaria de datos de respuesta forzada. En un bloque de respuesta forzada, el sujeto debe ordenar los n ítems de acuerdo con sus preferencias. Siguiendo con el ejemplo de un bloque de $n = 4$ ítems (A, B, C y D) expuesto anteriormente, el sujeto debe asignar la posición de cada ítem en relación con sus preferencias, según se muestra en la figura 2.

Respuesta forzada				Codificación binaria					
A	B	C	D	(A,B)	(A,C)	(A,D)	(B,C)	(B,D)	(C,D)
3	1	2	4	1	1	0	0	0	0

Figura 3. Formato de codificación de un bloque de respuesta forzada.

En cualquier bloque de n ítems es posible codificar las respuestas a través de una codificación binaria usando $\tilde{n} = \tilde{n}(n-1)/2$ pares de comparaciones. En nuestro ejemplo, en un bloque de $n = 4$ podemos realizar $\tilde{n} = 4(4-1)/2 = 6$ comparaciones binarias entre los 4 ítems del bloque. En todos los pares de comparaciones podemos utilizar el siguiente formato de codificación:

$$y_l = \begin{cases} 1 & \text{si se prefiere el ítem } i \text{ sobre el ítem } k, \\ 0 & \text{si se prefiere el ítem } k \text{ sobre el ítem } i. \end{cases} \quad (5)$$

Por tanto, en nuestro ejemplo la codificación del bloque de $n = 4$ quedaría de la forma presentada en la figura 3.

Ya hemos visto que según la ley del juicio comparativo de Thurstone, cada comparación por pares genera una utilidad para cada objeto que se compara según la ecuación (4). A partir de esta ecuación el proceso de codificación puede describirse de manera alternativa:

$$y_l^* = t_i - t_k, \quad (6)$$

donde y_l^* corresponde a una variable continua que representa la diferencia entre las *utilidades* de los ítems presentes en cada comparación. Por tanto, la vinculación entre la respuesta de comparación por pares y_l y la respuesta de comparación latente y_l^* es:

$$y_l = \begin{cases} 1 & \text{si } y_l^* \geq 0, \\ 0 & \text{si } y_l^* < 0. \end{cases} \quad (7)$$

Expresado en términos matriciales, sea t un vector $n \times 1$ de *utilidades* latentes, e y^* un vector $\tilde{n} \times 1$ de diferencias latentes, donde $\tilde{n} = n(n-1)/2$ corresponde a la cantidad de comparaciones por pares, las \tilde{n} ecuaciones quedan expresadas como:

$$y^* = At, \quad (8)$$

donde A es una matriz de diseño que vincula las diferencias latentes con las *utilidades* latentes. Cada fila de A corresponde a cada uno de los \tilde{n} pares de comparaciones y cada columna a cada uno de los n ítems. Para el caso de un bloque de $n = 4$ ítems:

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

La matriz de diseño especifica cuáles son los ítems que participan en cada comparación. En el ejemplo, cada columna corresponde a cada uno de los 4 ítems (A, B, C y D) y cada fila a cada una de las 6 posibles comparaciones de los 4 ítems del bloque. La primera fila compara los ítems {A, B}, la segunda los ítems {A, C}, la tercera los ítems {A, D}, la cuarta los ítems {B, C}, la quinta los ítems {B, D}, y la sexta los ítems {C, D}.

Para el caso de cuestionario de múltiples bloques, tenemos p número de bloques, n número de ítems por bloque y $n \times p = m$ número de ítems en total. Por tanto, la matriz de diseño tendrá m columnas correspondientes al total de ítems del cuestionario y $p \times \tilde{n}$ filas correspondientes a los \tilde{n} pares de comparaciones realizadas en todos los p bloques. Para el ejemplo presentado anteriormente, el cuestionario de $p = 3$ bloques, de $n = 4$ ítems por bloque, $m = 12$ ítems en total, la matriz de diseño quedaría así:

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Por otro lado, siguiendo el modelo de AFC clásico, que se expresa mediante (Jöreskog y Sörbom, 1981):

$$y = \mu_y + \Lambda \eta + \varepsilon, \quad (9)$$

Nuestro modelo de AFC, que mide d factores comunes a través de las *utilidades* latentes t , quedaría expresado como:

$$t = \mu_t + \Lambda \eta + \varepsilon, \quad (10)$$

Donde μ_t es el vector que contiene las m medias de las *utilidades* latentes t , Λ es la matriz $m \times d$ de coeficientes factoriales, η es el vector de factores comunes y ε es un vector m - dimensional de factores únicos (errores de medida). Como en un AFC clásico, se asume que: a) cada ítem, y por

tanto cada *utilidad*, mide solo un factor; b) los factores no están correlacionados con las unicidades; c) la media de los factores comunes es 0; d) la correlación entre los factores (matriz Φ) son parámetros libres que se han de estimar, y sus varianzas se fijan a 1 para la identificación del modelo; e) la matriz ψ de unicidades es diagonal, lo cual indica que las unicidades no están correlacionadas, y finalmente, f) tanto factores comunes como unicidades presentan una distribución normal.

La figura 4 ilustra cómo se modelan los resultados binarios, las *utilidades* y los factores latentes. El modelo corresponde a una estructura sencilla de $d = 4$ factores correlacionados, los cuales se definen mediante 3 ítems por factor. Los $m = 12$ ítems del cuestionario se presentan en bloques de $n = 4$ ítems. Nótese que no hay más de un ítem por bloque que mida el mismo factor latente. El cuestionario está formado por $p = 3$ bloques. El primer factor está medido por los ítems 1, 5 y 9, el segundo por los ítems 2, 6 y 10, el tercero por los ítems 3, 7 y 11, y el cuarto por los ítems 4, 8 y 12. La tarea consiste en ordenar los 4 ítems de cada bloque de acuerdo con su preferencia (desde 4, el que más le representa; a 1, el que menos le representa).

Los resultados del ordenamiento se transforman en 6 resultados binarios por bloque, 18 en total, los cuales se modelan como diferencias de pares de *utilidades* subyacentes a través de la ecuación (8). Como se observa, cada resultado binario ($y_{i,k}$) es resultado de 2 ítems y depende de las 2 *utilidades* latentes de cada ítem. A su vez, cada *utilidad* latente es función de un factor latente. Nótese que los 18 resultados binarios están medidos sin error; sin embargo, la varianza de las 12 *utilidades* latentes son parámetros libres que se han de estimar, las cuales corresponden a las unicidades de cada *utilidad*. De este modo, estamos frente a un modelo factorial de segundo orden que se define mediante 18 variables observadas que corresponden a los resultados binarios medidos sin error, 12 factores de primer orden que corresponden a las *utilidades* latentes asociadas a un error de predicción (unicidad), y 4 factores de segundo orden que equivaldrían a los factores latentes de primer orden en un AFC convencional.

Para la identificación del modelo, hemos seguido las recomendaciones sugeridas por Maydeu-Olivares y Böckenholt (2005) para ajustar datos provenientes de formatos de respuesta forzada. Para el caso de modelos multidimensionales, basta con 2 tipos de restricciones: a) fijar la varianza de los factores latentes en 1, como en un modelo factorial convencional, y b) puesto que las *utilidades* son variables latentes, hay que fijar la métrica de sus errores; para ello, es necesario fijar el error de una *utilidad* a 1 en cada bloque. En este caso, el error de una de las *utilidades* por cada bloque se puede fijar en ,50, ya que la *utilidad* latente depende de 2 resultados binarios, para que cumpla $\psi^2_i = \psi^2_{i_1} + \psi^2_{i_2}$.

Un estudio de simulación

Para ilustrar el funcionamiento del modelo de Maydeu-Olivares y Böckenholt (2005), presentaremos los resultados de un pequeño estudio de simulación que evalúa la recuperación de los parámetros y las pruebas de ajuste global. Hemos utilizado la misma estructura del ejemplo anterior; es decir, un modelo

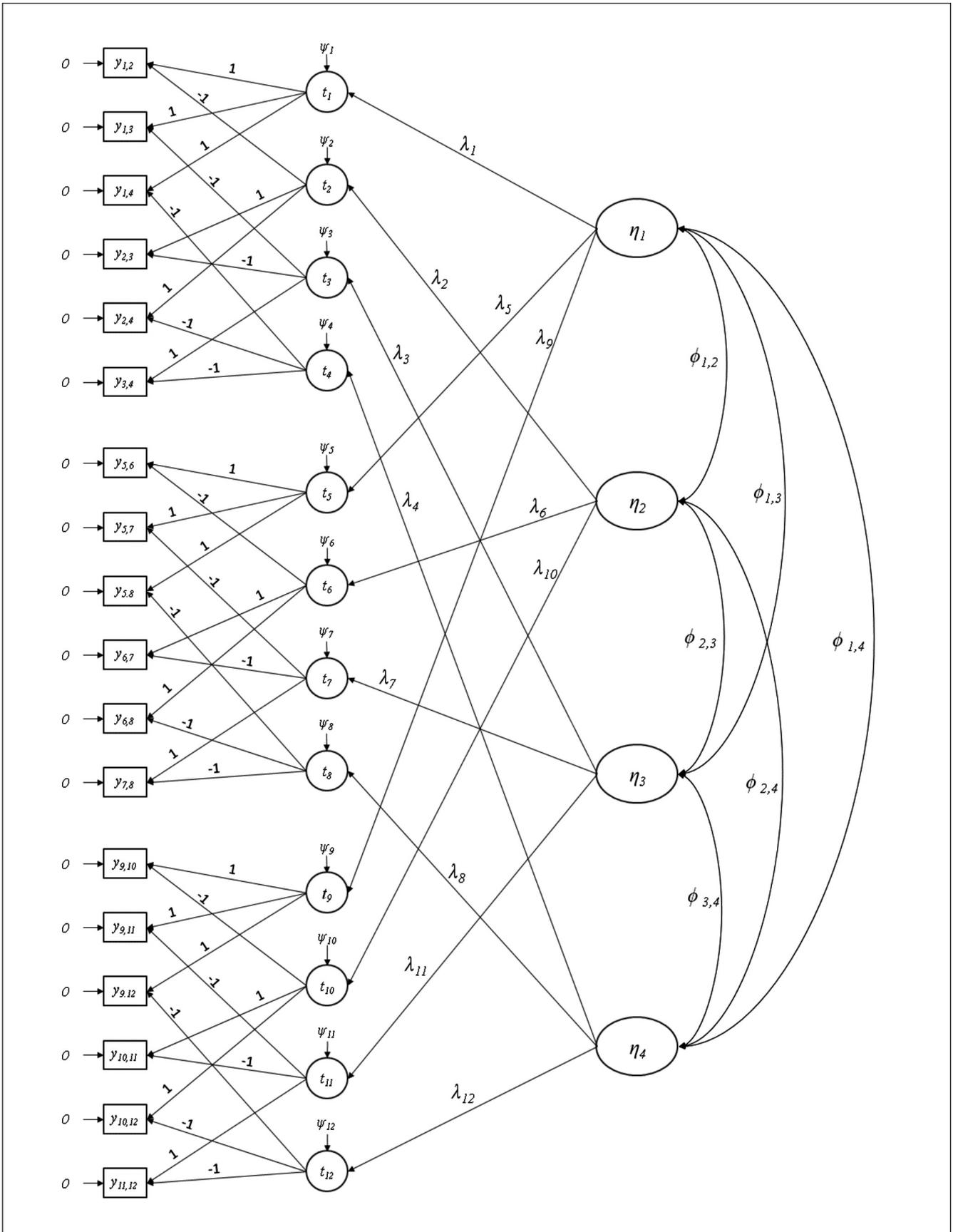


Figura 4. Modelo factorial Thurstoniano para un cuestionario de respuesta forzada de 3 factores con 3 bloques de 4 ítems por bloque. (Basado en Brown y Maydeu-Olivares, 2011, p. 467).

de 4 factores no correlacionados medidos por 3 *utilidades* latentes. Se generaron 200 muestras aleatorias de 1.000 casos cada una, con una distribución normal multivariante, con media μ y matriz de covarianzas Σ . Posteriormente, las variables normativas se transformaron en sus correspondientes datos ipsativos de respuesta forzada. El modelo de respuesta forzada, como en el ejemplo anterior, quedó constituido por $p = 3$ bloques de $n = 4$ ítems por bloque, los cuales miden $d = 4$ dimensiones con un total de $m = 12$ ítems. Las puntuaciones fueron codificadas como datos binarios. Para el análisis de los datos, hemos utilizado el método de estimación de mínimos cuadrados ponderados robustos disponible en el programa Mplus v4 de Múthen y Múthen (1998, 2010). Finalmente, y con el propósito de identificar el modelo, hemos fijado la varianza de los factores a 1 y la varianza de una de las *utilidades* en cada bloque a ,50. Dada la complejidad del modelo y la cantidad de réplicas realizadas, hemos restringido las correlaciones entre los factores a 0. Los resultados de la estimación de los parámetros aparecen en la tabla 3.

Como se observa, las medias tanto de las saturaciones como de los errores de medida alcanzan valores cercanos a los valores teóricos, lo que indica que el modelo logra recu-

perar los parámetros de manera satisfactoria, aunque con cierto sesgo. Esto contrasta con la desviación típica de los parámetros y con la media de los errores de estimación, los cuales alcanzan valores elevados con respecto al modelo normativo. Esto sugiere que, a pesar de que el modelo logra estimar los parámetros adecuadamente, la precisión de las estimaciones se ve afectada. Esto es coherente con otros estudios (Maydeu-Olivares y Brown, 2010), incluso con distintos tipos de datos ipsativos (Chan y Bentler, 1993; Cheung, 2004; Ximénez y Calderón, 2012). Este problema se debe esencialmente a la pérdida de información que presentan los datos ipsativos en comparación con las puntuaciones normativas, dada la propiedad de comparación intraindividual de los primeros (Chan, 2003).

En relación con la evaluación del ajuste, la tabla 3 muestra los estadísticos descriptivos para el índice de ajuste χ^2 , así como la proporción empírica de rechazo de la hipótesis de ajuste para ambos tipos de datos (cuando el número de ítems por bloque es > 2 es necesario realizar una corrección en los grados de libertad para el contraste del ajuste del modelo, debido a que en cada bloque existen $r = n(n-1)(n-2)/6$ redundancias en la matriz de correlaciones tetracóricas esti-

Tabla 3 Valores teóricos y estadísticos descriptivos de los parámetros y de los errores de estimación para los datos normativos y datos binarios de respuesta forzada

	ML datos normativos				WLSM datos binarios de respuesta forzada		
	Parámetros reales*	Estimación		SE Media	Estimación		SE Media
		Media	SD		Media	SD	
$\gamma_{1,1}$,707	,707	,046	,047	,720	,099	,088
$\gamma_{2,2}$,707	,704	,045	,047	,713	,083	,087
$\gamma_{3,3}$,707	,709	,045	,047	,723	,099	,089
$\gamma_{4,4}$,707	,704	,046	,047	,730	,114	,108
$\gamma_{5,1}$,707	,708	,046	,047	,713	,084	,087
$\gamma_{6,2}$,707	,708	,046	,047	,709	,090	,087
$\gamma_{7,3}$,707	,704	,047	,047	,717	,090	,087
$\gamma_{8,4}$,707	,712	,051	,047	,716	,109	,105
$\gamma_{9,1}$,707	,706	,049	,047	,730	,092	,091
$\gamma_{10,2}$,707	,708	,050	,047	,719	,086	,090
$\gamma_{11,3}$,707	,708	,048	,047	,732	,093	,091
$\gamma_{12,4}$,707	,705	,050	,047	,733	,116	,110
$\psi_{1,1}$,50	,495	,052	,050	,523	,144	,131
$\psi_{2,2}$,50	,501	,051	,050	,508	,147	,129
$\psi_{3,3}$,50	,495	,050	,050	,527	,141	,132
$\psi_{4,4}$,50	,501	,049	,050	,50*	---	---
$\psi_{5,5}$,50	,499	,052	,050	,508	,139	,125
$\psi_{6,6}$,50	,496	,049	,050	,519	,139	,127
$\psi_{7,7}$,50	,499	,051	,049	,516	,132	,127
$\psi_{8,8}$,50	,494	,052	,050	,50*	---	---
$\psi_{9,9}$,50	,500	,052	,050	,522	,139	,133
$\psi_{10,10}$,50	,497	,051	,050	,533	,134	,134
$\psi_{11,11}$,50	,495	,049	,050	,517	,137	,132
$\psi_{12,12}$,50	,497	,048	,050	,50*		
χ^2	Media	SD	PER	Media	SD	PER	
	48,111	9,490	,05	132,533	31,867	,20	

PER: proporción empírica de rechazo de H_0 ; SD: desviación típica; SE: error típico.

* Las unicidades han sido fijadas a ,50 para la identificación.

madras de las variables binarias. Es así que en un cuestionario de p bloques es necesario restar $p \times r$ a los grados de libertad del modelo). Los resultados indican que el estadístico χ^2 da cuenta de un pobre ajuste del modelo de respuesta forzada en comparación con el ajuste de los datos normativos. Del mismo modo, la proporción de rechazo de H_0 ($,20$) se encuentra por encima del valor esperado ($\alpha = ,05$), lo cual proporcionalmente a una tasa de error tipo I bastante elevada. Esto nos lleva a concluir que el análisis del ajuste de esta clase de modelos debe ser llevado a cabo con cautela, debido al inadecuado funcionamiento del estadístico χ^2 .

El método que proponen Brown y Maydeu-Olivares (2011) presenta un avance sustantivo que intenta solucionar los problemas en el ajuste de este tipo de modelos. Dicho método constituye una aproximación desde la Teoría de Respuesta al Ítem que supera muchos de los problemas técnicos de los trabajos anteriores. Este procedimiento considera la reparametrización del modelo de segundo orden en uno de primer orden matemáticamente equivalente. Ello supone un ahorro importante de recursos informáticos debido a la sencillez de la reparametrización del modelo y la obtención de las medidas de bondad de ajuste.

Más importante aún es la ventaja que aporta el abordaje desde la Teoría de Respuesta al Ítem: entre otras, la información que aporta el análisis de la curva característica de los ítems y del test, el estudio de la fiabilidad y la estimación del nivel de rasgo de los sujetos. Asimismo, los estudios de simulación con esta clase de modelos sugieren importantes recomendaciones prácticas con respecto al diseño de esta clase de test (cantidad de ítems, diseño de los bloques, dirección de los ítems, etc.), cuya descripción sobrepasa los propósitos de este manuscrito. De esta manera, el modelamiento de test de respuesta forzada desde la Teoría de Respuesta al Ítem se transforma en un avance importante para el estudio de la estructura interna de los datos provenientes de los test de respuesta forzada.

Comentarios finales

El presente artículo busca presentar el fundamento analítico que hace impropio el uso de las técnicas de análisis convencionales en la evaluación psicométrica de los test de formato de respuesta forzada. El principal problema radica en la utilización de un formato de codificación tradicional que lleva a la generación de puntuaciones ipsativas. Este tipo de puntuaciones posee características indeseables desde el punto de vista analítico.

Desde la perspectiva del estudio de la fiabilidad, estos rasgos radican fundamentalmente en el incumplimiento de uno de los supuestos básicos para la Teoría Clásica de los Test, como es la independencia de las puntuaciones, lo cual hace imposible la utilización de análisis de fiabilidad convencionales tales como el estadístico alfa de Cronbach y el test-retest.

En cuanto al estudio de la validez, las puntuaciones ipsativas generan matrices de covarianzas singulares debido a la dependencia lineal entre las variables, lo que hace imposible la aplicación de métodos de estimación de análisis factorial convencional como lo son los métodos de máxima verosimilitud o mínimos cuadrados generalizados. Por otro lado, la utilización de otro tipo de técnicas, como el ACP, es

desaconsejada, ya que lleva a la obtención de estructuras factoriales erróneas producto de la generación de factores bipolares artificiales debido a la presencia de términos negativos en las matrices de covarianzas, dada por la naturaleza ipsativa de las puntuaciones. Lo anterior nos lleva a la imposibilidad de la aplicación de técnicas convencionales para la búsqueda de evidencias de validez a través del análisis de la estructura interna de esta clase de datos.

Un avance prometedor en la solución de los problemas analíticos de los test de respuesta forzada es la propuesta de Maydeu-Olivares y Böckenholt (2005), así como la de Brown y Maydeu-Olivares (2011). Estos autores proponen un análisis sobre datos binarios producto de la comparación por pares entre ítems de un mismo bloque y la especificación de un modelo confirmatorio. Sin pretender presentar en detalle la formulación (para ello el lector puede consultar directamente la referencia), nuestro propósito ha sido únicamente el de esbozar a grandes rasgos la forma en que se está abordando el análisis de esta clase de test.

En síntesis, el propósito del presente trabajo ha sido presentar los errores que se cometen al utilizar técnicas convencionales en el análisis psicométrico de test con ítems de formato de respuesta forzada. A pesar de que estas dificultades son bien conocidas y que se las ha discutido ampliamente (Dunlap y Cornwell, 1994; Loo, 1999; Meade, 2004; McLean y Crissom, 1986), actualmente se sigue analizando y validando esta clase de test con técnicas convencionales diseñadas para datos normativos sin restricción ipsativa. Uno de los casos más emblemáticos corresponde al LSI de Kolb (1985), el cual se sigue validando y utilizando pese a las fuertes críticas que ha recibido (por ejemplo, Cornwell, Manfredo y Dunlap, 1991; Geiger, Boyle y Pinto, 1993; Henson y Hwang, 2002; Loo, 1996; Metallidou y Platsidou, 2008; Ruble y Stout, 1990). Es de suma importancia que los investigadores estén al tanto de los problemas y limitaciones de la evaluación psicométrica convencional de esta clase de instrumentos, así como de las actuales propuestas en el abordaje de los problemas asociados a los datos resultantes del formato de respuesta forzada.

Financiación

Este trabajo ha sido parcialmente financiado por el proyecto 2012-31958 del Ministerio de Economía y Competitividad de España.

Bibliografía

- Aitchinson, J. (1986). *The statistical analysis of compositional data*. Nueva York: Chapman & Hall.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56. doi: 10.1111%2Fj.2044-8325.1996.tb00599.x
- Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, 15, 263-272. doi: 10.1111%2Fj.1468-2389.2007.00386.x
- Brown, A. y Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71, 460-502. doi: 10.1177%2F0013164410375112

- Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review*, *51*, 292-303. doi: 10.1037%2Fh0057299
- Chan, W. (2003). Analyzing ipsative data in Psychological research. *Behaviormetrika*, *30*, 99-121. doi: 10.2333%2Fbhmk.30.99
- Chan, W. y Bentler, P. M. (1993). Covariance structure analysis of ipsative data. *Sociological Method and Research*, *22*, 214-247. doi: 10.1177%2F0049124193022002003
- Cheung, M. (2004). A direct estimation method on analyzing ipsative data with Chan and Bentler's (1993) method. *Structural Equation Modeling*, *11*, 217-243. doi: 10.1207%2Fs15328007sem1102_5
- Cheung, M. y Chan, W. (2002). Reducing the uniform response biases with ipsative measurement in multi-group confirmatory factor analysis. *Structural Equation Modeling*, *9*, 55-77. doi: 10.1207%2Fs15328007SEM0901_4
- Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs*, *14*. Richmond, VA: Psychometric Society.
- Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, *69*, 41-47. doi: 10.1111%2Fj.2044-8325.1996.tb00598.x
- Cornwell, J. M. y Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Seville and Willson (1991). *Journal of Occupational and Organizational Psychology*, *67*, 89-100. doi: 10.1111%2Fj.2044-8325.1994.tb00553.x
- Cornwell, J., Manfredo, P. y Dunlap, W. (1991). Factor analysis of the 1985 revision of Kolb's Learning Style Inventory. *Educational and Psychological Measurement*, *51*, 455-462. doi: 10.1177%2F0013164491512020
- Dunlap, W. P. y Cronwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, *29*, 115-126. doi: 10.1207%2Fs15327906mbr2901_4
- Geiger, M., Boyle, E. y Pinto, J. (1993). An examination of ipsative and normative versions of Kolb's revised Learning Style Inventory. *Educational and Psychological Measurement*, *53*, 717-726. doi: 10.1177%2F0013164493053003016
- Henson, R. y Hwang, D. (2002). Variability and prediction of measurement error in Kolb's Learning Style Inventory scores reliability: Generalization study. *Educational and Psychological Measurement*, *62*, 712-727. doi: 10.1177%2F0013164402062004011
- Hicks, L. E. (1970). Some properties of ipsative, normative and forced choice normative measures. *Psychological Bulletin*, *74*, 167-184. doi: 10.1037%2Fh0029780
- Johnson, C. B., Wood, R. y Blinkhorn, S. F. (1988). Spuriousness and spuriousness: The use of ipsative personality tests. *Journal of Occupational Psychology*, *61*, 153-162. doi: 10.1111%2Fj.2044-8325.1988.tb00279.x
- Jöreskog, K. G., y Sörbom, D. (1981). LISREL: *Analysis of linear structural relationships by the method of maximum likelihood* (versión V). Chicago: National Educational Resources, Inc.
- Kolb, D. A. (1985). *Learning Style Inventory*. Boston, MA: McBec and Company.
- Loo, R. (1996). Construct validity and classification stability of the revised Learning Style Inventory (LSI-1985). *Educational and Psychological Measurement*, *56*, 529-536. doi: 10.1177%2F0013164496056003015
- Loo, R. (1999). Issues in factor-analyzing ipsative measurement: The Learning Style Inventory (LSI-1985) example. *Journal of Business and Psychology*, *14*, 149-154. doi: 10.1023%2FA%3A1022918803653
- McLean, J. y Crisson, B. (1986). *Multivariate analysis of ipsative data: problems and solutions*. Trabajo presentado en el Annual Meeting of the Mid-South Educational Research Association. Memphis, Tennessee.
- Maydeu-Olivares, A. y Böckenholt, U. (2005). Structural equation modeling of paired comparisons and ranking data. *Psychological Methods*, *64*. 325-340. doi: 10.1037%2F1082-989X.10.3.285
- Maydeu-Olivares, A. y Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research*, *45*, 935-974. doi: 10.1080%2F00273171.2010.531231
- Meade, A. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*, 531-552. doi: 10.1348%2F0963179042596504
- Metallidou, P. y Platsidou, M. (2008). Kolb's Learning Style Inventory-1985: Validity issues and relations with metacognitive knowledge about problem-solving strategies. *Learning and Individual Differences*, *18*, 114-119. doi: 10.1016%2Fj.lindif.2007.11.001
- Múthen, L. K. y Múthen B. O. (1998, 2010). *Mplus User's Guide*. 6.ª ed. Los Angeles, CA: Muthén&Muthén. Disponible en: <http://www.statmodel.com>
- Ruble, T. L. y Stout, D. E. (1990). A psychometric analysis of Kolb's revised Learning Style Inventory. *Developments in Business Simulation & Experiential Exercises*, *17*, 147-149.
- Saville, P. y Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, *64*, 219-238. doi: 10.1111%2Fj.2044-8325.1991.tb00556.x
- Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C. y Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*, 943-953. doi: 10.1037%2F0021-9010.86.5.943
- Tenopyr, M. L. (1988). Artfactual reliability of forced-choice scales. *Journal of Applied Psychology*, *74*, 749-751. doi: 10.1037%2F0021-9010.73.4.749
- Thurstone, L. L. (1931). Rank order as a psychological method. *Journal of Experimental Psychology*, *14*, 187-201. doi:10.1037/h0070025
- Ximénez, C. y Calderón, C. (2012). El análisis factorial de datos ipsativos: un estudio de simulación. *Psicothema*, *24*, 302-309.