

Efectos de método en las escalas de Ryff: Un estudio en población de personas mayores

José Manuel Tomas^{*}, Juan Carlos Meléndez, Amparo Oliver,
Esperanza Navarro y Gustavo Zaragoza

Universitat de València, Spain

A pesar de su amplio uso, las escalas de Ryff han presentado problemas de validez factorial y en recientes investigaciones han mostrado efectos de método que, sin embargo, no han sido exhaustivamente investigados en la literatura, especialmente en poblaciones especiales, como la de mayores, estudiada en este trabajo. El objetivo de este trabajo es doble: estudiar la estructura factorial de la escala de Ryff en sus distintas versiones; y evaluar la presencia de efectos de método asociados a los ítems formulados en negativo. Para ello se usa una muestra de 556 personas mayores de 65 años no institucionalizadas de la Comunidad Valenciana. Se han probado de forma exhaustiva todos los modelos factoriales confirmatorios encontrados en la literatura para la escala, mediante análisis factorial confirmatorio estimado por máxima verosimilitud con correcciones robustas, y el de mejor ajuste fue el de seis factores original de los autores de la escala. Respecto a los factores de método, los resultados muestran un efecto consistente en todas las versiones del factor asociado a ítems negativos, que al ser contemplado en el modelo, ofrece una mejora del ajuste.

La varianza de método se ha definido tradicionalmente como una forma de error sistemático o contaminación, debido al método de medida y no al rasgo objeto de la medición (Campbell y Fiske, 1959). Por lo tanto, en una medida observada X , se pueden diferenciar tres componentes:

$$X = R + M + E \text{ donde}$$

R es el rasgo que se pretende medir (puntuación verdadera en teoría clásica de tests), M es el componente de método y E sería el error aleatorio

* Correspondencia: José Manuel Tomas Miguel, Departamento de Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Universitat de València, Av. Blasco Ibáñez, 21, 46010, Valencia (Spain), E-mail: tomasjm@uv.es

Asumiendo que los tres componentes son independientes, la varianza de la medida observada se descompone de la siguiente forma:

$$V(X) = V(R) + V(M) + V(E)$$

Esta varianza de método resulta preocupante, dado que puede distorsionar, y de hecho lo hace, las medidas de asociación, como las correlaciones, dejando de reflejar las verdaderas relaciones entre las variables de interés. Esta distorsión se conoce como sesgo de método (Conway, 2002), y para formularlo matemáticamente en términos de correlaciones, en este contexto, la relación entre dos variables es función de la correlación verdadera entre los rasgos, la correlación entre los métodos y la proporción de varianza de rasgo y de método en cada medida. De este modo, para dos variables observadas X e Y (Cote y Buckley, 1988) se cumple:

$$r_{xy} = [\rho_{ii,tj} (\sqrt{t_x} \cdot \sqrt{t_y})] + [\rho_{mk,ml} (\sqrt{m_x} \cdot \sqrt{m_y})]$$

donde r_{xy} es la correlación observada entre las variables de interés y, $\rho_{ii,tj}$ y $\rho_{mk,ml}$ son las correlaciones verdaderas entre los rasgos y los métodos, respectivamente, y t_x , t_y , m_x , y m_y son las proporciones de varianza de rasgo y de método en las medidas X e Y respectivamente.

La fórmula muestra claramente que si los métodos se encuentran correlacionados el segundo sumando es mayor de cero, inflando las correlaciones. Sin embargo, resulta también obvio que aún siendo no correlacionados, con el segundo sumando cero, la correlación sigue sin reflejar la relación verdadera, ya que entonces la varianza de método actúa como error aleatorio, atenuando las correlaciones (por ejemplo, Bollen, 1989).

Uno de los potenciales sesgos de método que se han propuesto, y han sido evaluados en la literatura, consiste en la aparición de factores de método asociados a los ítems formulados en negativo. La práctica de utilizar tanto ítems formulados en positivo como en negativo, para evitar el sesgo de aquiescencia tiene una larga tradición en psicología y otras ciencias sociales, y pueden trazarse sus inicios en los estudios sobre personalidad de los años 50 (Bass, 1955; Jackson y Messick, 1957; Leavitt, Hax, y Roche, 1955; Messick y Jackson, 1957). De hecho esta tradición ha dado lugar a una regla usada habitualmente, según la cual se escogen aproximadamente la mitad de ítems en positivo y la mitad en negativo

(Nunnally, 1978). Sorprendentemente, un formato de pregunta originalmente pensado para evitar un sesgo de respuesta, parece que produce varianza de método asociada a la aparición de factores de método en los ítems en negativo, teniendo entre otros efectos el de complicar los análisis factoriales de las escalas, problema éste que subrayamos no por ser el más relevante, sino por ser el objeto de este estudio.

Efectivamente, existe abundante evidencia en la literatura científica de que el uso de ítems formulados en negativo genera varianza de método, lo que hace aparecer factores de método negativos no asociados con los rasgos de interés (Williams, Ford y Nguyen, 2002). Este efecto se observa en distintos campos aplicados de la psicología. Por ejemplo, en el contexto de la psicología organizacional se ha encontrado este efecto de método en muchas de las escalas más ampliamente utilizadas, como la *Job Diagnostic Survey* (Cordery y Sevastos, 1993; Harvey, Billings y Nilan, 1985; Idaszak y Drasgow, 1987; Kulik, Oldham y Langner, 1988). De la misma forma, en diversas medidas de compromiso con la organización se ha encontrado este efecto (Ladd, Gordon, Beauvais, and Morgan, 1982; Magazine, Williams y Williams, 1996). Sin ánimo de ser exhaustivos, también se han encontrado en medidas de liderazgo (por ejemplo, Schriesheim y Eisenbach, 1995). Estos efectos de método también se han encontrado en escalas de personalidad y, en general, de auto-evaluaciones personales, como la que es objeto de análisis en este trabajo, así, se han encontrado en medidas de autoestima, de personalidad o de orientación vital (Carmines y Zeller, 1979; Di Stefano y Motl, 2006, 2008; Horan, DiStefano y Motl, 2003; Marsh, 1996; Meharabian y Ljunggren, 1997; Motl y DiStefano, 2002; Quilty, Oakman y Risko, 2006; Scheirer, Carver y Bridges, 1994; Tomás y Oliver, 1999; Vautier, Raufaste y Cariou, 2003; Wang, Siegal, Falck y Carlson, 2001).

Diversos de estos autores apuntan también a que la aparición de estos efectos de método es diferencialmente relevante en función de ciertas características personales, incluyendo el grado en que se cae en la deseabilidad social, el nivel educativo, etc. Independientemente de los factores causantes, resulta evidente que existe un amplio consenso en la literatura de que la formulación de los ítems, y muy especialmente la formulación en negativo y/o invertido de éstos, puede llevar a la aparición de factores de método, con la consecuencia de distorsiones en la estructura factorial de las escalas analizadas. Uno de los dominios donde es potencialmente posible que aparezcan estos efectos asociados a los ítems negativos es en las escalas de bienestar, que son casi exclusivamente medidas de autoinforme, de evaluación personal, percepciones de uno

mismo, y que prácticamente siempre se han medido con combinaciones de ítems positivos y negativos.

De entre los cuestionarios empleados para la medición del bienestar psicológico uno de los de más amplio uso es el desarrollado por Ryff desde 1989 (Van Dierendonck, 2005). El cuestionario original contaba con 120 ítems, balanceados en cada una de las seis dimensiones en que la teoría subyacente dividía el bienestar: autoaceptación, relaciones positivas con los otros, dominio del ambiente, autonomía, crecimiento personal y propósito en la vida. Dada la longitud de dicho instrumento, se han venido proponiendo diferentes versiones de menor tamaño. Así, Ryff, Lee, Essex y Schmutte (1994) desarrollaron una versión de 14 ítems por escala, seleccionados en función de sus correlaciones ítem-total y su ajuste al modelo teórico propuesto. Otra versión acortada incluye 9 ítems por factor (Van Dierendonck, 2005), y además, Ryff y Keyes (1995) desarrollaron una versión de 3 ítems por escala, seleccionando aquéllos que mejor ajustaban con el modelo teórico propuesto de seis factores.

En cuanto al funcionamiento de las escalas, los autores han indicado que la validez factorial de la versión de 14 ítems por dimensión muestra un mal ajuste al modelo teórico propuesto; que la versión de 9 ítems ofrece ajustes factoriales confirmatorios más bajos de lo deseable; y que la versión de 3 ítems muestra unos indicadores de ajuste ligeramente superiores, pero muy baja consistencia interna debido, probablemente, a su longitud y al criterio de selección de ítems (Keyes, Shmotkin y Ryff, 2002).

En relación a la estructura factorial de las Escalas de Ryff, éstas han sido analizadas mediante análisis factoriales confirmatorios en su versión de 54 ítems por Ryff y Keyes (1995), encontrando que el modelo confirmatorio de mejor ajuste era el de seis factores de primer orden y uno de segundo orden que explicaría a los anteriores, aunque el ajuste resultó pobre. Posteriormente Keyes, Shmotkin y Ryff (2002), también mediante análisis confirmatorios, encontraron que a la base de estos seis factores de primer orden se encontrarían dos constructos de nivel superior, bienestar subjetivo y bienestar psicológico, si bien esto último lo pusieron a prueba tomando los seis factores como indicadores observables, y realizando un nuevo análisis factorial confirmatorio con estos dos constructos como factores latentes. Aún así, el ajuste de este modelo tampoco resultó satisfactorio.

La estructura factorial de las escalas de Ryff ha sido posteriormente analizada en sus diferentes versiones también por otros autores. Así, por ejemplo, en muestras holandesas, van Dierendonck (2005) encuentra que el mejor modelo es el de seis factores con un factor de segundo orden, si bien

los ajustes tampoco eran satisfactorios y exigieron la eliminación de ítems, dejando la escala en 39 ítems. Por su parte Abbott, Ploubidis, Huppert, Kuh, Wadsworth y Croudace (2006), encuentran también una estructura de seis factores, pero con un factor de segundo orden afectando a cuatro de los factores de primer orden, si bien el ajuste solamente resultó satisfactorio con la reducción de ítems y la modelización de factores que, al margen del contenido del ítem, agrupaban a los formulados positiva y negativamente. También Springer, Hauser y Freese (2006) evaluaron diversos modelos factoriales confirmatorios para las escalas de Ryff, en tres grandes masas de datos estadounidenses de carácter nacional, concluyendo que la distinción de seis factores es dudosa dadas sus altas intercorrelaciones. Este artículo ha sido posteriormente replicado por Ryff y Singer (2006), señalando que el mejor ajuste global sigue siendo el de seis factores, lo que no obstante no resuelve el problema de las altas inter-correlaciones entre factores (Springer, Hauser y Freese, 2006).

En nuestro contexto, diferentes estudios han utilizado estas escalas como base para la investigación empírica del bienestar en el envejecimiento (Navarro, Meléndez y Tomás, 2008; Triadó, 2003), y algunos han puesto a prueba su estructura factorial y propiedades psicométricas. Así, el trabajo de Díaz, Rodríguez-Carvajal, Blanco, Moreno-Jiménez, Gallardo, Valle y van Dierendonck (2006) ha probado diversos modelos para poner a prueba la estructura de estas escalas, intentando mejorar sus propiedades psicométricas, reduciendo ampliamente el número de ítems hasta una versión de 29. También Tomás, Meléndez y Navarro (2008) han estudiado la estructura factorial de las escalas, concretamente en muestra de personas jubiladas, no empleando directamente los ítems, sino agregados de éstos, y sin obtener ajustes plenamente satisfactorios. En un trabajo reciente, y también con muestra de personas mayores, Triadó, Villar, Solé y Celdrán (2007) han puesto a prueba las diversas estructuras factoriales encontradas en la literatura para la versión de 54 ítems, y las conclusiones son muy similares, aunque con ajustes altamente insatisfactorios.

Un problema adicional de los resultados de todos los trabajos que han empleado análisis factorial confirmatorio, es que resulta difícil escoger entre los modelos de cinco y seis factores, pues sus índices de ajuste son muy similares, algunas veces a favor del modelo de cinco y otras del de seis, lo que probablemente se encuentra relacionado con las elevadas correlaciones entre las dimensiones, que no favorecen la validez discriminante que los autores defienden de las escalas (Ryff y Singer, 2006).

Dada esta situación, la posible presencia de efectos de método en la escala, y sus versiones, no puede descartarse y un único trabajo, hasta la fecha, la ha puesto a prueba en el contexto anglosajón y en población

general (Springer y Hauser, 2006). Efectivamente, estos autores encontraron que los modelos factoriales de mejor ajuste (que no de ajuste satisfactorio) mejoraban claramente al introducir un efecto de método asociado a los ítems formulados en negativo.

A la vista de todos estos resultados el objetivo de este trabajo es doble. Por un lado estudiar la estructura factorial de la escala de Ryff en sus distintas versiones. Por otro lado, evaluar la presencia de efectos de método asociados a los ítems formulados en negativo.

MÉTODO

Muestra y procedimiento. La muestra estuvo compuesta por un total de 556 personas de más de 65 años, jubiladas, de la Comunidad Valenciana (España), no institucionalizadas y sin deterioro cognitivo grave ya que debían responder por sí mismos a las escalas, siendo el muestreo de tipo incidental. La recogida de datos se realizó de manera directa e individual entre septiembre de 2006 y marzo de 2007 obteniéndose previamente el consentimiento informado de los participantes en el estudio. El tiempo estimado de aplicación fue aproximadamente de 40 minutos por persona. Los entrevistadores fueron previamente entrenados en la aplicación de los instrumentos. El 59% de la muestra está compuesta por mujeres. La media de edad es de 76.46 años (DT= 7.22), con un mínimo de 65 años, por criterios de inclusión, y un máximo de 95 años. Por estado civil, un 53.5% están casados, solteros fueron el 11%, viudos el 32.7% y en otros el 2.8% restante. Por nivel de estudios, un 18.4% de la muestra no tenía estudios, un 54.6% primarios, un 17.5% de secundaria y un 9.5% eran universitarios. Por criterios de inclusión, las personas mayores debían vivir en domicilio y no en residencia, pero en casa propia vivían un 75%, un 14.5% en vivienda alquilada, y con familiares un 10.6%. Un 87.3% tienen hijos, la media de hijos es 2.6 (DT= 1.68).

Instrumentos. Para la recogida de datos de esta investigación, se utilizó una plantilla con diferentes elementos sociodemográficos y varios instrumentos: Índice Barthel (Mahoney y Barthel, 1965), Life Satisfaction Index (LSI-A) de Neugarten, Havighurst y Tobin (1961) y las escalas de Bienestar Psicológico de Ryff (1989), sobre las que versa este trabajo.

Las escalas de Ryff están compuestas por ítems redactados tanto en positivo como en negativo, aunque para su posterior análisis se invirtieron los formulados en negativo. El formato de respuesta que se utilizó tiene puntuaciones comprendidas entre 1 (totalmente en desacuerdo) y 6

(totalmente de acuerdo). Dado el procedimiento (encuesta) la no respuesta ha sido prácticamente inexistente, en algún caso algún ítem no ha sido contestado por un participante. Los ítems se agrupan teóricamente en seis factores, que se presentan a continuación, así como el significado de las puntuaciones altas en ellos: a) autonomía, si se es independiente y autodeterminado, capaz de resistir presiones sociales para pensar y actuar de ciertos modos, se regula la conducta internamente y las autoevaluaciones se realizan desde estándares personales; b) dominio del ambiente, si se posee sentido de dominio y competencia manejando el ambiente, se controlan múltiples conjuntos de actividades externas, haciendo un uso efectivo de las oportunidades del contexto y siendo capaz de elegir o crear contextos adaptados a las necesidades y valores personales; c) crecimiento personal, cuando se mantiene el sentimiento de desarrollo continuado y de mejora a través del tiempo, ve el yo en expansión y crecimiento, abierto a nuevas experiencias; sentido de desarrollo del propio potencial y de cambiar en modos que reflejan más autoconocimiento y efectividad; d) relaciones positivas con otros al mantener relaciones interpersonales cercanas, satisfactorias, de mutua confianza, se preocupa por el bienestar de otros, siendo capaz de desarrollar empatía, afecto, intimidad y desde la comprensión del dar y tomar de las relaciones humanas; e) autoaceptación si posee una actitud positiva hacia sí mismo, reconociendo y aceptando los múltiples aspectos del yo, incluyendo lo negativo y lo positivo, además existe un sentimiento positivo hacia lo ya vivido; y f) propósito en la vida cuando tiene metas en la vida y sentido de dirección, sentimiento de significado en el presente y en el pasado, manteniendo creencias que proporcionan significado a la vida.

Análisis de datos. Se han empleado análisis factoriales confirmatorios, todos ellos estimados mediante el paquete estadístico EQS 6.1. Los modelos confirmatorios se han estimado mediante máxima verosimilitud con correcciones robustas de Satorra-Bentler (Bentler, 1995) dado el alejamiento de la normalidad multivariada en los indicadores, coeficiente de Mardia multivariado de 36.74, siendo éste el procedimiento recomendado en estas situaciones (Finney y DiStefano, 2006). Para evaluar el ajuste de los modelos confirmatorios se han utilizado una variedad de índices, tal y como se recomienda en la literatura. Así se emplea el CFI, GFI para los que se suele indicar 0.9 o superior como indicativo de un mínimo para aceptar el modelo; el RMSEA y el SMSR, medidas de cuantía del error, con valores menores de 0.08 considerados como aceptables; y el AIC con el objetivo de comparar modelos; además de la prueba de ji-cuadrado (ver, por ejemplo, Hoyle y Panter, 1995; Hu y Bentler, 1995, 1999; Marsh,

Balla, y Hau, 1996; Kaplan, 2000). Para evaluar los efectos de método asociados a ítems negativos se ha optado por modelar los factores de rasgo y con un factor adicional que se asocia a los ítems formulados en negativo (ver figura 1). Este método de análisis se ha recomendado frente al de unicidades correlacionadas (Tomás, Hontangas y Oliver, 2000).

RESULTADOS

En un primer momento, se han calculado todos los modelos confirmatorios substantivos que han aparecido en la literatura psicométrica sobre la escala de Ryff, y en sus distintas versiones. Se ha optado por aplicar solamente los modelos de primer orden, y no los modelos confirmatorios en que se incluían factores de segundo orden, pues estos se han planteado por las altas intercorrelaciones entre factores, que podrían, como hemos visto, estar debidas a un efecto de método. Por tanto, los pasos a dar son: calcular primero los modelos factoriales de primer orden encontrados en la literatura, para ver cuál de ellos presenta el mejor ajuste por comparación de modelos; y segundo modelar sobre el modelo de mejor ajuste un factor de método adicional. Los modelos substantivos probados serán:

- Modelo de un factor. Todos los ítems de la escala se agrupan en un único factor de bienestar que funcionaría como un modelo de línea base.
- Modelo de dos factores, que incluye los factores de bienestar subjetivo y psicológico defendidos por Keyes, Shmotkin y Ryff (2002).
- Modelo de cinco factores, que responde a la modificación de la estructura original de seis factores de las escalas, con una única modificación, la unión de los factores de autoaceptación y dominio del ambiente en un único factor, dada su elevada correlación, según los análisis realizados por van Dierendonck (2005) y Díaz et al. (2006).
- Modelo de seis factores. Los seis componentes originales del bienestar basados en la división teórica de la autora de la escala correlacionados entre sí (Ryff, 1989).

Estos modelos se han puesto a prueba en las distintas versiones de la escala, cada una de ellas de longitudes diferentes. Los resultados de los modelos factoriales sin efecto de método (ver tabla 1) muestran, en primer lugar, que los índices de ajuste no son adecuados para ninguno de ellos y,

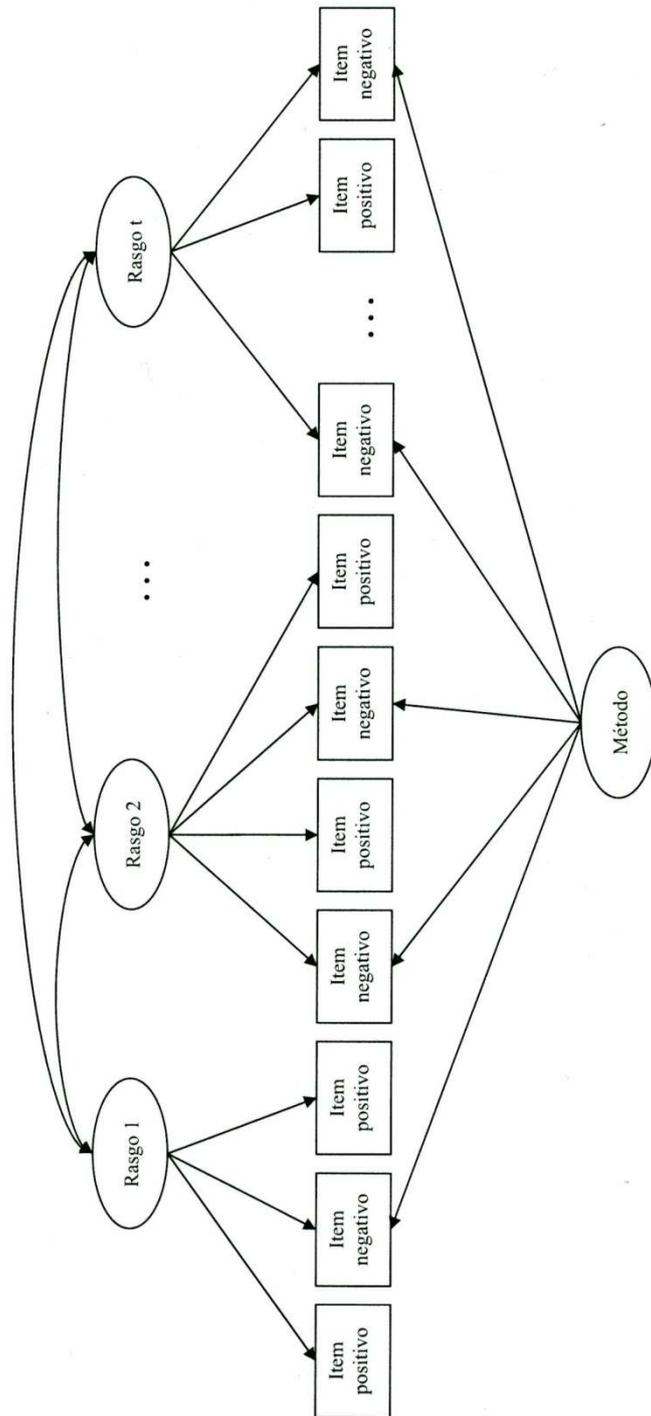


Figura 1. Modelo factorial confirmatorio empleado para probar los efectos de método asociados a los ítems formulados en negativo en la escala de bienestar de Ryff.

en segundo lugar, que los ajustes de todos ellos son similares, si bien el de cinco factores y, especialmente, el de seis factores presentan un ligero mejor ajuste. Este patrón de índices de ajuste ocurre en todas las versiones de la escala, tanto en las versiones de 54 ítems, como en la versión corta de 18 ítems, y también en las modificaciones presentadas por otros autores de 29 y 39 ítems, que se basaban en la lógica de eliminar los ítems con peor comportamiento psicométrico, pero siempre sin evaluar los efectos de método.

Tabla 1. Ajuste global de los modelos confirmatorios para las distintas versiones de las escalas de bienestar de Ryff.

Versión	Modelo	χ^2	d.f.	AIC	GFI	SRMR	CFI	RMSEA
18 ítems	1 factor	414.65	135	144.65	0.910	0.064	0.684	0.061
	2 f	399.14	134	131.14	0.913	0.063	0.700	0.059
	5f	36.81	125	11.81	0.920	0.061	0.733	0.058
	6f	312.48	120	72.47	0.931	0.057	0.782	0.053
	6f 1 met	226.13	112	2.13	0.952	0.046	0.871	0.043
29 ítems	1f	1075.7	377	321.79	0.849	0.065	0.727	0.057
	2f	1067.5	376	315.50	0.851	0.065	0.730	0.057
	5f	964.94	367	23.93	0.864	0.063	0.767	0.054
	6f	924.89	361	202.89	0.867	0.062	0.780	0.053
	6f 1 met	799.64	406	95.64	0.892	0.055	0.825	0.048
39 ítems	1f	2139.4	702	735.49	0.775	0.074	0.621	0.060
	2f	2117.0	701	715.01	0.779	0.074	0.627	0.060
	5f	1948.5	692	564.51	0.801	0.073	0.669	0.057
	6f	1928.8	687	554.84	0.806	0.073	0.673	0.057
	6f 1 met	1584.3	670	244.26	0.845	0.062	0.759	0.049
54 ítems	1f	3847.2	1377	1093.2	0.711	0.075	0.516	0.057
	2f	3827.0	1376	1075.0	0.714	0.076	0.519	0.057
	5f	3665.2	1367	931.26	0.722	0.075	0.549	0.055
	6f	3637.1	1362	913.10	0.724	0.075	0.554	0.055
	6f 1 met	2789.3	1334	121.27	0.811	0.059	0.715	0.044

Dados los resultados de los modelos con diversas dimensiones de bienestar, y a la vista de la abundante literatura, se escogió el modelo de seis factores, el de mejor ajuste, para a partir de éste añadir el factor adicional de método. Al observar los ajustes de los modelos hay que notar que el ajuste no es claramente satisfactorio para ninguno de ellos. Este

resultado es general en todos los trabajos confirmatorios realizados con la escala y en todos los contextos culturales probados. Los índices de ajuste están cercanos a aceptables, pero no son plenamente adecuados (Tomás, Meléndez y Navarro, 2008). Como puede verse en la tabla 1, el añadir un factor de método al modelo de seis factores correlacionados ha mejorado de forma clara el modelo. Esta mejora del ajuste es especialmente aparente en el índice AIC, de comparación de modelos, que se rebaja considerablemente, y además, es de carácter general a través de versiones, incluyendo aquéllas como las de 29 y 39 ítems. En todos los casos, las saturaciones factoriales asociadas al método fueron estadísticamente significativas ($p < 0.05$).

Visto que el factor de método mejora el ajuste, y que las saturaciones asociadas son estadísticamente significativas, cabe calcular la cuantía de este efecto de método. El análisis de las saturaciones de método ofrece un promedio de saturación estandarizada de 0.31 ($S_x = 0.096$), en absoluto despreciable. Es posible que el efecto de método esté diferencialmente presente en las cuatro versiones de la escala. La figura 2 muestra las medias e intervalos de confianza para las saturaciones estandarizadas de método en las cuatro versiones de la escala de Ryff. Puede verse que las medias son muy similares y los intervalos solapados. Una prueba de Kruskal-Wallis para evaluar posibles diferencias no resultó significativa ($\chi^2_{(9)} = .478$, $p = 0.924$).

Si los efectos de método son significativos, de cuantía no despreciable, y se encuentran por igual en las cuatro versiones de la escala y, dado que los ítems formulados en negativo están en todos los factores, quizá las altas correlaciones entre los factores de rasgo queden disminuidas una vez se modela el factor de método. Para comprobarlo, se analizaron las correlaciones de los rasgos *sin* y *con* el factor de método. A la vista de la figura 3 resulta evidente que las altas correlaciones no desaparecen al modelar el efecto de método. De hecho una prueba de Mann-Whitney no encontró diferencias significativas entre ambos tipos de correlaciones, con y sin factor de método ($z = 0.427$, $p = 0.427$).

Por último, cabía preguntarse si el introducir los efectos de método afecta a la cuantía de las saturaciones de rasgo. El primer detalle es que las saturaciones factoriales siguen siendo estadísticamente significativas. Pero queremos saber más, queremos saber si la relación del ítem con el factor de rasgo disminuye al modelar el factor de método. Para comprobarlo se han analizado las diferencias entre las saturaciones estandarizadas de rasgo, sin modelar el factor de método, con esas mismas saturaciones al modelar el factor de método. Una prueba no paramétrica de Mann-Whitney no mostró

diferencias significativas ($z = -1.17$, $p = .24$). De hecho la media de las saturaciones cuando no se modela el factor de método es de 0.4, mientras que si se modela solamente baja a 0.38. La figura 4 muestra los intervalos de confianza de las medias en esas dos situaciones.

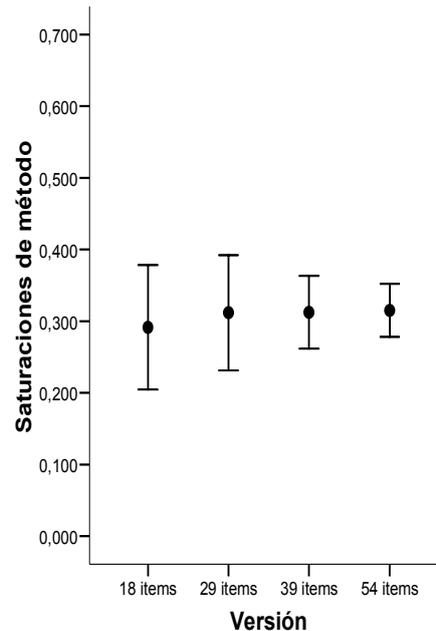


Figura 2. Medias y sus intervalos de confianza del 95% para las saturaciones factoriales estandarizadas de método de las cuatro versiones de la escala de Ryff.

DISCUSIÓN

Como se ha visto en la introducción, los objetivos de este trabajo eran dos: estudiar la estructura factorial de la escala de Ryff en sus distintas versiones, y evaluar la presencia de efectos de método asociados a los ítems formulados en negativo. Creemos que nuestro trabajo es de interés, porque aún en un solo trabajo evidencia que se encontraba relativamente dispersa en la literatura, y porque hace esto por primera vez en nuestro contexto cultural y en una muestra de población especial, geriátrica, que no procede de la población general.

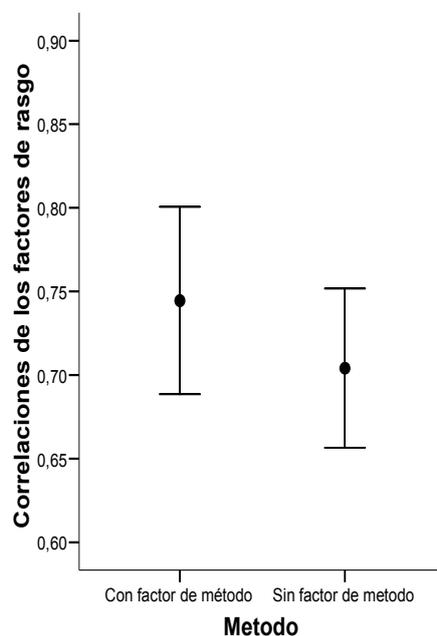


Figura 3. Medias y sus intervalos de confianza del 95% para las correlaciones entre factores de rasgo en todos los modelos con y sin factor de método.

En cuanto al primer objetivo, se ha estudiado la estructura factorial de la escala de Ryff, de forma sistemática, a través de versiones, por primera vez en muestra española con población geriátrica. Se han analizado las versiones empleadas más habitualmente de la escala, las de 54 y 18 ítems planteadas por la autora, y las versiones depuradas de éstas por razones psicométricas realizadas por otros autores (29 y 39 ítems). En el contexto español, se tiene evidencia empírica de la estructura factorial de esta escala en población general (Díaz et al., 2006), y además en población geriátrica (Tomás et al., 2008; Triadó et al., 2007). Esta evidencia apuntaba en la misma dirección de los resultados de este trabajo. Cuando se plantean modelos solamente de rasgo, sustantivos, el ajuste confirmatorio es inadecuado. Esto, como hemos visto ocurre también en los trabajos con muestras anglosajonas (Abbott et al., 2006; Keyes et al., 2002; Ryff y Keyes 1995; Ryff y Singer, 2006; Springer et al., 2006; Van Dierendonck, 2005). Por tanto la evidencia es que las escalas originales no tienen una estructura factorial que pueda defenderse, y son necesarias reducciones, a veces drásticas. En nuestro contexto solamente tras reducir de 54 a 29

ítems, Díaz et al. (2006) encuentran un ajuste razonable. No obstante, hay que considerar que siguen teniendo una muy baja validez discriminante en los factores, y que no ha habido estudios de validación cruzada, y el ajuste puede deberse a la capitalización del azar.

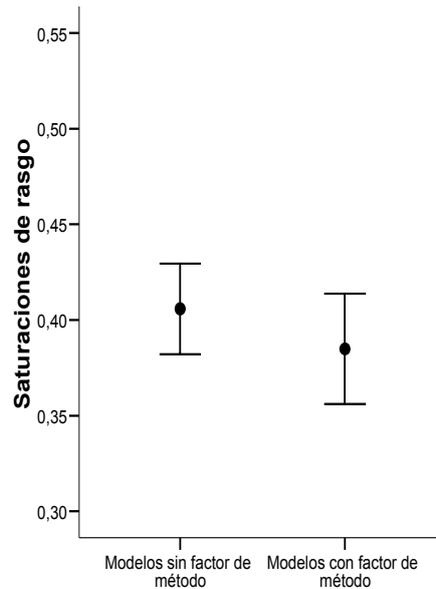


Figura 4. Medias y sus intervalos de confianza del 95% para las saturaciones estandarizadas de rasgo en todos los modelos con y sin factor de método modelado.

Adicionalmente, algunos de estos trabajos (Díaz et al., 2006; Van Dierendonck, 2005) ya ponían sobre la pista, aunque indirectamente, de que podían existir efectos de método asociados a los ítems negativos, pues al probar un modelo de dos factores correlacionados asociados a los ítems positivos y a los negativos obtenían ajustes muy similares a los de los modelos sustantivos de cinco y seis factores. No obstante, esta evidentemente no es la mejor forma de obtener evidencia sobre efectos de método, sino modelarlo explícitamente, lo que solamente se ha realizado en el trabajo de Springer y Hauser (2006), y en muestra de población general estadounidense. Las conclusiones de este trabajo coinciden plenamente con las del nuestro, pues la mejora en términos de ajuste al introducir el factor

de método asociado a los ítems negativos es muy clara y sistemática, en todas las versiones. A nivel incluso más específico, los resultados son también plenamente coincidentes, pues las cuantías de los efectos de método resultan muy similares, y el efecto de modelar un factor de método es también similar, pues las saturaciones y correlaciones de rasgo permanecen prácticamente iguales, de forma que la varianza explicada por el método se añade a la explicada por el rasgo en los ítems. En resumen, que a pesar del salto de población estadounidense a española y de personas de todas las edades a mayores de 65, los resultados son coherentes con la evidencia de la literatura.

El trabajo presentado, evidentemente, no supone un punto final a la investigación, y algunas líneas de investigación futura pueden claramente atisbarse. Especialmente tres: a) todavía queda por dar respuesta a la estructura de rasgo más adecuada para la escala y la repercusión del efecto de método en la estimación de los niveles de esos rasgos; b) resultaría interesante evaluar qué posibles variables explican la mayor o menor presencia de efectos de método asociado a los ítems negativos (nivel educativo, deseabilidad social, dificultades lectoras,...); y c) es necesario avanzar en el efecto que puede tener modelar o no este efecto de método cuando se relacionan los factores de rasgo de la escala con otras variables de interés en investigaciones empíricas.

ABSTRACT

Method effects in Ryff's scales: a study in an elderly population.

Although widely used, Ryff's scales have had some problems of factorial validity and recent investigations have also shown method effects. However, these method effects have not been deeply investigated in special populations such as the elderly, the population under study in this paper. The aim of this research is twofold: to study the factorial structure of Ryff's scale in its different versions; and to evaluate the presence of method effects associated to negatively worded items. In order to accomplish these aims, a sample of 556 Valencian elderly people living in their homes has been surveyed. All confirmatory models used in the literature to test for factorial validity have been tested. Maximum likelihood with robust corrections has been used to estimate the models. Best fitting model posited a six factor structure, as originally proposed by the author. With respect to method bias, results shown a consistent and non-trivial method factor effect associated to negatively worded items.

REFERENCIAS

- Abbott, R. A., Ploubidis, G. B., Huppert, F. A., Kuh, D., Wadsworth, M. E. J. y Croudace, T. J. (2006). Psychometric evaluation and predictive validity of Ryff's psychological well-being items in a UK birth cohort sample of women. *Health and Quality of Life Outcomes*, 4, 76.
- Bass, B. M. (1955). Authoritarianism or Acquiescence? *Journal of Abnormal and Social Psychology*, 51, 616-623.
- Bentler, P. M. (1995). *EQS program manual*. Encino, CA: Multivariate Software, Inc.
- Bollen, K. A. (1989). *Structural equation modelling with latent variables*. New York: Wiley.
- Campbell, D. T. y Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Conway, J. M. (2002). Method variance and method bias in Industrial and Organizational Psychology. En S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 344-365). Malden, MS: Blackwell Publishers Inc.
- Cordery, J. L. y Sevastos, P. P. (1993). Responses to the regional and revised Job Diagnostic Survey. *Journal of Applied Psychology*, 78, 141-143.
- Cote, J. A. y Buckley, M. R. (1988). Measurement error and theory testing in consumer behavior research: An illustration of the importance of construct validation. *Journal of Consumer Research*, 14, 579-582.
- Di Stefano, C. y Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440-464.
- Di Stefano, C. y Motl, R. W. (2008). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*, 46, 309-313.
- Díaz, D., Rodríguez-Carvajal, R., Blanco, A., Moreno-Jiménez, B., Gallardo, I., Valle, C. y van Dierendonck, D. (2006). Adaptación española de las escalas de bienestar psicológico de Ryff. *Psicothema*, 18, 572-577.
- Finney, S. J. y DiStefano, C. (2006). Non-normal and categorical data in SEM. En G. R. Hancock y R. O. Mueller (Eds.), *Structural Equation Modeling: A second course*. Greenwich, CO: Information Age Publishing.
- Harvey R., Billings R. y Nilan K. (1985). Confirmatory factor analysis of the Job Diagnostic Survey: Good news and bad news. *Journal of Applied Psychology*, 70, 461-468.
- Horan, P. M., Di Stefano, C. y Motl, R. W. (2003). Wording effects in self esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, 10, 435-455.
- Hoyle, R. H. y Panter, A. T. (1995). Writing about structural equation models. En R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 159-176). Thousand Oaks, CA: Sage.
- Hu, L. y Bentler, P. M. (1995). Evaluating model fit. En R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

- Hu, L. y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Idaszak, J. R. y Drasgow, F. (1987). A revision of the Job Diagnostic Survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, 72, 69-74.
- Jackson, D. N. y Messick, S. J. (1957). A note on ethnocentrism and acquiescent response sets. *Journal of Abnormal and Social Psychology*, 54, 132-134
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Keyes, C., Shmotkin, D. y Ryff, C. D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology*, 82, 1007-1022.
- Kulik, C. T., Oldham, G. R. y Langner, P. H. (1988). The measurement of job characteristics: A comparison of the original and the revised Job Diagnostic Survey. *Journal of Applied Psychology*, 73, 426-466.
- Ladd, R. T., Gordon, M. E, Beauvais, L. L. y Morgan, R. L. (1982). Union Commitment: Replication and extension, *Journal of Applied Psychology*, 67, 640-644.
- Leavitt, H. J., Hax, H. y Roche, J. H. (1955). Authoritarianism and agreement with things authoritative. *Journal of Psychology*, 40, 215-221.
- Magazine, S. L., Williams, L. J. y Williams, M. L. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's Affective and Continuance Commitment Scales. *Educational and Psychological Measurement*, 56, 241-250.
- Mahoney, F. I. y Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal*, 14, 61-65.
- Mehrabian, A. y Ljunggren, E. (1997). Dimensionality and content of optimism-pessimism analyzed in terms of the PAD Temperament Model. *Personality and Individual Differences*, 23, 729-737.
- Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70, 810-819.
- Marsh, H. W., Balla, J. R. y Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. En G. A. Marcoulides y R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315-353). Mahwah, NJ: Erlbaum.
- Messick, S. y Jackson, D. (1957). Authoritarianism or acquiescence in Bass' data. *Journal of Abnormal and Social Psychology*, 54, 424-426.
- Motl, R. W. y Di Stefano, C. (2002). Longitudinal invariance of self-esteem and methods effects associated with negatively worded items. *Structural Equation Modeling*, 9, 562-578.
- Navarro, E., Meléndez, J. C. y Tomás, J. M. (2008). Análisis del bienestar en la vejez según la edad. *Revista Española de Geriatria y Gerontología*, 43, 90-95.
- Neugarten, B. L., Havighurst, R. J. y Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology*, 16, 134-143.
- Nunnally, J. C. (1978). *Psychometric theory*. 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, 99-117.
- Ryff, C. D. (1989). Happiness is everything: or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069-1081.

- Ryff, C. D. y Keyes, C. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.
- Ryff, C. D. y Singer, B. H. (2006). Best news yet on the six-factor model of well-being. *Social Science Research*, 35, 1103-1119.
- Ryff, C. D., Lee, Y. H., Essex, M. J., y Schmutte, P. S. (1994). My children and me: Midlife evaluations of grown children and of self. *Psychology and Aging*, 9, 195-205.
- Scheirer, M. F., Carver, C. S. y Bridges M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 1063-1078.
- Schriesheim, C. A. y Eisenbach, R. J. (1995). An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *Journal of Management*, 21, 1177-1193.
- Springer, K. W., Hauser, R. M. y Freese, J. (2006). Bad news indeed for Ryff's six-factor model of well-being. *Social Science Research*, 35, 1119-1130.
- Springer, K. W. y Hauser, R. M. (2006). An assessment of the construct validity of Ryff's scales of psychological well-being: Method, mode, and measurement effects. *Social Science Research*, 35, 1080-1102.
- Tomás, J. M. y Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6, 84-98.
- 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., Meléndez, J. C. y Navarro, E. (2008). Modelos factoriales confirmatorios de las escalas de Ryff en una muestra de personas mayores. *Psicothema*, 20, 298-304.
- Triadó, C. (2003). *Envejecer en entornos rurales*. Madrid: IMSERSO, Estudios I+D+I, nº 19.
- Triadó, C., Villar, F., Solé, C. y Celdrán, M. (2007). Construct validity of Ryff's scale of psychological well-being in Spanish older adults. *Psychological Reports*, 100, 1151-1164.
- Van Dierendonck, D. (2005). The construct validity of Ryff's scale of psychological well-being and its extension with spiritual well-being. *Personality and Individual Differences*, 36, 629-644.
- Vautier, S., Raufaste, E. y Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the status of filler items. *International Journal of Psychology*, 38, 390-400.
- 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.
- Williams, L. J., Ford, L. R. y Nguyen, N. (2002). Basic and advanced measurement models for confirmatory factor analysis. En S. Rogelberg (Ed.), *Research Methods in Industrial and Organizational Psychology* (pp. 366-389). Oxford: Blackwell Publishers.