

Evidencia acumulada sobre los efectos de método asociados a ítems invertidos

Tomás, J. M.* , Galiana, L., Hontangas, P., Oliver, A., Sancho, P.

Universitat de València

La recomendación de invertir ítems al medir constructos psicológicos mediante escalas ha estado presente casi desde siempre en la práctica psicométrica. Pese a esto, esta estrategia parece interferir en el examen de la estructura latente de las escalas, tal y como se ha demostrado en diversas investigaciones. Este trabajo pretende aportar evidencia empírica que permita clarificar el efecto de la formulación invertida de los ítems, utilizando dos versiones de la Escala de Autoestima de Rosenberg (1965), una en su formato original con ítems positivos e invertidos y otra transformada de modo que todos los ítems sean positivos, para comprobar si se mantienen o desaparecen los efectos de método encontrados en la literatura asociados a los ítems invertidos. Para ello, se ha contado con una muestra incidental de 390 estudiantes de instituto y universidad de la ciudad de Valencia. Se han llevado a cabo diversos modelos de ecuaciones estructurales (incluyendo CTCM) alternativos, comparando los índices de ajuste de éstos. A partir de los resultados, se obtienen dos puntos relevantes de información: por una parte, se observa de nuevo un efecto de método asociado a los ítems invertidos; por otra, se observa como estos ítems, cuando son formulados en positivo, no presentan tal efecto, de manera que la Escala de Autoestima de Rosenberg presenta una estructura de un solo factor: el factor de autoestima. Las implicaciones de estos resultados se recogen en las conclusiones y discusión.

La práctica y recomendación de invertir ítems al medir constructos psicológicos mediante escalas ha estado presente casi desde siempre en la medición psicológica y educativa (DeVellis, 1991; Nunnally, 1978). Esta recomendación se basa fundamentalmente en el intento de disipar los

* Agradecimientos: La investigación ha sido parcialmente financiada por un proyecto de investigación del Ministerio de Educación concedido al primer autor (Proyecto PSI 2010-21334) y con una beca del gobierno de España (proyecto PSI2010-21334). Autor para correspondencia: Dr. José Manuel Tomás Miguel, Departamento de Metodología de les Ciències del Comportament, Facultat de Psicologia, Universitat de València, Av. Blasco Ibáñez, 21, 46010 València (Spain). E-mail: tomasjm@uv.es

efectos del sesgo de aquiescencia, que puede definirse como la tendencia a responder de forma afirmativa con independencia del contenido de los ítems (Morales, 2006), o como un estilo de respuesta que hace un uso desproporcionado de las opciones de respuesta positivas (Weijters, Geuens, y Schillewaert, 2010). Ferrando y Lorenzo-Seva (2010) muestran analíticamente que el sesgo debido al estilo de respuesta aquiescente influye en la calibración de los ítems y en la puntuación de los participantes, para el caso de que las escalas sean unidimensionales y los ítems invertidos se encuentren balanceados con los de sentido positivo. En la calibración, en presencia de aquiescencia, se espera un peor ajuste del modelo unidimensional si no se considera el factor de aquiescencia. En la puntuación se espera una estimación insesgada, ya que los ítems invertidos compensarían la mayor puntuación de los positivos. Sin embargo, si no se invirtieran ítems, se esperaría una estimación sesgada positiva. El modelo factorial se ha empleado para evaluar y controlar la aquiescencia, separando las influencias separadas del rasgo que se pretende medir y de la aquiescencia (Billiet y McClendon, 2000; Ferrando y Condon, 2006). Estos modelos son generalmente bidimensionales, conceptualizando la aquiescencia como un factor añadido. Ferrando, Lorenzo-Seva, y Chico, (2003), y posteriormente, Ferrando y Lorenzo-Seva (2010), aplicaron estos modelos de dos factores a escalas de personalidad, encontrando efectivamente un factor de aquiescencia donde los ítem formulados en el sentido del rasgo cargaban positivamente, mientras los invertidos lo hacían en negativo. Por todo ello, se ha recomendado construir escalas con un balance o equilibrio entre el número de ítems positivos, en el sentido del constructo a medir, e invertidos, en la dirección opuesta. Esta práctica era, y todavía sigue siendo, un rasgo común a la construcción de todo tipo de escalas en psicología y educación, ya sean de actitudes, opinión, personalidad, auto-concepto, etc.

Pese a ser práctica común, ya Carmines y Zeller (1979), en un manual sobre evaluación de la fiabilidad y la validez, mostraron que esta estrategia podía tener un precio: interferir con el examen de la estructura latente de las escalas en las que se empleaba. Para mostrar este efecto emplearon la escala de autoestima de Rosenberg en su versión original de 10 ítems con cinco invertidos. Como técnicas para valorar la interferencia que defendía emplearon análisis factorial exploratorio (AFE) y coeficientes de validez criterial. En el AFE encontraron una solución de dos factores que agrupaban a los ítems positivos, en el sentido de mayor puntuación mayor autoestima, por un lado, y a los invertidos por otro. Tras esto, correlacionaron los dos factores (autoestima positiva y negativa) con 16 criterios y contrastaron si las correlaciones de estos dos factores con cada uno de los criterios eran

iguales o diferentes. En ningún caso se encontraron diferencias estadísticamente significativas ($p > .25$). Evidentemente, este último análisis hacía difícil contemplar ambos factores como constructos sustantivamente diferentes, por lo que concluyeron que a la base de esta escala había un factor de autoestima contaminado por un artefacto debido al método de medida (la inversión).

Desde este trabajo seminal, muchos otros investigadores han encontrado estos efectos de método asociados a los ítems invertidos (e.g. Bachman y O'Malley, 1986; Bagozzi, 1993; Carmines y Zeller, 1974, 1979; Corwyn, 2000; Goldsmith, 1986; Hensley y Roberts, 1976; Kaufman, Rasinski, Lee, y West, 1991; Kohn, 1977; Marsh, 1996; Marsh y Grayson, 1995; Salgado e Iglesias, 1995; Tomás y Oliver, 1999, 2004; Tomás et al. 2012; Wang, Siegal, Falck, y Carlson, 2001). Todos estos trabajos han empleado modelos factoriales para considerar los efectos de método asociados a los ítems invertidos. Otro rasgo común ha sido que la mayoría de estos trabajos han estudiado una escala concreta, siguiendo los pasos de Carmines y Zeller: la escala de autoestima de Rosenberg (RSES, 1965). La escala, pensada para medir el aspecto de autoaceptación de la autoestima (Crandall, 1973), es un instrumento en principio unidimensional, elaborado para medir un factor de autoestima global. Rosenberg (1965, 1989) la presentó originariamente con 10 ítems, 5 positivos y 5 invertidos, como un escalograma de Gutmann. No obstante, prácticamente todas las investigaciones y estudios posteriores la han tratado como si de escalamiento Likert se tratara.

Como hemos visto entre los estudios que han empleado la escala de Rosenberg destacan, en función del objetivo de la presente investigación, aquellos que han analizado los efectos de método asociados a los ítems invertidos. En la actualidad, el estudio de este tipo de efectos de método se aborda desde un diseño de investigación, una metodología y un análisis de datos específico. Es decir, presenta un contexto de investigación bastante definido. Este contexto de investigación para el estudio de efectos de método coincide con el del análisis de la validez convergente-discriminante, que se traduce en diversos tipos de matrices multirrasgo-multimétodo (multitrait-multimethod matrices, MTMM). Estas matrices se han analizado desde diversos modelos estadísticos, pero el Análisis Factorial Confirmatorio (Confirmatory Factor Analysis, CFA) es hoy por hoy el modelo estadístico de referencia (Millsap, 1995; Schmitt y Stults, 1986; Wothke, 1996).

Entre los modelos CFA empleados para ello, destacan el CFA con rasgos y unicidades correlacionados (correlated traits and correlated

uniqueness, CFA-CTCU; Marsh, 1988, 1989) y el CFA con rasgos y métodos correlacionados (correlated traits and correlated methods, CFA-CTCM). En el modelo CFA-CTCU el factor de método no se modela explícitamente, sino que se infiere correlacionando las unicidades de todos los indicadores de un mismo método. Por tanto, no conocemos el porcentaje de varianza asociado al efecto de método “inferido”. El modelo CFA-CTCM, por su parte, plantea la hipótesis de que la varianza de las variables se puede describir como una combinación lineal de los rasgos, el método o métodos, y los efectos de error o unicidad (Jöreskog, 1974). Este modelo dispone lo siguiente: (a) una explicación de la matriz multirrasgo-multimétodo en términos de los factores subyacentes, en lugar de las variables observadas; (b) la evaluación de la validez convergente y discriminante en la matriz, así como la estimación de la cuantía de los parámetros, (c) las pruebas de hipótesis relacionadas con esos parámetros que definen la validez convergente y discriminante, (d) estimaciones bien definidas de la varianza debido a rasgos, métodos y unicidades, y (e) estimaciones (desatenuadas) de las correlaciones tanto para métodos como para factores de rasgo (Byrne y Goffin, 1993). Este modelo se ha mostrado claramente superior en estudios de simulación para el caso de varias variables observables por combinación de rasgo-método, lo que es el caso en el estudio de efecto de método de la escala de Rosenberg (Tomás, Hontangas y Oliver, 2000). El modelo CFA-CTCM se muestra predominante en la literatura actual, pues es el más utilizado para estudiar la estructura factorial de las medidas de autoinforme, concretamente la estructura de la Escala de Autoestima de Rosenberg, (Di Stefano y Motl, 2009a, 2009b; Horan, DiStefano y Motl, 2003; Quilty, Oakman y Risko, 2006; Tomás y Oliver, 1999; TsuruMaki, Sato, y Nihei, 2009; Wang et al., 2001). Frente a éste, el CFA-CTCU no siempre ofrece una clara interpretación de las unicidades correlacionadas como estimación de los efectos de método, asumiendo además que los métodos no están correlacionados entre sí (Bagozzi, 1993; Kenny y Kashy, 1992).

Por otra parte, Greenberger, Chen, Dmitrieva y Farrugia (2003), si bien también analizaron si al invertir ítems la escala de Rosenberg se produce un artefacto metodológico, presentaron rasgos distintivos. Este trabajo es, hasta lo que conocemos, el único en que se han presentado diversas versiones de la escala de autoestima de Rosenberg con el objetivo de valorar el posible efecto de método asociado a los ítems invertidos. Para ello elaboraron dos nuevas versiones de la escala, una con todos los ítems formulados en el sentido de la autoestima (en positivo) y otra con todos los ítems en el sentido contrario a la autoestima (invertidos), además de la escala original de Rosenberg con ítems de ambos tipos balanceados.

Administraron una de estas tres versiones a cada uno de los participantes en el estudio, 741 estudiantes universitarios de una universidad pública californiana. Los autores concluyen que en la escala original emergía una estructura de dos factores, mientras que en las otras dos versiones el mejor modelo era el unifactorial, un solo factor de autoestima. Además no encontraron diferencias estadísticamente significativas entre las medias de las escalas, lo que implicaba ausencia de sesgo de aquiescencia, y las tres versiones no correlacionaban de forma diferente con los criterios empleados.

Esta investigación se basa en las anteriormente recogidas, tanto en el aspecto sustantivo que nos ocupa, los efectos de método, como en las técnicas analíticas que emplea. El presente trabajo amplía los anteriores estudios, y en concreto el de Greenberger et al. (2003) para intentar dar nueva luz sobre la naturaleza de los efectos de método como artefacto y sus posibles soluciones. Se emplean dos versiones de la escala de autoestima de Rosenberg (1965), una en su formato original con ítems positivos e invertidos y otra transformada de modo que todos los ítems sean positivos. No se emplea la tercera versión en negativo dado que se consideró de escaso valor aplicado construir una escala con todos los ítems en el sentido que no pretende medir la escala. Estas dos escalas se presentan a todos los participantes del estudio, de forma que se tiene la posibilidad de estudiar todos los ítems conjuntamente y de comparar ambas escalas en los mismos participantes, frente a los grupos independientes utilizados por Greenberger et al. (2003). Además, se considera explícitamente el efecto de método como un factor independiente al rasgo. Y finalmente, se extiende el trabajo de estos autores al emplear un idioma diferente, el castellano, y evaluar si los efectos encontrados son, o no, los mismos.

MÉTODO

Diseño y participantes. El diseño es correlacional y de encuesta, con una muestra de carácter incidental, compuesta por 390 estudiantes de instituto y universidad de la ciudad de Valencia. El 50.8% eran estudiantes de secundaria, 17.9% estudiantes de Bachillerato, el 15.9% estudiantes del grado de Psicología y el 15.4% estudiantes de grado de Fisioterapia. La media de edad fue 16.82 (DT= 3.69) y el 56.7% eran mujeres.

Instrumentos. Se utiliza la Escala de Autoestima de Rosenberg (1965), en el marco de una batería más amplia que incluye factores actitudinales y de personalidad, así como características sociodemográficas.

Los participantes respondieron a dos versiones de ésta. La primera es la traducción al castellano de la versión de 10 ítems original, con escalamiento Likert que incluía cinco ítems en la dirección positiva de la autoestima (ítems 1, 2, 4, 6, y 7), y cinco en el sentido opuesto o invertidos (3, 5, 8, 9, y 10). Adicionalmente, todos los participantes han contestado a otra versión en que los ítems invertidos se pasaron a formato positivo. Es decir, los ítems invertidos se parafrasearon en positivo (ver Tabla 2). Las dos versiones de la escala se colocaron en lugares de la batería no consecutivos y se contrabalancearon para que la mitad de participantes contestara una versión primero y el resto la otra.

Análisis estadísticos. Todos los modelos estimados son modelos de análisis factorial confirmatorio. Dentro del análisis factorial confirmatorio, parte de los modelos se modelan mediante el CTCM (Correlated Traits Correlated Methods) explicado en la introducción. En estos modelos CTCM se modela uno o dos factores de rasgo y un factor de método asociado a los ítems invertidos y/o originalmente invertidos. La explicación detallada de todos los modelos factoriales confirmatorios estimados se presenta en la sección de resultados y en la figura 1. Los modelos se han estimado en EQS versión 6.1 (Bentler, 1995), y al tratarse de medidas de carácter ordinal, mediante el cálculo de correlaciones policóricas y estadísticos robustos de Satorra-Bentler.

La plausibilidad de los modelos, su ajuste, se ha evaluado en base a distintos índices y estadísticos de diversas familias, tal y como se recomienda en la literatura (Hu & Bentler, 1999; Tanaka, 1993). En concreto, y dado que se empleaban correlaciones policóricas y estimaciones robustas, se emplearon: (a) el estadístico χ^2 robusto de Satorra y Bentler (Kline, 1998; Ullman, 1996); (b) tres índices robustos de comparación con modelo nulo, el NFI (Normed Fit Index), el NNFI (Non-Normed Fit Index) y el CFI (Comparative Fit Index) (Bentler, 1990) cuyos valores de .9 e idealmente más de .95 son indicativo de excelente ajuste (Hu, & Bentler, 1999); (c) la raíz del error cuadrático medio (RMSEA) con valores de menos de .08 indicando buen ajuste y de menos de .05 como ajuste excelente (Steiger, & Lind, 1980); Adicionalmente, y dado que se presenta una secuencia de modelos alternativos, se empleará el Criterio Informativo de Akaike (AIC, Burnham, y Anderson, 1998), especialmente diseñado para comparar modelos, donde los valores menores se asocian a los modelos más adecuados.

RESULTADOS

En primer lugar, se puso a prueba si existía un efecto de aquiescencia operando en la escala. De ser así, las medias de ambas versiones debieran diferir significativamente, con valores más altos en la escala con todos los ítems en el sentido de la autoestima. Esta comparación mediante pruebas t relacionadas se ha efectuado tanto sobre las puntuaciones extraídas mediante la suma de ítems, como sobre las puntuaciones factoriales del factor de autoestima de ambas escalas, y con resultados no significativos en ambos casos: $t_{371} = 1.21, p = .226$ y $t_{321} = -1.61, p = .108$. Al poder descartarse un efecto de aquiescencia se pasó a la prueba de los modelos que evalúan el potencial efecto de método asociado a los ítems invertidos.

Los modelos propuestos para las dos versiones de la escala de autoestima de Rosenberg se muestran en la Figura 1. Los modelos planteados son alternativos y su lógica se detalla a continuación:

- *Modelo 1.* Especifica una solución de un solo factor de autoestima subyacente a los 20 ítems. Esto es, a la versión original y a la versión con todos los ítems en sentido positivo. Este modelo pone a prueba si es razonable pensar que una única dimensión (autoestima) subyace a la contestación de los ítems. Dado que los ítems en posición 1, 2, 4, 6 y 7 de la escala de Rosenberg son idénticos de formulación en ambas escalas, se han liberado las covarianzas entre los errores por pares entre estas dos versiones. Estas covarianzas de error se mantienen libres a estimar para todos los modelos propuestos.
- *Modelo 2.* Se especifican dos factores, uno para los 10 ítems originales y otro para la versión de ítems en positivo. Permite detectar posibles efectos de orden y/o de conversión de los ítems negativos en positivos. Si la posición dentro de la batería de test y/o el cambio en los ítems invertidos produjera un cambio importante en las respuestas de los sujetos, este modelo podría ajustar mejor que el de un solo factor. Los factores se asumen en todo caso correlacionados. Este modelo permitirá evaluar la validez convergente entre ambas medidas.
- *Modelo 3.* Presenta un solo factor de rasgo (autoestima) que explica las puntuaciones de todos los ítems, junto a un factor de método asociado a los 5 ítems invertidos de la escala original. Es un CTCM con un factor de rasgo y uno de método. En definitiva, pretende detectar el impacto de la formulación negativa de los ítems, efecto de método repetidamente encontrado en la literatura.
- *Modelo 4.* Se especifica un factor de rasgo asociado a todos los ítems y un factor de método asociado a los ítems invertidos de la escala original

(3, 5, 8, 9 y 10) y a los mismos ítems parafraseados para ser positivos. Pretende detectar si existe un rasgo común a estos ítems y por tanto descartar un efecto de método asociado a la inversión. De resultar en un ajuste inadecuado este modelo permitiría descartar que los ítems en la escala original covaríen debido a rasgo y no a su inversión, tal y como se defiende en las soluciones de dos factores de rasgo, autoestima positiva y negativa.

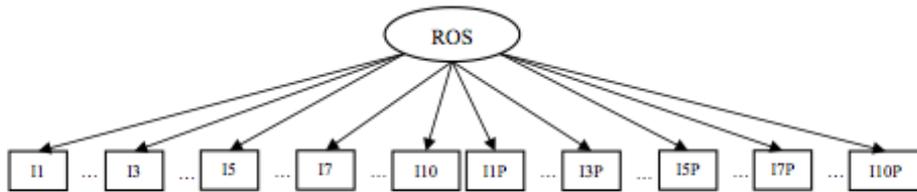
- *Modelo 5.* Especifica la misma estructura que el modelo 2, salvo que añaden un efecto de método asociado a los cinco ítems invertidos de la escala original.
- *Modelo 6.* Especifica la misma estructura que el modelo 5, pero con un factor de método asociado a los 5 ítems invertidos en la escala original y a estos mismos ítems parafraseados en positivo, con la misma lógica que el modelo 4.

Los resultados de ajuste de los distintos modelos se presentan en la tabla 1. El modelo más simple, el de un solo factor a la base de las covarianzas de los 20 ítems (originales y parafraseados en positivo) presenta un ajuste razonable a la luz de los diversos índices empleados. Sin embargo, es el de peor ajuste de todos los modelos competitivos presentados, como puede verse en los valores de los AIC. El modelo 2, que supone que las escalas dan lugar a dos factores de autoestima relacionados, mejora el ajuste, si bien la correlación entre ambos factores de autoestima es muy elevada ($r = .84$). Esto por lo que respecta a los modelos que no modelan efectos de método.

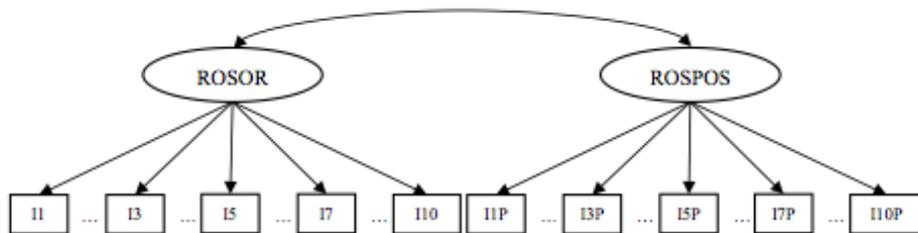
Los modelos que, además de factores de rasgo, modelan un factor de método son los numerados 3 al 6. Todos ellos mejoran a sus correspondientes modelos solo de rasgo en ajuste global. Los modelos 3 y 5 son los que modelan el efecto de método *solamente* en los ítems invertidos, es decir en los ítems 3, 5, 8, 9 y 10 de la escala original. Ambos modelos mejoran a sus correspondientes modelos sólo de rasgo, y en todos los casos las saturaciones factoriales de método son estadísticamente significativas ($p < .05$). Las saturaciones de estos dos modelos, tanto de rasgo como de método, pueden verse en la tabla 2. Al considerar ambos ajustes analíticos hay diferencias apreciables entre las saturaciones de rasgo del modelo 3, promedio de .58, frente a una media de rasgo en el modelo 5 de .62, lo que concuerda con la mejora del ajuste global del modelo 5. Estas diferencias entre saturaciones, sin embargo, son muy bajas entre ambos modelos en saturaciones de método: media de .39 para el modelo 3 y de .38 para el modelo 5. Adicionalmente, es importante resaltar que la correlación entre

los dos factores de autoestima del modelo 5, pese a modelar los efectos de método, sigue siendo positiva y muy elevada ($r = .803$).

Modelo 1



Modelo 2



Modelo 3

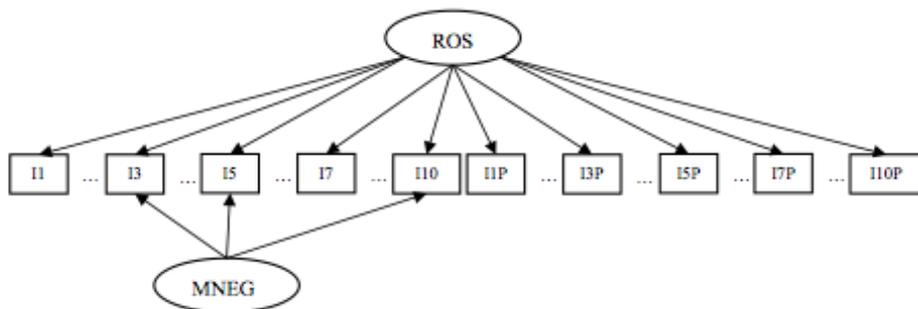
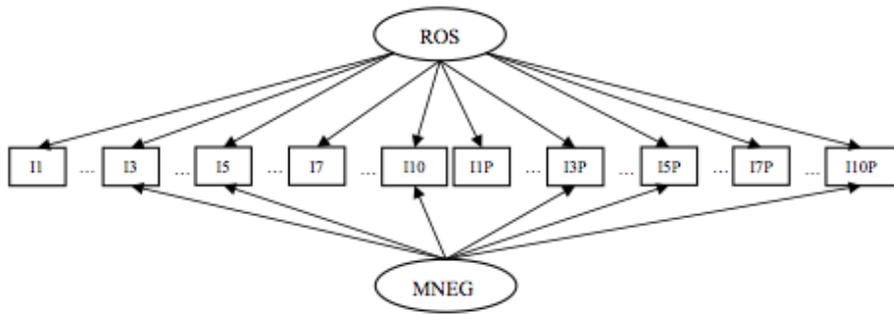
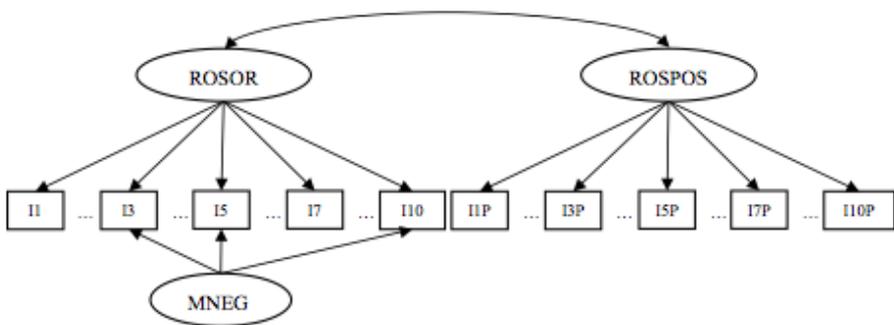


Figura 1. Modelos estructurales puestos a prueba.

Modelo 4



Modelo 5



Modelo 6

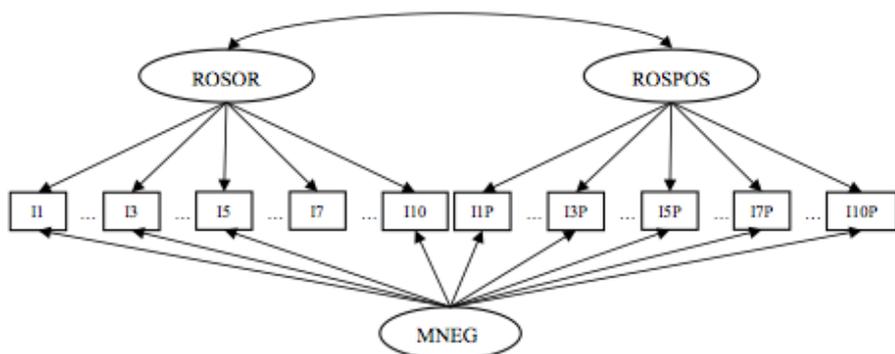


Figura 1 (continuación).

Tabla 1. Ajuste de los modelos teóricos puestos a prueba.

	NFI	NNFI	CFI	RMSEA	AIC	χ^2	χ^2_{rob}	gl	<i>p</i>
<i>Modelo 1</i>	.931	.940	.948	.093	290.92	1106.04	620.92	160	<.01
<i>Modelo 2</i>	.957	.971	.975	.065	58.46	609.61	386.46	164	<.01
<i>Modelo 3</i>	.939	.947	.956	.087	227.72	982.43	547.62	160	<.01
<i>Modelo 4</i>	.941	.948	.957	.087	217.68	947.27	517.68	155	<.01
<i>Modelo 5</i>	.963	.977	.981	.057	9.59	519.32	327.59	159	<.01
<i>Modelo 6</i>	.964	.977	.981	.058	9.69	505.37	317.69	154	<.01

Para finalizar, resulta relevante analizar lo que ocurre en los modelos 4 y 6 frente a los 3 y 5, respectivamente. Los modelos 4 y 6 modelan un factor de rasgo a los ítems 3, 5, 8, 9, y 10 de ambas escalas, esto es a los ítems invertidos de la escala original y a los parafraseados para ser en positivo. Si estos modelos supusieran un mejor ajuste, sería indicativo de que pesa más lo que los ítems tienen en común, el contenido, que su formulación, dado que unos son invertidos y otros positivos. Sin embargo, estos modelos no ajustan mejor que los modelos más parsimoniosos. De hecho, si se entra en el detalle de las saturaciones de los ítems parafraseados, ninguna de ellas es estadísticamente significativa, en ninguno de ambos modelos.

CONCLUSIONES Y DISCUSIÓN

En este trabajo se han utilizado dos versiones de la escala de autoestima de Rosenberg (una con ítems positivos e invertidos, y otra sólo con ítems positivos) para clarificar el efecto de la formulación invertida de los ítems, utilizando una secuencia de modelos de medida. Una primera conclusión que nos ha permitido contestar el diseño de investigación es si se apreciaba un sesgo debido a la aquiescencia. Los resultados muestran que no opera en esta escala y para esta muestra. Este resultado es coincidente con el encontrado por Greenberger et al. (2003), que al pasar tres versiones de la escala de Rosenberg (original, todos los ítems en positivos, todos en invertido) a tres muestras de estudiantes universitarios tampoco encontraron diferencias en las medias de las tres versiones. Estos resultados no coinciden con los hallados por otros autores (Ferrando, Lorenzo-Seva, y Chico, 2003; Ferrando y Lorenzo-Seva, 2010). Sin embargo, los presentes hallazgos, tomados conjuntamente, irían en línea con la argumentación de

Tabla 2. Saturaciones factoriales.

Ítem	λ		$\lambda_{\text{método}}$	
	Modelo3	Modelo5	Modelo3	Modelo5
1. Siento que soy una persona de valor, por lo menos en un plano de igualdad con los demás.	.642	.679		
2. Siento que tengo una serie de buenas cualidades.	.646	.715		
3. En definitiva, me inclino a pensar que soy un fracasado.	.567	.673	.399	.234
4. Soy capaz de hacer las cosas tan bien como la mayoría de la gente.	.579	.691		
5. Siento que no tengo mucho de lo que estar orgulloso.	.486	.549	.384	.240
6. Tengo una actitud positiva hacia mí mismo.	.678	.720		
7. En general, estoy satisfecho conmigo mismo.	.725	.803	.278	.300
8. Me gustaría poder tenerme más respeto.	.448	.411		
9. A veces me siento inútil.	.447	.453	.609	.737
10. A veces pienso que no soy bueno en todo.	.087	.076	.298	.340
1. Siento que soy una persona de valor, por lo menos en un plano de igualdad con los demás.	.700	.711		
2. Siento que tengo una serie de buenas cualidades.	.719	.749		
3. En definitiva, me inclino a pensar que soy una persona de éxito.	.586	.596		
4. Soy capaz de hacer las cosas tan bien como la mayoría de la gente.	.671	.721		
5. Siento que tengo mucho de lo que estar orgulloso.	.683	.715		
6. Tengo una actitud positiva hacia mí mismo.	.733	.722		
7. En general, estoy satisfecho conmigo mismo.	.762	.760		
8. Me tengo respeto a mí mismo.	.695	.702		
9. Me siento útil.	.718	.728		
10. A veces pienso que soy bueno en todo.	.209	.176		

Roer (1965), basada en el examen de un gran número de estudios sobre estilos de respuesta, de que éstos son específicos de los tests o escalas que se pasan, frente a otros autores que arguyen que es una disposición de los participantes (por ejemplo, Cronbach, 1942, 1950).

En cuanto a los resultados que propiamente tienen que ver con el objetivo del trabajo, discernir la presencia de efectos de método asociado a los ítems invertidos, la evidencia indica que la conversión de los ítems invertidos en positivos no altera las cualidades psicométricas de la escala, salvo en que desaparecen los efectos de método. Al parecer la formulación de ítems en sentido positivo frente a invertido no supone una diferencia sustantiva respecto al constructo medido por el instrumento; es decir, una escala con ítems positivos e invertidos y otra compuesta sólo por ítems positivos, con el mismo contenido, se pueden considerar equivalentes respecto al constructo evaluado. Es cierto que los modelos de mejor ajuste son los de dos factores de autoestima, uno asociado a la escala original, y otro a la formulada toda en positivo, pero la correlación tan elevada entre ambos factores hace muy improbable que sean constructos sustancialmente diferentes.

En segundo lugar, se constata claramente la existencia de un efecto de método asociado a los ítems invertidos que, al parecer, es independiente de su contenido (autoestima), puesto que no desaparece cuando se emplean los mismos ítems en sentido positivo, tal como ha quedado patente en los resultados de las versiones de los modelos 3 y 4, por un lado, y 5 y 6, por otro. Este efecto de método, que seguimos detectando, probablemente sea debido a que los enunciados invertidos evoquen una representación cognitiva o afectiva distinta de los ítems positivos; no obstante, la naturaleza del fenómeno está todavía por determinar. Una forma de avanzar en esta línea sería intentar buscar variables relacionadas con este efecto. En consecuencia, en lugar de limitarse a constatar su existencia, o además de esto, las futuras investigaciones deberían considerar algún aspecto relacionado con el procesamiento cognitivo y/o afectivo de los enunciados por parte de los sujetos para explicar qué está afectando a este tipo de ítems.

Un tercer resultado, no esperado pero relevante, es que los modelos con dos factores de autoestima (véase los modelos 2, 5 y 6) tengan mejor ajuste que los equivalentes que solamente presentan un factor de autoestima asociado a los 20 ítems de ambas versiones de la escala de Rosenberg. Aunque como se ha visto, la correlación entre estos dos factores de autoestima es muy elevada, no es menos cierto que la presencia de dos factores separados mejora el ajuste y debe, por tanto, tomarse en consideración. Al estar las dos escalas separadas físicamente dentro de la

encuesta, y pese a estar presentadas físicamente igual y ocupando cada una de ellas una sola página y contrabalanceadas, parece claro que existe una mayor congruencia (covariación) entre los ítems que se encuentran físicamente juntos. Este resultado parece ser un efecto conocido de orden, la tendencia a contestar de forma similar o congruente conjuntos de ítems que miden lo mismo y están cercanos (Sudman, Bradburn y Schwarz, 1996) o la tendencia de los informantes a ser consistentes o estar en línea con las respuestas de ítems inmediatamente anteriores (Schuman y Preser, 1996). Algún efecto similar podría estar a la base de este resultado, abriendo la posibilidad de futuras investigaciones que traten de valorar este punto.

ABSTRACT

Evidence gathered of method effects associated to negatively worded items. Negatively worded items have been long used in the measurement of psychological constructs, since almost the beginning of the psychometric practice. However, this strategy interferes in the examination of the scales' latent structures, as it has been shown in several studies. This paper aims to provide new evidence on this topic, in order to clarify the negatively worded items effect, by using two versions of the Rosenberg' Self-Esteem Scale (1965), one in its original form, with positive and negatively worded items, and the other one in a transformed version, in which all the items are positive, in order to test if, in this latest version, the method effect associated to negatively worded items still remains. The sample was composed by 390 high-school and university students from Valencia, Spain. Several alternative (including CTCM) structural equation models have been tested, and their fit indices are compared. Results offer empiric evidence on the method effect associated to negatively worded items. This effect disappears when items are worded in a positive way and, thus, Rosenberg' Self-Esteem Scale structure is composed by just one factor: the self-esteem factor. Implications of these results are pointed out in the discussion.

REFERENCIAS

- Bachman, J. G., y O'Malley, P. M. (1986). Self-concept, self-esteem, and educational experiences: The frog pond revisited (again). *Journal of Personality and Social Psychology*, 50(1), 35-46.
- Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measures of self-esteem. *Journal of Research in Personality*, 27, 49-87.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M. (1995). *EQS program manual*. Encino, CA: Multivariate Software.
- Billiet, J. B., & McClendon, M. J. (2000). Modelling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628.

- Billiet, J. B., & McClendon, M. K. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628.
- Burnham, K. P., y D. R. Anderson (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer-Verlag.
- Byrne, B., y Goffin, R. (1993). Modeling multitrait-multimethod data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research*, 28, 67-96.
- Carmine, E. G., y Zeller, R. A. (1974). On establishing the empirical dimensionality of theoretical terms: An analytical example. *Political Methodology*, 1, 75-96.
- Carmine, E. G., y Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Corwyn, R. F. (2000). The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality*, 34(4), 357-379.
- Crandall, R. (1973). The measurement of self-esteem and related constructs. En J. P. Robinson y P. R. Shaver (Eds.), *Measures of social psychological attitudes* (2nd ed., pp. 45-167). Ann Arbor, MI: Institute for Social Research.
- Crawford, J. R. y Henry, J. D. (2003). The Depression Anxiety Stress Scales: Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, 42, 111-131.
- Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology*, 33, 401-415.
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3-31.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- DiStefano, C., y Motl, R. W. (2009a). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*, 46(3), 309-313.
- DiStefano, C., y Motl, R. W. (2009b). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling*, 16, 134-146.
- Ferrando, P. J., & Condon, L. (2006). Assessing acquiescence in binary responses: IRT-related item-factor analytic procedures. *Structural Equation Modeling*, 13, 96-115.
- Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, 63, 427-448.
- Ferrando, P.J., Lorenzo-Seva, U., y Chico, E. (2003). Unrestricted factor analytic procedures for assessing acquiescent responding in balanced, theoretically unidimensional personality scales. *Multivariate Behavioral Research*, 38, 353-374.
- Goldsmith, R. E. (1986). Dimensionality of the Rosenberg Self-Esteem Scale. *Journal of Behavior and Personality*, 1(2), 253-264.
- Greenberger, E., Chen, C., Dmitrieva, J., Farrugia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter? *Personality and Individual Differences*, 35, 1241-1254.
- Hensley, W. E., y Roberts, M. K. (1976). Dimensions of Rosenberg's Self-Esteem Scale. *Psychological Reports*, 38, 583-584.
- Horan, P. M., DiStefano, C., y Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, 10(3), 435-455.

- Hu, L., y Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. En R. C. Atkinson, D. H. Krantz, R. D. Luce, y P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1-56). San Francisco: Freeman.
- Kaufman, P., Rasinski, K. A., Lee, R., y West, J. (1991). *National Education Longitudinal Study of 1988. Quality of the responses of eight-grade students in NELS88*. Washington, DC: U.S. Department of Education.
- Kenny, D. A. y Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- Kohn, M. L. (1977). *Class and conformity: A study of values* (2nd ed.). Chicago: University of Chicago Press.
- Marsh, H. W. (1988). *Self-description Questionnaire-I*. San Antonio, USA: The Psychological Corporation.
- Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to Early Adulthood. *Journal of Educational Psychology*, 81(3), 417-430.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, 70(4), 810-819.
- Marsh, H. W., y Grayson, D. (1995). Latent variable models of multitrait-multimethod data. En R. H. Hoyle (Ed.), *Structural equation modeling: Concept, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
- Millsap, R. E. (1995). The statistical analysis of method effects in multitrait-multimethod data: A review. En P. E. Shrout y T. Fiske (Eds.), *Personality research, methods and theory: A festschrift honoring Donald W. Fiske*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Morales, P. (2006). *Medición de actitudes en psicología y educación. Construcción de escalas y problemas metodológicos*. Madrid: Universidad Pontificia Comillas.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Quilty, L. C., Oakman, J. M., y Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling*, 13(1), 99-117.
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63, 129-156.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. New Jersey, USA: Princeton University Press.
- Rosenberg, M. (1989). *Society and the adolescent self-image* (Rev. ed.). Middletown, CT: Wesleyan University Press.
- Salgado, J. F., e Iglesias, M. (1995). Estructura factorial de la escala de autoestima de Rosenberg: Un análisis factorial confirmatorio. *Psicológica*, 16, 441-454.
- Satorra, A. y Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schmitt, N. y Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, 10, 1-22.
- Steiger, J. H., y Lind, C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. En K. A. Bollen (Ed.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.
- Tomás, J. M., Hontangas, P. M. y Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research*, 35, 469-500
- Tomás, J. M., Sancho, P., Oliver, A., Galiana, L., Meléndez, J. C. (2012). Efecto de método asociado a ítems invertidos vs. ítems en negativo. *Revista Mexicana de Psicología*, 29, 105-115.
- Tomás, J. M., y Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6(1), 84-98.
- Tomás, J. M., y Oliver, A. (2004). Análisis psicométrico confirmatorio de una medida multidimensional del autoconcepto en español. *Interamerican Journal of Psychology*, 38(2), 285-293.
- TsuruMaki, M., Sato, T. y Nihei, Y. (2009). The effect of negatively worded measures of self-esteem on children. *Social Behavior and Personality*, 37(10), 1383-1384.
- Ullman, J. B. (1996). Structural equation modeling. En B. Tabachnick y L. Fidell (Eds.), *Using multivariate statistics* (3ª ed., pp. 709-812). New York: HarperCollins.
- Wang, J., Siegal, H. A., Falck, R. S., y Carlson, R. G. (2001). Factorial structure of Rosenberg's self-esteem scale among crack-cocaine drug users. *Structural Equation Modeling*, 8, 275-286.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, 34, 105-121.
- Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. En G. A. Marcoulides y R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

(Manuscrito recibido: 20 Julio 2012; aceptado: 10 Diciembre 2012)