

Efectos de método y formatos de ítems en la medición de constructos bipolares*

Ana Hernández, Vicente González-Romá y Begoña Espejo.

Universitat de València.

El Análisis Factorial Confirmatorio (AFC) es un método que puede emplearse para analizar matrices multirrasgo multimétodo (MRMM) a fin de evaluar la incidencia de los métodos empleados en la medición de constructos psicológicos. El objetivo del presente trabajo fue estimar y comparar, mediante la parametrización de Marsh (1988,1989) los efectos de método introducidos por 3 formatos de ítem empleados en la medición del bienestar psicológico (formato de diferencial semántico, formato de respuesta bipolar con anclajes verbales, y formato tipo Likert). La matriz input consistió en las correlaciones entre medidas de tres dimensiones del bienestar psicológico obtenidas a través de los tres métodos considerados. La muestra de estudio estaba compuesta por 143 sujetos empleados en un Organismo de la Administración del Estado. De los tres formatos considerados, fue el formato tipo Likert el que introdujo efectos de método más importantes, seguido del método con formato de diferencial semántico. El formato de respuesta bipolar con anclajes verbales no introdujo efectos de método relevantes.

Palabras clave: Matrices multirrasgo-multimétodo, Análisis Factorial Confirmatorio, Efectos de método, formatos de ítems.

Para medir constructos psicológicos se ha recurrido a distintos métodos de escalamiento (Likert, Thurstone, Diferencial Semántico, etc.), si bien puede afirmarse que, debido a su sencillez, el método de escalamiento unidimensional de Likert ha sido el más utilizado. Sin embargo, además de la sencillez del procedimiento, otros aspectos deberían ser considerados a la hora de elegir el método de medición a utilizar. Uno de esos aspectos es la varianza en las respuestas que se puede atribuir al método de medición empleado. Como señalan Campbell y Fiske (1959) cualquier instrumento de medición psicológico puede concebirse como una unidad rasgo-método, de modo que tanto el rasgo medido como el método empleado eliciten varianza

* Los autores agradecen los comentarios al manuscrito original realizados por dos revisores anónimos.

sistemática en las respuestas de los sujetos. La determinación de los efectos introducidos por los métodos de medición es de suma importancia, ya que, en primer lugar, son deseables tests que reflejen adecuadamente las diferencias en el rasgo medido y en los que el método empleado explique sólo una pequeña parte de la varianza observada; de otro modo, la validez del test quedaría seriamente cuestionada (Campbell y Fiske, 1959). En segundo lugar, cabe señalar que la varianza debida al método puede distorsionar las correlaciones obtenidas entre rasgos diferentes medidos con el mismo método (Marsh, 1989), lo que daría pie a conclusiones erróneas.

Para evaluar los efectos de método que introducen los procedimientos de medición psicológica, Campbell y Fiske (1959) propusieron el análisis de matrices multirrasgo-multimétodo. Siguiendo los criterios propuestos por Campbell y Fiske (1959) se puede evaluar el efecto que introduce cada método de medida comparando la magnitud de las correlaciones entre distintos rasgos medidos con un mismo método (bloque heterorrasgo-monométodo), con las correlaciones entre esos mismos rasgos medidos a través de distintos métodos (bloque heterorrasgo-heterométodo). Es decir, se trata de comparar la magnitud de las correlaciones heterorrasgo-monométodo con las correlaciones "paralelas" en el triángulo heterorrasgo-heterométodo correspondiente. La existencia de efectos de método queda determinada cuando las correlaciones heterorrasgo-monométodo son más elevadas en términos generales que las correspondientes heterorrasgo-heterométodo. Esta aproximación ha sido utilizada frecuentemente (Ostrom, 1969; Byrne y Shavelson, 1986; Schiersheim, Solomon y Kopelman, 1989), si bien presenta una serie de importantes limitaciones. La primera de ellas se refiere al elevado número de comparaciones que requiere la evaluación de los efectos de método (Marsh, 1989; Marsh y Bailey, 1991; Widaman, 1985). En este punto cabe señalar que "estas comparaciones no son pruebas de significación estadística, y pruebas de significación apropiadas serían difíciles de idear para tantas comparaciones no independientes" (Marsh, 1988; pp. 571-572). La segunda limitación hace referencia a la arbitrariedad de los criterios de decisión, ya que es el investigador el que decide en qué medida existen o no efectos de método. La tercera alude a la incapacidad de ofrecer estimaciones de la proporción de varianza de las variables que es atribuible a los rasgos y a los métodos, y la imposibilidad de descomponer las correlaciones observadas entre las medidas (Bagozzi, 1978; Widaman, 1985; Marsh, 1988; Marsh y Hocevar, 1983). En cuarto lugar cabe destacar la conveniencia de que las medidas tengan una fiabilidad similar, ya que si no, las intercorrelaciones pueden quedar distorsionadas, e invalidar las inferencias realizadas (Widaman, 1985; Marsh, 1988). Y en quinto lugar la limitación de los supuestos en que se basa esta aproximación (Schmitt y Stults, 1986), ya que asume que tanto las correlaciones entre los rasgos y los métodos de medida, como las correlaciones entre los métodos son nulas, suponiendo que todos los rasgos son influidos de igual forma por los métodos de medida (Althauser y Heberlin, 1970; Alwin, 1974; Kalleberg y Kluegel, 1975).

Para solucionar algunas de las limitaciones que plantean los criterios de Campbell y Fiske (1959), se ha utilizado frecuentemente el análisis factorial

confirmatorio (AFC) en el análisis de las matrices multirrasgo-multimétodo. En el AFC, los efectos que introduce cada método pueden ser operacionalizados a partir de distintos tipos de parametrizaciones: la parametrización estándar, la propuesta por Rindskopf (1983), y la estructura del método 'E' o parametrización de las unicidades correlacionadas de Marsh (1988, 1989), cada una de las cuales tiene una serie de ventajas e inconvenientes. Mediante las dos primeras se puede descomponer la varianza de cada variable observada en tres partes: 1. varianza explicada por los factores de rasgo; 2. varianza explicada por los factores de método; y 3. varianza residual. Sin embargo, el uso de estas parametrizaciones genera con gran frecuencia "soluciones pobremente definidas", siendo muy frecuente la aparición de casos Heywood (varianzas de error negativas o iguales a cero) (Jöreskog, 1981; Dillon, Kumar y Mulani, 1987; Marsh, 1989; Marsh y Bailey, 1991). Desde una conceptualización distinta de la varianza asociada al método (no se asume la unidimensionalidad ni la estructura congenérica de los efectos de método), la parametrización de Marsh permite superar los problemas mencionados. Puesto que en esta parametrización no se incluyen factores de método, la contribución de dichos efectos se estima a partir de las correlaciones entre los residuales de las variables medidas con el mismo método, quedando las correlaciones entre los residuales de las variables medidas con métodos distintos fijadas a 0 (ver figura 1). Así, al igual que sucede con los criterios de Campbell y Fiske (1959) la parametrización de Marsh asume que las correlaciones entre los métodos son nulas.

Al hablar de los efectos de método, se puede hacer referencia a los efectos introducidos por distintos tests, por distintos métodos de escalamiento, por distintos observadores, en diferentes momentos temporales, o también por distintos formatos de items (Marsh y Butler, 1984; Bagozzi y Yi, 1991; Goffin y Jackson, 1992). Por lo que se refiere a los efectos de método introducidos por distintos formatos de items, dichos efectos no han sido estudiados previamente de forma explícita. Sin embargo, sí existen estudios donde se han comparado los efectos introducidos por distintos métodos de escalamiento (Bagozzi, 1978; Byrne y Shavelson, 1986), que en algunos casos también comportaban items con formato diferente. Concretamente Marsh (1989) analizó mediante distintas parametrizaciones -la parametrización estándar, la propuesta por Rindskopf (1983) y la parametrización de las unicidades correlacionadas (Marsh, 1988; 1989)- la matriz multirrasgo-multimétodo del estudio de Byrne y Shavelson (1986). Dicha matriz, contenía las correlaciones entre tres dimensiones monopulares del autoconcepto medidas a través de tres métodos de escalamiento (Likert, diferencial semántico y Guttman). Cabe destacar que tanto el método Likert como el método Guttman, consistían en escalas cuyos items tenían formato de respuesta tipo Likert. Los resultados mostraron que la parametrización estándar convergía en una solución adecuada, mostrando los mejores índices de bondad de ajuste, si bien el método de las unicidades correlacionadas también mostró un ajuste adecuado, ofreciendo estimaciones y conclusiones similares a las obtenidas mediante la parametrización estándar. Tanto para el método de diferencial semántico como para el método de Guttman todas las saturaciones presentadas en los factores de método

correspondientes fueron estadísticamente significativas, mientras que para el método Likert sólo una de las tres saturaciones en el factor de método correspondiente resultó estadísticamente significativa. Desde aquí se observa que el formato de respuesta tipo Likert introduce diferentes efectos de método dependiendo del método de escalamiento empleado, oscilando entre efectos moderados, en el caso del método de Likert, y altos, en el caso del método de Guttman. Los métodos de escalamiento de diferencial semántico y de Guttman explicaron respectivamente el 17.54% y al 17.71% de la varianza observada, mientras que el método de Likert explicó el 11.56% de la varianza observada.

Teniendo en cuenta la ausencia de estudios acerca de los efectos de método que introducen los distintos formatos de respuesta de los ítems, el objetivo del presente estudio es contribuir a clarificar, mediante análisis factorial confirmatorio, la magnitud de los efectos de método asociados a distintos formatos de ítems frecuentemente utilizados en la medida de constructos bipolares, y más concretamente del bienestar psicológico. En concreto se trata de un formato de diferencial semántico, un formato de respuesta bipolar con anclajes verbales, y un formato tipo Likert.

METODO

Muestra. La muestra utilizada en este estudio está compuesta por 143 sujetos pertenecientes a un organismo de la Administración Central del Estado, con sede en Andalucía. Del total de la muestra, el 36.4% de los sujetos son varones y el 53.1% mujeres (esta variable presenta un 10.5% de datos faltantes), con una edad media muestral igual a 35.46 años (d.t.=8.50). En cuanto a la antigüedad en la empresa, la media muestral es igual a 9.90 años (d.t.= 8.98), siendo la antigüedad profesional media igual a 10.11 años, (d.t.= 8.79). Respecto al nivel jerárquico, el 4.2% de los sujetos que componen la muestra son directivos, el 7.7% son jefes de sección o agencia, el 13.3% jefes de negociado, el 11.9% jefes de equipo, y el 52.4% son empleados.

Instrumentos de medida. El bienestar psicológico es un componente de carácter multidimensional de la salud psicológica. Su evaluación se ha realizado mediante la medición de tres de los ejes o dimensiones bipolares propuestas en el modelo de salud mental de Warr (1987): el eje ansioso-calmado, el eje cansado-llevo de energía y el eje deprimido-entusiasmado. Los métodos de medida empleados han sido los siguientes:

- METODO 1.- Se trata de una pregunta con formato similar al del diferencial semántico, donde se formula un ítem bipolar para cada dimensión. Los sujetos deben situarse en uno de los siete espacios que existen entre cada par de adjetivos. Por ejemplo: "Indique cómo le ha hecho sentirse su trabajo, en términos generales, durante las últimas semanas, situándose en uno de los siete espacios que se hallan entre cada par de adjetivos en función de su

mayor
DEPR

que de
ello un
ansios
muy d
seman
muy
entusi
7. mu

referen
grado
escala
sentid
2. bas
acuere

a una
Máxin
Sörbo
ya qu
altern
pobres
Marst

TABI

R1M1
R2M1
R3M1
R1M2
R2M2
R3M2
R1M3
R2M3
R3M3
R1: eje
M1: mé

correl
unidit
Así, e
que l:

mayor o menor proximidad a uno de los adjetivos que forman cada par: DEPRIMIDO _.:.:.:.: ENTUSIASMADO"

- MÉTODO 2.- Consiste en una pregunta en la que se pide a los sujetos que describan cómo se han sentido en las últimas semanas, utilizando para ello una escala de respuesta de 7 puntos que oscila, en cada caso, desde muy ansioso a muy calmado, desde muy cansado a muy lleno de energía, y desde muy deprimido a muy entusiasmado. Por ejemplo: "Pensando en las últimas semanas, y en términos generales, ¿cómo le ha hecho sentirse su trabajo? 1. muy deprimido, 2. bastante deprimido, 3. algo deprimido, 4. ni entusiasmado ni deprimido, 5. algo entusiasmado, 6. bastante entusiasmado, 7. muy entusiasmado."

- MÉTODO 3.- Se trata de tres afirmaciones cada una de las cuales hace referencia a cada uno de los rasgos medidos. Los sujetos deben indicar el grado de acuerdo que muestran con cada afirmación, utilizando para ello una escala de respuesta tipo Likert con siete categorías. Por ejemplo: "Me he sentido deprimido durante las últimas semanas. 1. totalmente en desacuerdo, 2. bastante en desacuerdo, 3. algo en desacuerdo, 4. incierto, 5. algo de acuerdo, 6. bastante de acuerdo y 7. totalmente de acuerdo."

Análisis. La matriz MRMM que se obtuvo (ver tabla 1) fue sometida a una serie de análisis factoriales confirmatorios empleando el método de Máxima Verosimilitud a través del programa LISREL VII (Jöreskog y Sörbom, 1989). En los análisis se utilizó la parametrización de Marsh (1989), ya que tanto mediante la parametrización estándar como mediante la alternativa propuesta por Rindskopf (1983), se suelen obtener soluciones pobremente definidas (Jöreskog, 1981; Marsh, 1989; Marsh y Bailey, 1991; Marsh, Byrne y Craven, 1992).

TABLA 1. Matriz de correlaciones.

	R1M1	R2M1	R3M1	R1M2	R2M2	R3M2	R1M3	R2M3	R3M3
R1M1	1.000								
R2M1	0.475	1.000							
R3M1	0.417	0.644	1.000						
R1M2	0.659	0.409	0.439	1.000					
R2M2	0.347	0.567	0.418	0.364	1.000				
R3M2	0.510	0.498	0.604	0.533	0.508	1.000			
R1M3	-0.518	-0.285	-0.306	-0.532	-0.235	-0.479	1.000		
R2M3	-0.366	-0.449	-0.330	-0.411	-0.400	-0.503	0.633	1.000	
R3M3	0.261	0.396	0.459	0.328	0.541	0.535	-0.122	-0.195	1.000

R1: eje 'ansioso-calmado'; R2: eje 'deprimido-entusiasmado'; R3: eje 'cansado-lleño de energía'

M1: método 1; M2: método 2; M3: método 3

La parametrización de Marsh o parametrización de las unicidades correlacionadas (Marsh 1988,1989, Marsh y Bailey, 1991), no asume la unidimensionalidad ni la estructura congénica de los efectos de método. Así, en el modelo se incluyen tres factores de rasgo correlacionados, mientras que la contribución de los efectos de método es estimada a partir de las

correlaciones entre los residuales de las variables referidas a los rasgos medidos con un mismo método. Si estas correlaciones son estadísticamente significativas, puede afirmarse la existencia de efectos de método relevantes. Como se indicó anteriormente, las correlaciones entre los residuales de variables medidas con distintos métodos son fijadas a cero, asumiéndose así la independencia de los efectos de método (ver figura 1). Sin embargo, diversos estudios (Marsh, Byrne y Craven, 1992; Kenny y Kashy, 1992) han mostrado como, incluso en el caso de que los factores de método estén altamente correlacionados, la parametrización de las unicidades correlacionadas converge en soluciones más adecuadas y precisas en relación al valor de los parámetros poblacionales, que la parametrización estándar del análisis factorial confirmatorio.

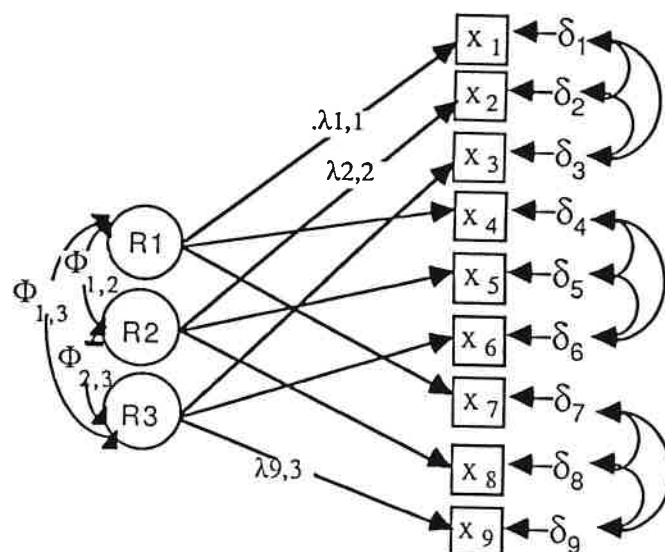


Figura 1. Parametrización de Marsh (1988,1989).

RESULTADOS

En primer lugar, por lo que se refiere a la bondad de ajuste del modelo, todos los índices pueden considerarse satisfactorios, ya que el estadístico χ^2 no alcanza la significación estadística ($\chi^2=12.6$; $p>0.05$), el índice de bondad de ajuste (GFI) y el índice de bondad de ajuste ajustado (AGFI) alcanzan valores superiores a 0.90 (GFI=0.97; AGFI=0.92), y el promedio de los residuales estandarizados (RMSR) es inferior a 0.05 (RMSR=0.044).

Además, la solución factorial es adecuada, ya que todos los parámetros del modelo presentan estimaciones dentro de su rango de valores permisibles, y presentan errores estándar con valores aceptables.

TABLA 2. Correlaciones entre los errores de las variables medidas con un mismo método en la parametrización de Marsh (en la diagonal aparecen las varianzas residuales).

		M 1			M 2			M 3		
		ANS	DEP	CAN	ANS	DEP	CAN	ANS	DEP	CAN
M 1	ANS	.405								
	DEP	.133*	.482							
	CAN	.040	.263**	.540						
M 2	ANS				.292					
	DEP				-.059	.387				
	CAN				-.046	-.067	.200			
M 3	ANS							.596		
	DEP							.359**	.630	
	CAN							.151*	.171**	.600

* $p < 0.05$; ** $p < 0.01$; ANS: eje 'ansioso-calmado'; DEP: eje 'deprimido-entusiasmado'; CAN: eje 'cansado-lleno de energía'. M1: método 1; M2: método 2; M3: método 3

En cuanto a los efectos de método, los resultados mostraron que, en relación al método 1 (formato de diferencial semántico), 2 de las 3 correlaciones entre unicidades estimadas resultaron ser estadísticamente significativas (ver tabla 2). Ninguna de las correlaciones entre los residuales de las variables que fueron medidas mediante el método 2 (formato de respuesta bipolar con anclajes verbales) fueron estadísticamente significativas. Asimismo, los resultados referidos al método 3 (formato tipo Likert) mostraron que todas las correlaciones entre los residuales de las variables medidas con dicho método resultaron ser estadísticamente significativas. Estos resultados indican que los métodos 1 y 3 introducen efectos relevantes, mientras que el método 2 no. Por otra parte, a juzgar por el número de correlaciones estadísticamente significativas y la magnitud de las mismas, es el método 3 el que produce efectos de mayor magnitud.

Para comprender en mayor grado la importancia de cada uno de los métodos empleados, se realizaron una serie de análisis factoriales confirmatorios para comparar pares de modelos anidados. En primer lugar, se contrastó un modelo en el que se asumía que el método 1 no introducía efectos de método, para lo cual se fijaron a cero las correlaciones entre los residuales de las variables referidas a los rasgos medidos con este método. El ajuste de este modelo se comparó con el ajuste del modelo inicial comentado anteriormente mediante un test de la diferencia entre los índices χ^2 asociados a cada modelo. Esta diferencia se distribuye como una χ^2 con tantos grados de libertad como la diferencia entre los grados de libertad de los modelos. Puesto que se trata de modelos anidados, si la diferencia entre las χ^2 de ambos modelos no resulta estadísticamente significativa, concluiremos que el método 1 no introduce efectos relevantes. Este mismo proceso de comparación se desarrolló en relación a los otros dos métodos utilizados. Los índices de bondad de ajuste de cada uno de los modelos con restricciones, así

como las diferencias entre los índices χ^2 de éstos y el modelo inicial, pueden observarse en la tabla 3.

TABLA 3. Índices de bondad de ajuste y comparación de modelos

MODELO	χ^2	g.l.	$\chi^2/g.l.$	G.F.I.	A.G.F.I.	R.M.S.R.	diferencia entre χ^2
$\theta_{M1}^\delta = 0$	50.42*	18	2.80	0.926	0.814	0.050	32.8*
$\theta_{M2}^\delta = 0$	20.42	18	1.13	0.968	0.919	0.044	2.8
$\theta_{M3}^\delta = 0$	73.52*	18	4.08	0.890	0.725	0.075	55.9*

* $p < 0.01$

θ_{Mi}^δ = modelo "inicial-Marsh" en el que se han fijado a 0 las correlaciones entre los residuales de las variables referidas a los rasgos medidos mediante el Método i.

Tanto el modelo que asume que el método 1 (formato de diferencial semántico) no introduce efectos de método, como el modelo que asume la inexistencia de efectos debidos al método 3 (formato de Likert), ajustan significativamente peor ($p < 0.01$) que el modelo inicial. Por otra parte, el ajuste del modelo en el que las correlaciones entre los residuales de los rasgos medidos con el método 2 (formato de respuesta bipolar con anclajes verbales) son fijadas a cero, no difiere significativamente de aquél en el que dichas correlaciones son estimadas libremente (modelo inicial), siendo además todos los índices de bondad de ajuste de ambos modelos muy similares.

Por último, a partir de las estimaciones ofrecidas por el modelo inicial se calcularon los porcentajes de varianza explicados por los factores de rasgo para cada uno de los métodos utilizados (ver tabla 4). La cantidad de varianza atribuible a cada factor de rasgo se obtiene elevando al cuadrado las saturaciones factoriales de cada factor de rasgo. La varianza de error de cada variable observada es estimada directamente en la diagonal de la matriz Θ , que en esta parametrización contiene tanto la varianza de error aleatorio como la varianza atribuible a los métodos de medición considerados. Los porcentajes obtenidos indicaron que fue el método 2 (formato de respuesta bipolar con anclajes verbales) el que medía más adecuadamente los factores de rasgo considerados, ya que fue el que ofreció un mayor porcentaje de varianza debida a los factores de rasgo, y un menor porcentaje de varianza de error (efectos de método más residuales).

TABLA 4. Porcentaje de varianza explicada por el rasgo y porcentaje de varianza no explicado (efectos de método+residuales)

	Rasgo	Error
Formato similar al del diferencial semántico		
Ansiedad	60%	40%
Depresión	54%	48%
Cansancio	48%	54%
Formato de respuesta bipolar		
Ansiedad	71%	29%
Depresión	61%	39%
Cansancio	80%	20%
Formato de respuesta tipo Likert		
Ansiedad	36%	60%
Depresión	38%	63%
Cansancio	41%	60%

DISCUSION

El objetivo del presente trabajo era determinar, mediante AFC de matrices multirrasgo-multimétodo, los efectos de método que introducen diferentes formatos de ítems en la medida de tres dimensiones bipolares del bienestar psicológico.

Los resultados obtenidos indicaron que tanto el método 1 (pregunta con formato de diferencial semántico) como el método 3 (pregunta con formato de respuesta tipo Likert) introducen efectos de método relevantes, mientras que el método 2 (pregunta con formato de respuesta bipolar con anclajes verbales) no parece introducir efectos de método en la medición del bienestar psicológico (ver tabla 2). A estas conclusiones se ha llegado tras observar que tanto el método 1 como el método 3 presentan correlaciones significativas entre los residuales de las variables referidas a los rasgos medidos con cada uno de estos métodos, mientras que con el método 2 dichas correlaciones no son significativas. Además, al comparar el ajuste de tres modelos -cada uno de los cuales hipotetizaba sucesivamente que cada uno de los métodos considerados no introducía efectos relevantes- con el ajuste del modelo inicial -que hipotetizaba la existencia de efectos introducidos por estos métodos-, se observó que es el modelo que asume que el método 3 no introduce efectos aquél cuyo ajuste difiere en mayor grado del ajuste del modelo inicial (diferencia de $\chi^2=55.9$; $p<0.01$), seguido del modelo que asume la inexistencia de efectos significativos para el método 1 (diferencia de $\chi^2=32.8$; $p<0.01$). En cuanto al modelo que hipotetiza que el método 2 no introduce efectos de método, su ajuste no difiere significativamente del ajuste del modelo inicial. Estos resultados indican que las hipótesis de que los métodos 1 y 3 no introducen efectos significativos no se pueden mantener, y ratifican que el método 2 no introduce efectos relevantes.

Los resultados obtenidos en el presente estudio son congruentes con los obtenidos por Marsh (1989) a partir de la matriz MRMM del estudio de Byrne y Shavelson (1986), donde tanto los ítems con formato de diferencial semántico como los ítems con formato de respuesta Likert, introdujeron efectos relevantes, aunque en este último caso, y dependiendo del método de escalamiento, los efectos oscilaban de moderados a altos.

En nuestro caso, el peor funcionamiento del formato tipo Likert, frente a los otros formatos, puede deberse en parte al hecho de que, si bien se pretende medir un constructo bipolar, el ítem es explícitamente monopolar, esto es, alude a uno solo de los polos de la dimensión. Este formato es frecuente en la medición de los constructos bipolares en cuestionarios que incluyen ítems referidos a cada uno de los polos de la dimensión (p. ej. Warr, 1990). Antes de sumar las puntuaciones de los ítems que componen la escala, las respuestas a los ítems de un polo son invertidas, asumiendo que la respuesta a un ítem referido a un polo, indica con exactitud la ubicación del sujeto en la dimensión bipolar. A este respecto cabe señalar que Lloret (1995) ha demostrado empíricamente que este supuesto no es sostenible.

En relación a los formatos de ítem explícitamente bipolares (métodos 1 y 2), los resultados obtenidos indican que el formato que ofrece anclajes verbales concretos para cada una de las alternativas de respuesta (método 2) es el más adecuado para medir las dimensiones bipolares del bienestar psicológico. Este formato fue el único que no introdujo efectos de método relevantes, y el que ofreció los porcentajes de varianza atribuibles a los rasgos más elevados, oscilando entre el 61 y el 80%. Los investigadores interesados en constructos bipolares (autoestima, bienestar psicológico, conservadurismo-liberalismo, etc.) deberían tener en cuenta estos resultados para mejorar la calidad de sus mediciones.

En el futuro, esta investigación podría ser objeto de una replicación operacional (Bedeian, Mossholder, Kernery y Armenakis, 1992), en la que en lugar de trabajar con medidas monoítem de cada rasgo, se emplearan escalas compuestas por ítems que respondieran a los tres formatos considerados.

ABSTRACT

Confirmatory Factor Analysis (CFA) is often used to analyze MTMM data in order to evaluate method effects in the measure of psychological constructs. The aim of this paper was to estimate and compare, by means of Marsh's parameterization (1988,1989), the method effects associated with three item formats used to measure psychological well-being (semantic differential format, items with bipolar response scale and verbal anchors and typical Likert format). The input matrix was composed of correlations among measures of three dimensions of psychological well-being obtained by means of the aforementioned methods. The sample was made up of 143 employees of a Public Administration Agency. The Likert format showed the greatest method effects. The Semantic Differential format showed

relevant method effects too, whereas the method effects of the bipolar response item format with verbal anchors were irrelevant.

Key words: Multitrait-multimethod matrices, Confirmatory Factor Analysis, Method Effects, Item Formats

REFERENCIAS

- Althausen, R. P. y Heberlin, T.A. (1970). Validity and the multitrait-multimethod matrix. In E. F. Borgotta & W. Bohrnstedt (Eds.), *Sociological methodology*. San Francisco CA: Jossey Bass.
- Alwin, D. F. (1974). Approaches to the interpretation of relationships and the multitrait-multimethod matrix. In E. F. Borgotta & W. Bohrnstedt (Eds.), *Sociological methodology*. San Francisco CA: Jossey Bass.
- Bagozzi, R. P. (1978). The construct validity of the affective, behavioral, and cognitive components of attitude by analysis of covariance structures. *Multivariate Behavioral Research*, 13, 9-31.
- Bagozzi, R. P. y Yi, Y. (1991). Multitrait-multimethod matrices in consumer research. *Journal of Consumer Research*, 17, 426-439.
- Bedeian, A. G.; Mossholder, K. W.; Kemery, E. R. y Armenakis, A. A. (1992). Replication requisites: a second look at Klenke-Hamel and Mathieu (1990). *Human Relations*, 45, 1093-1117.
- Byrne, B. M., y Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology* 78, 474-481.
- Campbell, D. T. y Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Dillon, W. R.; Kumar, A. y Mulani, N. (1987). Offending estimates in covariance structures analysis: comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, 101, 126-135.
- Goffin, R. D. y Jackson, D. N. (1992). Analysis of multitrait-multirater performance appraisal data: composite direct product method versus confirmatory factor analysis. *Multivariate Behavioral Research*, 27, 363-385.
- Jöreskog, K. G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics*, 8, 65-92.
- Jöreskog, K. G. y Sorböm, D. (1989). *LISREL VII User's reference Guide*: Mooresville, Inc: Scientific Software, Inc.
- Kalleberg, A. L. y Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: some limitations and an alternative. *Journal of Applied Psychology*, 60, 1-9.
- Kenny, D. A. y Kashy, D. A. (1992). The analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Lloret, S. (1995). *Medición de constructos afectivos bipolares mediante escalas tipo Likert: el desajuste entre el modelo de escalamiento, el método de análisis y el objeto de medida*. Tesis Doctoral. Universidad de Valencia.
- Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: an international handbook*. Oxford: Pergamon.
- Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: many problems and a few solutions. *Applied Psychological Measurement*, 13, 335-361.

- Marsh, H. W. y Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: a comparison of alternative models. *Applied Psychological Measurement, 15*, 47-70.
- Marsh, H. W. y Butler, S. (1984). Evaluating reading diagnostic tests: an application of Confirmatory Factor Analysis to Multitrait-Multimethod data. *Applied Psychological Measurement, 8*, 307-320.
- Marsh, H.W., Byrne, B. M. y Craven, R. (1992). Overcoming problems in Confirmatory Factor Analyses of MTMM data: the correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research, 27*, 489-507.
- Marsh, H. W. y Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20*, 231-248.
- Ostrom, T. M.(1969). The relationship between the affective, behavioral and cognitive components of attitude. *Journal of Experimental Social Psychology, 5*, 12-30.
- Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika, 48*, 73-83.
- Schmitt, N. y Stults, D.M.(1986). Methodology review: analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1-22.
- Warr, P. (1987). *Work, Unemployment and Mental Health*. New York: Oxford University Press.
- Warr, P. (1990). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology, 63*, 190-210.
- Widaman, K.F.(1985). Hierarchicall nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.

(Revisión aceptada: 18/6/97)