

Propiedades psicométricas del test de optimismo Life Orientation Test

Pere Joan Ferrando, Eliseo Chico y Josep M^a Tous*
Universidad Rovira i Virgili y * Universidad de Barcelona

El presente estudio evalúa las propiedades psicométricas de un test de optimismo (LOT-R) adaptado al castellano mediante el modelo multidimensional de respuesta graduada (MMRG). El estudio tiene dos vertientes: sustantiva y metodológica. Desde el punto de vista sustantivo interesa evaluar las propiedades de un test que, en su versión original, muestra interesantes propiedades para la investigación aplicada en personalidad. Desde el punto de vista metodológico se pretende ilustrar el uso de un modelo que, hasta ahora, no ha sido prácticamente utilizado en investigación aplicada, así como evaluar sus potenciales ventajas teóricas. Los resultados sustantivos indican que la adaptación del LOT-R tiene propiedades muy similares a las de la versión original. Desde el punto de vista metodológico, el MMRG presenta una serie de ventajas respecto al análisis tradicional. Los resultados, tanto sustantivos como metodológicos, se discuten desde un punto de vista aplicado.

Psychometric properties of the 'Life Orientation Test' (LOT): An empirical application of the Multidimensional Graded Response Model. This study examines the psychometric properties of the LOT-R Spanish adaptation using the multidimensional graded response model (MGRM). The study has two purposes: substantive and methodological. From a substantive point of view we want to assess the behaviour of a test which, in its original version, has interesting properties in applied personality measurement. From a methodological point of view our study illustrates the application of the MGRM in the personality domain, and allows us to assess whether the theoretical advantages of the MGRM are important from an applied point of view. The substantive results suggest that the properties of the Spanish adaptation are similar to those of the original test. The methodological results show that the model has some advantages with respect to traditional analyses. Both type of results are discussed from an applied point of view.

Aunque los conceptos tales como 'pensamiento positivo' u 'optimismo' han formado parte de lo que podríamos denominar 'Psicología popular' desde hace muchos años, tan sólo recientemente han empezado a ser objeto de investigación rigurosa por parte de la Psicología de la personalidad. Sin embargo, a pesar de este tardío inicio, la cantidad de investigación generada por el tema en las dos últimas décadas ha sido muy importante. La revisión de la literatura indica, además, que el interés por el constructo de optimismo no se debe tanto a preocupaciones teóricas como al hecho de que dicho constructo permite predecir variables de notable importancia en Psicología clínica. De esta forma, el optimismo parece jugar un importante papel en el uso de conductas de afrontamiento adaptativas (Scheier y Carver, 1985; Scheier Weintraub y Carver, 1986), así como en el bienestar psicológico y físico (Chang, D'Zurilla y Maydeu, 1994), y parece ser un importante predictor de la enfermedad, tanto psicológica como física (Mroczek, Spiro, Aldwin, Ozer y Bosse, 1993; Peterson, Seligman y Vaillant, 1988).

De las diferentes aproximaciones al concepto de optimismo, este trabajo se centra en el enfoque de Scheier y Carver (1985). Es-

tos autores toman como punto de partida un modelo de autorregulación de conducta. Brevemente, la parte relevante de este modelo asume que, cuando surgen dificultades, las expectativas favorables incrementan los esfuerzos de las personas para alcanzar objetivos, en tanto que las expectativas desfavorables reducen tales esfuerzos, a veces hasta el punto de desentenderse totalmente de la tarea. Dentro de este modelo, el optimismo y el pesimismo serían considerados como expectativas generalizadas (favorables y desfavorables respectivamente) acerca de las cosas que le suceden a uno en la vida. Tales expectativas se consideran además como disposiciones estables (es decir, rasgos). De aquí que Scheier y Carver hablen de optimismo disposicional.

La evolución del trabajo de Scheier y Carver es típica de las teorías de la personalidad que Hampson (1986) denomina 'de rasgo único'. La teoría se gesta en un laboratorio y recibe un cierto apoyo empírico inicial. En una segunda etapa se pretenden generalizar los hallazgos a otros dominios de conducta, así como obtener una evidencia más 'ecológica', para lo cual se construye un test. En este caso además, el test se hace muy popular, hasta el punto de llegar a ser casi sinónimo de la dimensión que pretende medir (Ahrens y Haaga, 1993; Anderson, 1996; Boland y Cappeliez, 1997; Olason y Roger, 2001; Sweetman, Munz y Wheeler, 1993).

El test desarrollado por Scheier y Carver es el «Life Orientation Test» (LOT, Scheier y Carver, 1985). La versión inicial fue revisada en 1994, dando lugar al LOT-R (Scheier, Carver y Bridges, 1994), que es la versión que consideraremos aquí. El LOT-R cons-

Fecha recepción: 29-1-02 • Fecha aceptación: 11-3-02

Correspondencia: Pere Joan Ferrando

Facultad de Psicología

Universidad Rovira i Virgili

43007 Tarragona (Spain)

E-mail: pjfp@fcep.urv.es

ta de 10 ítems en escala Likert de 5 puntos. 6 ítems pretenden medir la dimensión de optimismo disposicional en tanto que los otros 4 ítems son «de relleno» y sirven para hacer menos evidente el contenido del test. De los 6 ítems de contenido, 3 están redactados en sentido positivo (dirección optimismo) y 3 en sentido negativo (dirección pesimismo). Los ítems redactados en sentido negativo se revierten y se obtiene una puntuación total orientada hacia el polo de optimismo. La adaptación del LOT-R utilizada en el presente trabajo se incluye en el apéndice 1. Junto a cada ítem aparece la clave con la que se presentarán los resultados de este ítem en la presente investigación (o1,o2,o3,p1,p2,p3).

En su estudio normativo sobre el LOT-R, Scheier et al. (1994) obtuvieron una estructura factorial muy similar en hombres y mujeres y una fiabilidad de las puntuaciones estimada en $\alpha = 0.78$ (aceptable en un instrumento tan corto y en el dominio de personalidad). Los estudios posteriores, sin embargo, han criticado al LOT-R por dos razones: (1) el test en realidad podría estar midiendo una tercera variable más amplia de personalidad y (2) la estructura del test es bidimensional. Respecto al punto (1), la evidencia parece indicar que la crítica es incorrecta (Chang, 1998b; Myers y Steed, 1999). Nos centraremos, pues, en la segunda crítica.

De acuerdo con la teoría de Scheier y Carver, el constructo de optimismo-pesimismo es unidimensional y bipolar, es decir, es un rasgo único con dos extremos. Sin embargo, prácticamente todos los estudios factoriales del LOT-R (incluidos los de Scheier y Carver) obtienen dos factores, formados, uno por los ítems de optimismo y el otro por los de pesimismo (Mroczek et al., 1993; Robinson-Whelan, Kim, MacCallum y Kiecolt-Glaser, 1997). Como es de esperar, los factores están negativamente correlacionados, obteniéndose valores que oscilan entre -0.47 y -0.57 (Dember, Martin, Hummer y Melton, 1989; Marshall, Wortman, Kusulas, Hervig y Vickers, 1992; Myers y Steed, 1999). Son valores sustanciales, desde luego, pero insuficientes para considerar un factor único.

El consistente resultado que acabamos de describir podría deberse a dos razones: (1) sesgos de respuesta debido al formato de los ítems (los de optimismo están redactados positivamente y los de pesimismo negativamente) o (2) al contenido diferencial de los ítems. Scheier y Carver (1985; Scheier, Carver y Bridges, 1994) defienden la primera explicación, y sugieren que los sesgos de respuesta se deben a que los ítems redactados en positivo son más cortos y directos. Otros autores, sin embargo, defienden que se trata de dos constructos parcialmente independientes (Plomin, Scheier, Bergeman, Pederson, Nesselroade y McCleam, 1992). Es éste un punto que analizaremos en este trabajo.

Respecto a la validez, como hemos dicho antes, el constructo parece tener importantes relaciones y existe una buena cantidad de estudios de validación. En el presente trabajo estudiaremos su relación con tres constructos: (1) afectividad negativa, (2) estrés percibido y (3) neuroticismo. Investigaciones previas han mostrado que el optimismo disposicional es un importante predictor de estas variables (Aherens y Haaga, 1993; Anderson, 1996; Vickers y Vogeltanz, 2000; Olason y Roger, 2001).

Para que nuestros resultados puedan compararse con los obtenidos anteriormente mediante el test original hemos utilizado la medidas más habituales en los estudios previos de validez. Para medir el Afecto positivo-Afecto negativo hemos escogido el PANAS (Watson, Clark y Tellegen, 1988). En contra de lo que pueda parecer a simple vista, las dimensiones de Afecto positivo y Afecto negativo tal como las mide este test deben considerarse como

esencialmente independientes, por lo que una reducción del Afecto negativo no implica necesariamente un incremento del afecto positivo (Avia y Sánchez, 1995).

Utilizando el LOT-R y el PANAS, Olason y Roger (2001) obtienen correlaciones entre optimismo y afectividad positiva de 0.33, y entre optimismo y afectividad negativa de -0.48, en tanto que Marsalt, Nortman, Kusulas, Hervig y Vickers (1992) obtienen correlaciones de 0.37 y -0.22, respectivamente.

La medida de estrés percibido que hemos utilizado en este estudio es la escala de Estrés Percibido PSS (Cohen et al., 1983). Utilizando el LOT-R y la PSS, Chang, Rand y Strunck (2000) obtienen una correlación de -0.41, Scheier y Carver (1985) de -0.55 y Chang (1998a) de -0.54. Finalmente, como medida de neuroticismo hemos utilizado la escala N de neuroticismo del EPQ-R (Eysenck, Eysenck y Barrett (1985). Utilizando el LOT-R y la escala N, Marshall, Wortman, Kusulas, Hervig y Vickers (1992) obtienen una correlación de -0.35, en tanto que Scheier, Carver y Bridges (1994) y Mehsabian y Ljunggren (1997) obtienen ambos una correlación de -0.36.

Objetivos

Los objetivos de este trabajo son sustantivos y metodológicos. Desde el punto de vista sustantivo pretendemos evaluar las propiedades de un instrumento psicométrico adaptado al castellano. En nuestra opinión este instrumento puede ser de utilidad, especialmente en investigación aplicada en el dominio de personalidad. Desde el punto de vista metodológico, la evaluación utiliza procedimientos que, si bien son conocidos en la literatura psicométrica, han sido muy poco utilizados en estudios empíricos. El presente trabajo, por tanto, puede verse también como un ejemplo ilustrativo del uso de estos procedimientos. Además, al aplicar estos procedimientos a un caso 'real' quizá podamos hacernos una idea de hasta qué punto sus potenciales ventajas teóricas resultan importantes en la práctica.

Justificación psicométrica y descripción de los procedimientos

Las puntuaciones en escalas de Likert se analizan casi siempre utilizando modelos lineales, bien el modelo descriptivo clásico de análisis de ítems o bien el modelo factorial común. Desde un punto de vista teórico estricto tales modelos son incorrectos, ya que asumen variables continuas, no limitadas y medidas por lo menos en escala de intervalo. Las puntuaciones en una escala de Likert no cumplen desde luego los dos primeros supuestos y tampoco puede demostrarse que cumplan el tercero. Así pues, en el mejor de los casos, los modelos lineales deben considerarse como una aproximación.

En situaciones aplicadas, sin embargo, y bajo condiciones no muy extremas (es decir, en la mayor parte de aplicaciones), la evidencia empírica indica que el uso de modelos lineales tiende a producir ajustes aceptables y estimaciones que son razonables y tienen sentido desde la teoría. Algunos autores (e.g. Bentler, 1995; Loehlin, 1992) piensan que la aproximación lineal es ya suficientemente buena para analizar medidas de Likert razonablemente bien construidas, y que no es necesario el uso de modelos más complejos.

Los modelos no lineales para el análisis de escalas de Likert se han desarrollado desde el marco de la Teoría de Respuesta a los Ítems (TRI). Para este tipo de escalas, los supuestos de estos modelos son más plausibles que los del modelo lineal, lo que los ha-

ce a priori teóricamente más correctos. Además, presentan respecto al modelo lineal las potenciales ventajas que tiene la TRI respecto a la teoría clásica del test (TCT, véase e.g. Muñiz, 1997). Sin embargo, los modelos no lineales tienen también importantes limitaciones: (1) son complejos matemáticamente y difíciles de estimar; (2) el software necesario para su ajuste es específico y no se encuentra disponible en los paquetes de uso general; (3) debido a las dificultades de estimación tienden a funcionar bien sólo en el caso de tests cortos administrados en muestras grandes; y (4) no está claro que sus ventajas teóricas sean tan importantes desde un punto de vista práctico.

Las limitaciones descritas arriba justifican que los modelos no lineales se hayan utilizado muy poco en la práctica. Sin embargo, el presente estudio es un caso donde su utilización puede ser ventajosa, ya que nos ocupamos de un test muy corto que ha sido administrado a una muestra bastante grande. Además, algunas de las ventajas potenciales de la TRI podrán ponerse aquí en evidencia. En particular, las ventajas que evaluaremos son: (1) una mayor información acerca de las características de los ítems, y (2) una mayor precisión en la estimación de las puntuaciones de los sujetos. Respecto a (1) se espera que el modelo no lineal nos permita hacer un análisis detallado de este pequeño conjunto de ítems más allá de lo que permitiría el modelo lineal. Respecto a (2) podría pensarse que la mayor precisión debería traducirse en inferencias más claras a partir de las puntuaciones del test; quizás en una mejora de la validez.

El modelo TRI no lineal que consideramos en este trabajo es el modelo multidimensional de respuesta graduada (MMRG, Muraki y Carlson, 1995). En el caso de que los ítems del LOT se comportasen en forma unidimensional, el modelo reduciría al modelo unidimensional de respuesta graduada propuesto inicialmente por Samejima (1969, 1997). En el caso bidimensional el modelo se define como sigue. Para un individuo con niveles en el rasgo θ_1 y θ_2 , la probabilidad de responder en la categoría k en un ítem j viene dada por

$$P_{jk}(\theta) = \int \frac{a_{j1}\theta_{i1} + a_{j2}\theta_{i2} - b_{j,k-1}}{a_{j1}\theta_{i1} + a_{j2}\theta_{i2} - b_{j,k}} \phi(t) dt$$

donde $\phi(t)$ es la función de densidad de una variable normal tipificada. Los parámetros ' a_{j1} ' y ' a_{j2} ' son los índices de discriminación. El parámetro ' b_{jk} ' es un índice de posición o dificultad. La expresión anterior corresponde a la versión de ojiva normal del modelo, que será la que utilizemos aquí. La correspondiente versión logística produce virtualmente los mismos resultados.

La interpretación del parámetro de discriminación (a) con respecto a un determinado rasgo es una extensión inmediata del modelo TRI de dos parámetros. Cuanto mayor es el valor de a , mejor discrimina cada una de las categorías de respuesta entre los participantes que escogen esta categoría y los que escogen categorías adyacentes. Dado que sólo hay un índice de discriminación por ítem y rasgo, esta capacidad de discriminación se considera constante a través de las diferentes categorías. De acuerdo con Samejima (1997) un modelo que cumple esta propiedad es 'homogéneo'.

Los parámetros de posición (b), sin embargo, no son tan fáciles de interpretar. Como señala Baker (1992) estos parámetros resultan de una transformación matemática que permite reducir el modelo graduado a una serie de ojivas de dos parámetros, facilitando con ello la estimación de parámetros. Sin embargo, la interpreta-

ción que resulta de esta transformación es bastante indirecta y, además, existe el problema de que siempre hay un parámetro menos que el número de categorías. A efectos de interpretación, puede ser más conveniente trabajar con los valores modales de cada categoría de respuesta (Samejima, 1969). El valor modal de las categorías primera y última se interpreta igual que en el modelo de dos parámetros: es el nivel en el rasgo en el que la probabilidad de responder en la categoría es de 0.5. Para las restantes categorías el valor modal es aquel nivel en el rasgo en el que la probabilidad de responder en la categoría es máxima.

Existen dos procedimientos generales para estimar los parámetros de los ítems en el MMRG. Podemos denominarlos 'de información plena' y 'de información parcial'. En el presente trabajo hemos utilizado procedimientos de información parcial por dos razones: (1) la muestra, aunque razonable, no nos parece aún lo bastante grande como para garantizar resultados estables con el primer tipo de procedimientos, y (2) los procedimientos de información parcial se integran en el marco general de los modelos de ecuaciones estructurales y tienen mejor resuelto el problema de la evaluación del ajuste del modelo, problema de capital importancia en el presente estudio.

El procedimiento de información parcial utilizado en nuestro estudio consiste en llevar a cabo un análisis factorial no lineal de ítems y después obtener las estimaciones del MMRG mediante reparametrización. La relación entre los correspondientes parámetros puede encontrarse por ejemplo en Takane y de Leeuw, 1987. La aplicación del método se describe con mayor detalle en Ferrando (1999).

Respecto a la estimación de los parámetros de los sujetos hemos adoptado una aproximación Bayesiana y hemos utilizado como estimador puntual del nivel en el rasgo a la media de la distribución posterior de dicho rasgo dado el vector de respuestas observado (EAP en la terminología anglosajona). Además, la desviación típica de la citada distribución (PSD) sirve como error típico de medida, y nos da pues una idea de la precisión que tiene una estimación particular (véase Bock y Mislevy, 1982). Dado que la aproximación Bayesiana requiere la especificación de una distribución a priori, el estimador EAP permite obtener estimaciones razonables incluso para patrones extremos. Éste es importante cuando la escala tiene muy pocos ítems, tal como es el caso en el presente estudio.

Etapas del estudio

Estructuraremos el estudio en tres etapas. En primer lugar ajustaremos el MMRG utilizando el modelo AF confirmatorio y evaluaremos la dimensionalidad de los ítems. Más en detalle, lo que evaluaremos será el ajuste de los dos modelos propuestos en la literatura: el unidimensional bipolar y el bidimensional. En segundo lugar, una vez decidido el modelo apropiado, transformaremos los parámetros AF en parámetros TRI (discriminaciones, posiciones y valores modales) e interpretaremos dichos parámetros. Asimismo, obtendremos estimaciones de los parámetros de los sujetos (puntuaciones factoriales en terminología AF) y evaluaremos la precisión de dichos estimadores a través de diferentes niveles de rasgo. En tercer lugar, estudiaremos la validez del LOT-R con relación a las diversas medidas descritas anteriormente. En esta tercera etapa compararemos los coeficientes de validez obtenidos mediante las estimaciones TRI con aquellos obtenidos de la TCT (es decir, la suma simple de los ítems).

Método

Participantes y procedimientos

La muestra para la calibración de los ítems del LOT-R se compone de 735 estudiantes de las Facultades de Psicología y Relaciones Laborales de la Universidad Rovira i Virgili de Tarragona (521 mujeres y 214 hombres), 576 de estos participantes respondieron también a las otras medidas detalladas a continuación y son los que componen la muestra para el estudio de validez.

La administración de las diferentes medidas se llevó a cabo dentro del aula y en el horario académico de los estudiantes, siempre con el margen de tiempo necesario para permitirles contestar con total libertad y sin premuras de tiempo. En la aplicación de cada una de las pruebas se dieron las directrices oportunas para poder realizarlas correctamente. En todo momento se señaló la voluntariedad, la finalidad y el anonimato de los resultados individuales.

Instrumentos

La versión española del LOT-R que se presenta en el apéndice 1 fue desarrollada por Otero, Luengo, Romero Gómez y Castro (1998). Las medidas utilizadas en el estudio de validez fueron: el cuestionario PANAS, la escala de estrés percibido PSS y la escala de neuroticismo del cuestionario EPQ-R (Eysenck, Eysenck y Barrett (1985) en su adaptación al castellano por Aguilar, Tous y Andrés (1990). La adaptación española del PSS fue llevada a cabo por uno de los autores del trabajo (que tradujo la escala del inglés al español) y por un estadounidense, profesor de inglés, con más de 30 años de residencia en España (que llevó a cabo la traducción inversa). Finalmente, los desacuerdos se resolvieron mediante la discusión entre ambos autores. La versión española del PANAS es la de Sandín, Chorot, Lostao, Joiner, Santed y Valiente (1999).

La calibración de los ítems se llevó a cabo mediante los programas PRELIS-2 y LISREL-8 (Jöreskog y Sörbom, 1996). Los parámetros de los sujetos se estimaron mediante el programa TH-Score (Ferrando y Lorenzo, 1998).

Análisis y resultados

Análisis preliminares

Se verificó en primer lugar el comportamiento de los ítems de relleno incluidos en el LOT-R (véase el apéndice 1). La mayor parte de las correlaciones entre los 4 ítems de relleno y los 6 ítems de contenido sustantivo eran muy próximas a cero y se hallaban por debajo de 0.10 en valor absoluto. Ninguna de las correlaciones entre los dos dominios de ítems fue superior a 0.20 en valor absoluto. Estos resultados sugieren que los ítems de relleno funcionan realmente como tales, por lo que dichos ítems no fueron ya considerados en ninguno de los análisis que se presentan a continuación.

En segundo lugar se obtuvieron las puntuaciones directas en cada uno de los ítems sustantivos y se estudió su estructura mediante la aproximación lineal. La matriz de correlaciones entre los ítems se presenta en el apéndice 2. Para interpretar esta matriz debe recordarse que las puntuaciones de todos los ítems están orientadas hacia el polo de optimismo.

La matriz de correlaciones interítem fue analizada mediante un AF exploratorio lineal. Se utilizó el procedimiento de máxima verosimilitud con rotación oblicua Oblimin directa. La solución en

un factor daba saturaciones altas y positivas en todos los ítems, pero el ajuste no era aceptable (chi-cuadrado: 102.92 con 9 grados de libertad). La solución en dos factores se presenta en la Tabla 1. Esta solución daba un buen ajuste (chi-cuadrado: 5.66 con 4 grados de libertad).

La solución de la Tabla 1 es muy clara y, de hecho, se aproxima bastante al criterio de estructura simple. Cada uno de los ítems puede considerarse como una medida bastante pura de una sola dimensión. La correlación estimada entre los factores fue de 0.58.

Evaluación de la dimensionalidad

Para el análisis factorial no lineal de los ítems se utilizó el procedimiento de mínimos cuadrados generalizados de Browne adaptado al caso en que las variables son categóricas (Muthén, 1984). La tabla 2 presenta los resultados del ajuste de los dos modelos propuestos en la literatura para el LOT. De acuerdo con los criterios al uso en el dominio de los modelos de ecuaciones estructurales (véase e.g. Hu y Bentler, 1999), el modelo en un solo factor común daría un ajuste inaceptable, en tanto que el ajuste del modelo en dos factores relacionados sería razonablemente bueno. El modelo en dos factores es el más parsimonioso posible, ya que plantea que los ítems son medidas factorialmente puras del correspondiente factor. De esta forma, cada ítem sólo tiene en la práctica un índice de discriminación, ya que el otro es cero.

Dado que el modelo en un factor está anidado en el segundo modelo, se puede obtener el test jerárquico de diferencia entre ambos. El valor del test jerárquico es $\Delta\chi^2_{(1)} = 32.67$, lo que indica que se produce una mejora sustancial en el ajuste al pasar del modelo unidimensional al de dos factores. Otra forma de ver el test jerárquico es como test de hipótesis de que la correlación desatenuada entre los dos factores es unidad (véase e.g. Ferrando, 2000). Esta hipótesis debe ser claramente rechazada, una decisión que ya podía preverse, puesto que la correlación estimada entre factores fue de 0.68 (la correlación es positiva, ya que los ítems de pesimismo se puntúan en forma revertida). Cabe hacer notar que los resultados descritos hasta ahora concuerdan notablemente tanto con los resultados de la aproximación lineal descrita en el apartado anterior como con aquellos obtenidos utilizando la versión original del test (véase la sección de introducción).

Tabla 1
Patrón factorial rotado del LOT-R. Solución lineal exploratoria en dos factores

	F1	F2
o1	-0.07	0.60
o2	-0.06	0.64
o3	0.22	0.49
p1	0.35	0.04
p2	0.63	0.05
p3	0.85	-0.08

Tabla 2
Evaluación del ajuste de los modelos

Modelo	χ^2	g.l.	RMSEA	NNFI
1 factor	50.97	9	0.079	0.85
2 factores	18.29	8	0.041	0.96

En suma, el modelo que parece ser más aceptable y que se interpretará a continuación es un modelo de dos factores bastante correlacionados, aunque no equivalentes, definidos por ítems que se comportan como medidas bastante puras. Para evitar redundancias no presentaremos la parametrización AF, sino que daremos únicamente la parametrización TRI que interpretaremos en la siguiente sección.

Calibración de los ítems y estimación de los parámetros de los sujetos

La tabla 3 muestra los valores estimados de los parámetros de los ítems de acuerdo con la parametrización TRI: parámetros de discriminación (a), parámetros de dificultad (b) y valores modales (β). De acuerdo con la justificación teórica hecha anteriormente la interpretación se basará en los valores de a y β .

En la métrica que utilizamos en el presente trabajo, los valores del parámetro a suelen moverse dentro del intervalo de 0.3 a 2.0 (Hulin, Drasgow y Parsons, 1983). Sin embargo, en test de rendimiento típico no suelen alcanzarse nunca valores por encima de 1.5 (Reise y Waller, 1990). Esta baja capacidad discriminativa de los ítems parece ser un problema endémico en la medición de la personalidad.

Teniendo en cuenta estas referencias, las capacidades discriminativas de los ítems del LOT parecen bastante aceptables. En otras palabras, los ítems permiten discriminar relativamente bien entre

los sujetos que responden en una categoría y los que responden en categorías adyacentes.

Respecto a los valores modales, hay varios puntos que merecen ser comentados: (1) todos los ítems están centrados en valores negativos y no en la media del rasgo (es decir, en $\theta = 0$); (2) en ninguno de los ítems las distancias entre valores modales son iguales, ni siquiera aproximadamente; (3) la distribución de los valores modales no es simétrica. Se extienden mucho más por el lado negativo que por el positivo. Los puntos (1) y (3) nos indican que los ítems son «fáciles». Respecto al punto (2), sugiere que sería erróneo tratar las respuestas como puntos de una escala de intervalos iguales como es práctica habitual.

Si se comparan los parámetros estimados en los ítems de ambas dimensiones no se observa ninguna diferencia importante. Por el contrario, la distribución y valores de los índices de discriminación y posición son notablemente parecidos en ambos casos. Este resultado sugiere que la división en dos factores no se debe a ningún artefacto (por ejemplo que los ítems de una dimensión sean mucho más extremos que los de la otra). Así pues, la división debería buscarse más bien en el contenido de los ítems.

La conclusión anterior merece cierta discusión sustantiva, ya que aparentemente va contra el sentido común. Después de todo, como señalan Scheier y Carver (1985), es difícil pensar que una persona pueda ser optimista y pesimista al mismo tiempo. Sin embargo, este razonamiento puede ser engañoso. Como señala Cronbach (1990), cuando critica las escalas bipolares, los opuestos lógicos no tienen por qué ser también opuestos psicológicos. Además, la incompatibilidad podría ser menor de lo que parece a primera vista. Si examinamos el contenido de los ítems en el apéndice 1, podríamos admitir que ser optimista respecto al futuro no es incompatible con no contar con que las cosas salgan al modo de uno (es decir, uno puede ser optimista y al mismo tiempo no dar por hecho que las cosas vayan a salir bien). Marshall y otros (1992) indican otro argumento: las personas pueden tener en mente objetivos próximos o lejanos y uno podría, por ejemplo, ser pesimista a corto plazo pero optimista a largo plazo.

Pasamos ahora a los parámetros de los sujetos. La distribución que se obtiene de los estimadores EAP en ambas dimensiones es asimétrica negativa, lo cual se deriva de las características de los ítems (fáciles) y del criterio de estimación (Bayesiano). Los errores de medida estimados mediante la PSD son muy similares en ambas dimensiones y, como era de esperar, se mantienen bastante constantes a lo largo de los diferentes valores del rasgo. En los valores extremos de ambos rasgos, digamos más allá de $\theta = 1.5$ en valor absoluto, los valores de PSD se mueven en torno a 0.65, en los restantes casos oscilan en torno a 0.55. Para interpretar estos valores podemos utilizar la medida de fiabilidad propuesta por Bock y Mislevy (1982): $\rho = 1 - [\text{PSD}]^2$, con lo que obtenemos una fiabilidad de alrededor de 0.58 en los extremos de ambos rasgos y de 0.70 para los restantes niveles. No es, desde luego, una gran precisión, pero no olvidemos que son medidas de personalidad y que sólo tienen 3 ítems cada una. En el caso binario, para alcanzar esta precisión en un test de aptitud se requerirían posiblemente el doble de ítems (Bock y Mislevy, 1982). De acuerdo con nuestra experiencia, en un test de personalidad harían falta tres veces más.

Análisis de validez

La tabla 4 presenta los coeficientes de validez del LOT con respecto a las medidas definidas anteriormente. A efectos de compa-

Tabla 3					
Valores estimados de los parámetros de los ítems					
Índices de discriminación (a)					
Ítem	Factor Op.		Factor Pes.		
o1	0.6471		0		
o2	0.9669		0		
o3	1.0582		0		
p1	0		0.4785		
p2	0		1.0402		
p3	0		1.3459		
Índices de dificultad o posición (b)					
Ítem	B1	b2	b3	b4	
o1	-1.7879	-0.6992	0.0119	1.4353	
o2	-2.5149	-0.9987	-0.1335	1.3813	
o3	-3.4141	-1.9306	-1.0191	0.6508	
p1	-1.3924	-0.4711	0.2594	1.2604	
p2	-2.8079	-1.4992	-0.4646	1.2799	
p3	-3.0031	-1.7321	-0.7428	1.0949	
Valores modales (β)					
Ítem	β_1	β_2	β_3	β_4	β_5
o1	-1.7879	-1.2435	-0.3437	0.7236	1.4353
o2	-2.5149	-1.7568	-0.5661	0.6239	1.3813
o3	-3.4141	-2.6723	-1.4749	-0.1841	0.6508
p1	-1.3924	-0.9318	-0.1059	0.7599	1.2604
p2	-2.8079	-2.1536	-0.9819	0.4077	1.2799
p3	-3.0031	-2.3676	-1.2374	0.1760	1.0949

ración, se presentan los coeficientes de validez utilizando las estimaciones EAP descritas en la sección anterior, y las puntuaciones totales obtenidas mediante la suma simple de las puntuaciones de los ítems que definen cada dimensión (la correlación entre las puntuaciones totales en ambas dimensiones fue de 0.45). Por último, y como referencia respecto a los resultados descritos en la introducción, se presentan también las correlaciones obtenidas utilizando la puntuación total en el LOT-R. Para interpretar los resultados de la tabla 4 debe recordarse de nuevo que todas las puntuaciones están orientadas hacia el polo de optimismo.

Medida	th(Op.)	th(Pes.)	PD(Op.)	PD(Pes.)	PD(Total)
Neuroticismo (N-EPQR)	-.404	-.359	-.392	-.357	-.439
Estrés percibido (PSS)	-.427	-.382	-.419	-.346	-.449
Afecto positivo (PANAS)	.340	.217	.325	.189	.298
Afecto negativo (PANAS)	-.331	-.329	-.324	-.311	-.374

Nota. th: estimación EAP; PD: puntuación directa

Desde el punto de vista sustantivo, los resultados de la tabla 4 son bastante claros. Los coeficientes de validez son relativamente elevados y en la dirección esperada en todos los casos. Si comparamos los coeficientes obtenidos utilizando la puntuación directa total con los resultados anteriores que siguen el mismo procedimiento (véase la sección de introducción), vemos que las correlaciones son muy similares en el caso del PANAS y del PSS y que resultan superiores en la versión española en el caso de la escala N. Así pues parece que en este aspecto la versión española se comporta al menos igual de bien que la escala original.

Un segundo aspecto que podría tener interés sustantivo es que, al considerar dos escalas por separado, estas escalas parecen tener cierta validez diferencial. Nótese que la escala de optimismo es claramente un mejor predictor que la de pesimismo con respecto a las tres primeras medidas, pero que funciona prácticamente igual que la de pesimismo con respecto al afecto negativo. La validez diferencial sugiere que la división en dos escalas no es arbitraria.

Con respecto a los resultados más metodológicos, la tabla 4 indica que los coeficientes de validez obtenidos a partir de las estimaciones EAP son sistemáticamente más elevados que aquellos obtenidos a partir de las sumas simples; sin embargo, las diferencias son muy pequeñas, posiblemente sin importancia desde un punto de vista práctico.

Discusión

Desde un punto de vista sustantivo, los presentes resultados sugieren que el LOT-R en su versión española tiene un comportamiento psicométrico bastante aceptable. La escala tiene una estructura clara con dos dimensiones parcialmente independientes. Todos los ítems tienen un aceptable poder discriminativo y, en conjunto, ambas subescalas miden con razonable precisión en un rango bastante amplio de niveles del rasgo. Por último, los coeficientes de validez con respecto a una serie de medidas relevantes son bastante elevados con relación a lo que se puede esperar. En

conjunto cabe destacar que las propiedades de la versión adaptada son muy similares a las del test original. En cuanto a las limitaciones, la escala parece ser 'fácil', es decir, muestra un cierto efecto techo. Sin embargo, este resultado podría ser debido a las características particulares del grupo en el que se han calibrado los ítems. Como dice Ramsay (2001) un tanto informalmente, 'en general, los estudiantes universitarios son gente bastante feliz'.

Si bien la versión adaptada del LOT-R tiene una estructura muy similar a la del test original, debe hacerse notar que prácticamente todos los estudios revisados basados en el test original utilizaban también muestras de estudiantes universitarios. Tan sólo un estudio de Scheier Carver y Bridges (1994) analizó el LOT-R en una muestra de universitarios y en una segunda muestra de pacientes que esperaban una operación quirúrgica. La estructura del test resultó ser muy similar en ambas muestras. Este resultado sugiere que la estructura del LOT-R es generalizable a otras muestras, pero se necesita mucha más evidencia para poder afirmarlo.

En lo que refiere a los aspectos metodológicos, el uso del MMRG presenta desde luego ciertas ventajas. En primer lugar, el modelo que se propone no es meramente descriptivo, sino que puede ponerse a prueba y ser rechazado por los datos. De esta forma, hemos podido poner a prueba los modelos en uno y dos factores e incluso hemos podido usar un test jerárquico entre ambos obteniendo resultados bastante claros. En segundo lugar, hemos podido estudiar con detalle las características de los ítems. En particular, hemos visto que tanto las discriminaciones como los valores modales son muy similares en ambas subescalas, lo cual sugiere que los dos factores del LOT-R no son artificiales. Por último, el MMRG nos ha permitido evaluar la precisión con que el test estima los niveles de los sujetos en distintos valores del rasgo.

Un punto en el que el uso del relativamente complejo MMRG no se traduce en grandes ventajas es en la validez. Los coeficientes obtenidos utilizando las estimaciones EAP o las puntuaciones directas son prácticamente los mismos. Ciertamente, aquellos obtenidos desde las puntuaciones EAP son ligeramente superiores, y quizá podrían aumentar algo más si también las otras medidas se calibraran mediante TRI. Sin embargo, nos tememos que las mejores prácticas seguirían siendo insignificantes. Tal y como se desprende del trabajo de Mokken y sus seguidores, si la relación entre los niveles en el rasgo y las puntuaciones en los ítems es monótona (creciente), entonces la puntuación directa tiene prácticamente toda la información necesaria para ordenar a los sujetos según sus niveles en el rasgo.

En conclusión, de los resultados que aquí se presentan se desprende que la adaptación española del LOT-R podría utilizarse ya en investigación empírica en personalidad. En estudios de tipo correlacional se podrían obtener puntuaciones en las dos subescalas mediante suma simple de los correspondientes ítems (como hemos visto, este procedimiento resulta ya suficiente), y estudiar las relaciones con medidas relevantes. Consideramos que este tipo de estudios es importante, y que sería de interés evaluar si las dos dimensiones que aquí hemos separado tienen validez diferencial con respecto a algún tipo de medida que no hemos considerado aquí (quizás extraversión). En cuanto a posibles estudios de tipo clínico, antes sería necesario posiblemente elaborar baremos con respecto a diferentes poblaciones de interés.

Apéndice 1:

Adaptación al castellano del LOT-R (Otero, Luengo, Romero, Gómez y Castro, 1998).

1. En tiempos difíciles, suelo esperar lo mejor (o1)
2. Me resulta fácil relajarme R
3. Si algo malo me tiene que pasar, estoy seguro de que me pasará (p1)
4. Siempre soy optimista en cuanto al futuro (o2)
5. Disfruto un montón de mis amistades R
6. Para mí es importante estar siempre ocupado R
7. Rara vez espero que las cosas salgan a mi manera (p2)
8. No me disgusto fácilmente R
9. Casi nunca cuento con que me sucedan cosas buenas (p3)
10. En general, espero que me ocurran más cosas buenas que malas (o3)

Apéndice 2:

Matriz de correlaciones entre los ítems del LOT-R

	o1	o2	o3	p1	p2	p3
o1	1	0.3707	0.3338	0.1673	0.1957	0.1865
o2	0.3707	1	0.4234	0.1539	0.3121	0.308
o3	0.3338	0.4234	1	0.193	0.33	0.3873
p1	0.1673	0.1539	0.193	1	0.2696	0.3004
p2	0.1957	0.3121	0.33	0.2696	1	0.524
p3	0.1865	0.308	0.3873	0.3004	0.524	1

Referencias

- Aguilar, A., Tous, J.M. y Andrés, A. (1990). Adaptación y estudio psicométrico del EPQ-R. *Anuario de Psicología*, 46, 101-118.
- Ahrens, A.H. y Haaga, D.A. (1993). The specificity of attributional style and expectations to positive and negative affectivity, depression, and anxiety. *Cognitive Therapy and Research*, 17, 93-98.
- Anderson, G. (1996). The benefits of optimism: a meta-analytic review of the life orientation test. *Personality and Individual Differences*, 21, 719-725.
- Avia, M.D. y Sánchez, M.L. (1995). *Personalidad: aspectos cognitivos y sociales*. Madrid: Pirámide.
- Baker, F.B. (1992). *Item response theory. Parameter estimation techniques*. New York: Marcel Dekker.
- Bentler, P.M. (1995). *EQS Structural equations program manual*. Encino: Multivariate Software.
- Bock, R.D. y Mislevy, R. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- Boland, A. y Cappeliez, P. (1997). Optimism and neuroticism as predictors of coping and adaptation in older women. *Personality and Individual Differences*, 22, 909-919.
- Chang, E.C. (1998a). Dispositional optimism and primary and secondary appraisal of a stressor: controlling for confounding influences and relations to coping and psychological and physical adjustment. *Journal of Personality and Social Psychology*, 74, 1.109-1.120.
- Chang, E.C. (1998b). Distinguishing between optimism and pessimism: a second look at the 'optimism-neuroticism' hypothesis. En R.R. Hoftman, M.F. Sherrick y J.S. Warm (Eds.), *Viewing psychology as a whole: the integrative science of William N. Dember* (pp. 415-432). Washington: American Psychological Association.
- Chang, E.C., D'Zurilla, T.J. y Maydeu-Olivares, A. (1994). Assessing the dimensionality of optimism and pessimism using a multimeasure approach. *Cognitive Therapy and Research*, 18, 143-160.
- Chang, E.C., Rand, K.L. y Strunk, D.R. (2000). Optimism and risk for job burnout among working college students: stress as a mediator. *Personality and Individual Differences*, 29, 2, 255-264.
- Cohen, S., Kamarch, T. y Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 386-396.
- Cronbach, L.J. (1990). *Essentials of psychological testing* (5ª edición). New York: HarperCollins
- Dember, W.N., Martin, S.H., Hummer, M.R., Howe, S.R. y Melton, R.S. (1989). The measurement of optimism and pessimism. *Current psychology: Research and Reviews*, 8, 102-119.
- Eysenck, S.B.G., Eysenck, H.J. y Barrett, P.T. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, 6, 21-29.
- Ferrando, P.J. (1999). Likert scaling using continuous, censored and graded response models: effects on criterion-related validity. *Applied Psychological Measurement*, 23, 161-175.
- Ferrando, P.J. (2000). Testing the equivalence among different item response formats in personality measurement: a structural equation modeling approach. *Structural Equation Modeling*, 7, 271-286.
- Ferrando, P.J. y Lorenzo, U. (1998). THSCORE: A program for obtaining ability estimates under different psychometric models. *Educational and Psychological Measurement*, 58, 841-845.
- Hampson, S.E. (1986). *La construcción de la personalidad*. Barcelona: Paidós.
- Hu, L. y Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hulin, C.L., Drasgow, F. y Parsons, C.K. (1983) *Item response theory. Application to psychological measurement*. Homewood: Dow Jones-Irvin.
- Jöreskog, K.G. y Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software.
- Jöreskog, K.G. y Sörbom, D. (1996). *PRELIS 2: User's Reference Guide*. Chicago: Scientific Software.
- Loehlin, J.C. (1992). *Latent variable models*. Hillsdale, L.E.A.
- Marshall, G., Wortman, C., Kusulas, J., Hervig, L. y Vickers, R. (1992). Distinguishing optimism from pessimism: Relation to fundamental dimensions of Mood and Personality. *Journal of Personality and Social Psychology*, 62(6), 1.067-1.074.
- 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(5), 729-738.
- Mroczek, D.K., Spiro III, A., Aldwin, C.M., Ozer, D.J. y Bosse, R. (1993). Construct validation of optimism and pessimism in older men: Findings from the normative aging study. *Health Psychology*, 12, 406-409.
- Muñiz, J. (1997). *Teoría de respuesta a los ítems (2ª ed.)*. Madrid: Pirámide
- Muraki, E. y Carlson, J.E. (1995). Full-Information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19(1), 73-90.
- Muthen, B. (1984). A general structural equation model with dichotomous, ordered, categorical and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Myers, L.B. y Steed, L. (1999). The relationship between dispositional optimism, dispositional pessimism, repressive coping and trait anxiety. *Personality and Individual Differences*, 27(6), 1.261-1.272.
- Olason, D.T. y Roger, D. (2001). Optimism, pessimism and «fighting spirit»: a new approach to assessign expectancy and adaptation. *Personality and Individual Differences*, 31, 755-768.

- Otero, J.M., Luengo, A., Romero, E. Gómez, J.A. y Castro, C. (1998). *Psicología de personalidad. Manual de prácticas*. Barcelona: Ariel Practicum.
- Peterson, C., Seligman, M.E.P. y Vaillant, G.E. (1988). Pessimistic explanatory style is a risk factor for physical illness: A third-five year longitudinal study. *Journal of Personality and Social Psychology*, 55, 23-27.
- Plomin, R., Scheier, M.F., Bergeman, C.S., Pedersen, N.L., Nesselroade, J.R. y McClearn, G.E. (1992). Optimism, pessimism and mental health: a twin/adoption study. *Personality and Individual Differences*, 13, 921-930.
- Ramsay, J.O. (2001). *TestGraf: a program for graphical analysis of multiple choice test and questionnaire data*. Montreal: McGill University.
- Reise, S.P. y Waller, N.G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14, 45-58.
- Robinson-Whelen, S., Kim, C., MacCallum, R y Keicolt-Glaser, J.K. (1997). Distinguishing optimism from pessimism in older adults: is it more important to be optimism or not to be pessimism? *Journal of Personality and Social Psychology*, 73, 1.345-1.353.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 4.2.
- Samejima, F. (1997). Graded response model. En: W.J. van der Linden y R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.
- Sandín, B., Chorot, P., Lostao, L., Joiner, T., Santed, M. y Valiente, R.M. (1999). Escala PANAS de afecto positivo y negativo: Validación factorial y convergencia transcultural. *Psicothema*, 11(1), 37-51.
- Scheier, M. y Carver, C.S. (1985). Optimism, Coping, and Health: Assessment and Implication of Generalized Outcome Expectancies. *Health Psychology*, 4, 219-247.
- Scheier, M., Weintraub, J. y Carver, C.S. (1986). Coping with stress: Divergent strategies of optimists and pessimists. *Journal of Personality and Social Psychology*, 51, 1.257-1.264.
- Scheier, 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, 1.063-1.078.
- Sweetman, M.E., Munz, D.C. y Wheeler, R.J. (1993). Optimism, hardiness, and explanatory style as predictors of general well-being among attorneys. *Social Indicators Research*, 29, 153-161.
- Takane, Y. y de Leeuw, J. (1987). On the relationship between Item Response Theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393-408.
- Vickers, K.S. y Vogeltanz, N.D. (2000). Dispositional optimism as a predictor of depressive symptoms over time. *Personality and Individual Differences*, 28(2), 259-272.
- Watson, D., Clark, L. y Tellegen, A. (1984). Cross-cultural convergence in the structure of mood: A Japanese replication and a comparison with U.S. finding. *Journal of Personality and Social Psychology*, 47, 127-144.