

METODOLOGÍA

Evaluación TRI de la escala I-E de Rotter: un nuevo enfoque y algunas consideraciones

Pere Joan Ferrando, Josep Demestre, Cristina Anguiano-Carrasco y Eliseo Chico
Universidad Rovira i Virgili

Se evalúa el funcionamiento y propiedades de medida de la versión española de la escala de Locus de Control de Rotter; dicha evaluación utiliza un nuevo enfoque basado en la teoría de respuesta al ítem. Nuestro procedimiento permite: (a) detectar el impacto de los sesgos de respuesta (acquiescencia y deseabilidad social); (b) evaluar de forma clara la dimensionalidad y estructura del test; (c) evaluar las posibles causas del deficiente funcionamiento de algunos ítems; y (d) incrementar la precisión de la medida. Globalmente, los resultados no contradicen lo obtenido en estudios previos, pero permiten matizar algunas de las críticas que ha recibido la escala. Se proponen algunas recomendaciones para su uso.

An IRT assessment of Rotter's I-E scale: A new approach and some considerations. This study assessed the functioning and measurement properties of Rotter's Locus of Control Scale in its Spanish version by using a new approach based on Item Response Theory. Our procedure allows us: (a) to detect the impact of response biases (acquiescence and social desirability); (b) to evaluate clearly the test dimensionality and structure; (c) to evaluate and understand the causes of some items' inefficiency; and (d) to increase measurement precision. Globally, the results do not contradict those obtained in previous research, but allow some of the criticisms the scale had received to be qualified. Some recommendations for use are proposed.

Durante las tres décadas que siguieron a su publicación, la escala de Locus de Control Interno-Externo (E I-E) de Rotter (1966) fue el instrumento de referencia para medir dicha dimensión de personalidad (Furnham, 1993; Lefcourt, 1991; Spector, 1988). Aunque actualmente las tendencias se dirigen más hacia la medida multidimensional en dominios más específicos (Furnham, 1993; Lefcourt, 1991), dicha escala se sigue usando ampliamente (e.g., Holland, Geraghty y Shah, 2010; Pérez-García, Sanjuán y Bermúdez, 2002) y todavía se siguen investigando sus propiedades básicas (e.g., Beretvas, Suizzo, Durham y Yarnell, 2008). Paralelamente a su masiva utilización, la escala ha sido duramente criticada tanto por sus características psicométricas como itemétricas (forma, enunciado y contenido de los ítems). Esta situación se puede explicar, quizá, por la utilidad que parece tener, aun con todas sus posibles deficiencias internas, en una amplia variedad

de situaciones y contextos (Furnham, 1993; Lefcourt, 1982, 1991; Nowicki y Duke, 1974; Rotter, 1990; Spector, 1988).

Tras 44 años de análisis, cabría pensar que los nuevos estudios sobre la escala poco pueden ya aportar. Sin embargo, los autores consideramos que la E I-E posee unas peculiaridades muy interesantes que no se han tenido en cuenta en las numerosas críticas que ha recibido. El presente trabajo sí que las tiene en cuenta, y pretende re-evaluar estas viejas críticas utilizando un nuevo enfoque basado en modelos y procedimientos de análisis no utilizados hasta ahora. Del conocimiento de las propiedades y limitaciones de la E I-E se espera que puedan derivarse una serie de recomendaciones para su uso.

La principal peculiaridad itemétrica de la E I-E es que intenta medir una dimensión de personalidad utilizando ítems de actitud, referidos, en su mayor parte, a opiniones y creencias de tipo general. La crítica más habitual que han recibido sus enunciados es que requieren un elevado nivel de comprensión lectora, del que podrían carecer los individuos con bajo nivel cultural (Nowicki y Duke, 1974). También han sido considerados como potencialmente afectables por la deseabilidad social (DS. Cone, 1971; Kes-tenbaum, 1976). Sin embargo, los 23 ítems que forman la versión final (véase tabla 1) fueron seleccionados a partir de un conjunto

Tabla 1
Ítems del Cuestionario del «Lugar de Control» (Tous, 1984)

Dim.	Nº ítem		Contenido
S. G.	1	A	La mayoría de las cosas tristes que suceden en la vida de las personas son debidas a la mala suerte
		B	Las desgracias que le suceden a la gente son debidas a los errores que ellos cometen
C. P.	2	A	Una de las principales razones por las cuales tenemos guerras es porque la gente no se toma el interés suficiente por la política
		B	Por mucho que se esfuerce la gente para impedir las siempre habrá guerras
C. I.	3	A	A la larga la gente consigue obtener en el mundo el respeto que se merece
		B	Desgraciadamente por más que se esfuerce el mérito de un individuo pasa a menudo desapercibido
C. A.	4	A	La idea de que los profesores son injustos con los estudiantes es una tontería
		B	La mayoría de los estudiantes no se dan cuenta de hasta qué punto sus puntuaciones están influidas por hechos fortuitos
E. I. P.	5	A	Sin las oportunidades adecuadas uno no puede ser un líder eficaz
		B	Las personas competentes que no consiguen ser líderes es porque no han aprovechado las oportunidades que han tenido
C. I.	6	A	Por mucho que te esfuerces, algunas personas no te tragan
		B	Las personas que no consiguen agradar a los demás es que no saben tratar a la gente
E. I. P.	7	A	He encontrado frecuentemente que lo que tiene que suceder sucede
		B	Confiar en el azar nunca es bueno
C. A.	8	A	En el caso de un estudiante bien preparado no existen nunca los exámenes injustos
		B	A menudo las preguntas de los exámenes tienden a tener tan poco que ver con el trabajo del curso que es inútil estudiar
E. I. P.	9	A	Conseguir el éxito es una cuestión de mucho trabajo, el factor suerte tiene poco o nada que ver con ello
		B	Conseguir un buen trabajo depende principalmente en estar en el lugar adecuado en el momento justo
C. P.	10	A	El ciudadano corriente puede tener influencia en las decisiones del gobierno
		B	Este mundo está gobernado por las pocas personas que están en el poder y el ciudadano de a pie no puede hacer mucho
E. I. P.	11	A	Cuando hago planes tengo la casi certeza de que saldrán bien
		B	No es siempre acertado hacer planes a largo plazo porque en cualquier caso muchas cosas dependen de la buena o mala suerte
S. G.	12	A	En mi caso el conseguir lo que quiero no tiene nada que ver con la suerte
		B	Muchas veces podemos decidir lo que vamos a hacer tirando una moneda al aire
S. G.	13	A	El conseguir ser el jefe muchas veces depende de quien tuvo la suerte de llegar primero al lugar adecuado
		B	El que la gente consiga que las cosas les salgan bien depende de su pericia, la suerte no tiene nada que ver con ello
C. P.	14	A	En lo que concierne a los asuntos mundiales, la mayoría de nosotros somos víctimas de fuerzas que no podemos entender ni controlar
		B	Las personas pueden controlar los asuntos mundiales tomando parte activa en los asuntos políticos y sociales
S. G.	15	A	La mayoría de la gente no sabe hasta qué punto sus vidas están sujetas a hechos fortuitos
		B	Realmente la suerte no existe
C. I.	16	A	Es difícil saber si le caemos bien o mal a una persona
		B	La cantidad de amigos que tengas depende de lo amable que seas
S. G.	17	A	A la larga las cosas malas que nos suceden están compensadas por las buenas
		B	La mayoría de las desgracias son el resultado de la falta de habilidad, ignorancia, pereza, o de las tres cosas juntas
C. P.	18	A	Con bastante esfuerzo podemos erradicar la corrupción política
		B	Es difícil que las personas tengan mucho control sobre las cosas que deciden los políticos
C. A.	19	A	Algunas veces no puedo comprender cómo los profesores alcanzan los niveles que imparten
		B	Existe una conexión directa entre lo que estudio y las puntuaciones que alcanzo
S. G.	20	A	Muchas veces tengo la sensación de que tengo poca influencia sobre las cosas que me suceden
		B	Para mí es imposible creer que la suerte o el azar juegan un importante papel en mi vida
C. I.	21	A	Las personas se sienten solas porque no intentan ser amigables
		B	No sirve de mucho esforzarse en agradar a la gente, si les gustas, les gustas
E. I. P.	22	A	Lo que me sucede a mí depende de mis propias acciones
		B	Algunas veces tengo la sensación de que no tengo el control suficiente sobre el rumbo que está tomando mi vida
C. P.	23	A	La mayoría de las veces no puedo comprender por qué los políticos se comportan de la manera que lo hacen
		B	A la larga, las personas son las responsables de que exista un mal gobierno a nivel local

inicial de 100, eliminando aquellos que correlacionaban sustancialmente con la escala de Marlowe-Crowne (Crowne y Marlowe, 1960; véase Rotter, 1966).

En nuestra opinión, el sesgo potencialmente más importante en la E I-E no es la DS (posiblemente ya controlada por la selección de ítems), sino la respuesta aquiescente (RA). La tendencia a la RA parece ser más fuerte en ítems de actitud que en ítems típicos de personalidad (Paulhus, 1991). Pero además, hay indicios de que la RA tiende a manifestarse sobre todo en ítems largos, verbalmente complejos (Condon, Ferrando y Demestre, 2006) y que presentan enunciados con pocas referencias personales (Schuman y Presser, 1981). Para controlar la RA, la E I-E utiliza un balanceo casi completo: en 12 de los ítems, el primer enunciado mide en la dirección de externalidad y en los 11 restantes en la dirección de internalidad.

Desde un punto de vista psicométrico, la principal peculiaridad de la E I-E es que utiliza un formato normativo de elección forzosa (Hicks, 1970) para medir una dimensión bipolar. Cada ítem contiene dos enunciados, de los que uno se supone que está más cerca del polo externo y el otro del polo interno, y el respondiente debe decidir con cuál de los dos está más de acuerdo. Aunque no está claro que los modelos convencionales sean directamente aplicables a este tipo de ítems, la práctica totalidad de los análisis de la E I-E se han basado en la teoría clásica del test o en el análisis factorial (AF) lineal. Las críticas han sido principalmente dos: baja fiabilidad y multidimensionalidad (e.g., Beretvas et al., 2008; Furnham, 1993; Lefcourt, 1991; Marsh y Richards, 1987).

Con respecto a la primera crítica, la E I-E se construyó como un instrumento 'de banda ancha', cuyo objetivo era maximizar la representatividad de sus ítems, intentando que muestrearan los distintos aspectos del rasgo a medir (Rotter, 1966, 1975). Con este enfoque se minimiza la redundancia, pero, en contrapartida, la consistencia interna no puede ser elevada. (e.g., Marsh y Richards, 1987). No parece, por tanto, demasiado correcto evaluar la utilidad psicométrica de la E I-E basándose en criterios de consistencia interna. Dado que, además, la escala solo tiene 23 ítems, no puede esperarse que sus puntuaciones alcancen una elevada precisión de medida. En nuestra opinión, el aspecto clave a evaluar es si las puntuaciones alcanzan el grado de precisión requerido para las finalidades con que se van a utilizar.

Además de la amplitud de banda, Gurin, Gurin, Lao y Beattie (1969) y Tyler y Gatz (1977) plantearon una distinción entre ítems personales e impersonales. En los primeros, los enunciados se refieren a uno mismo y vienen expresados habitualmente en primera persona. En los segundos, los enunciados se refieren a los demás y suelen expresarse en tercera persona. Gurin y Tyler consideraron que la escasa precisión de algunos ítems se debía a su carácter impersonal.

Con respecto a la dimensionalidad, y también de forma intencional, los ítems de la E I-E se refieren a una serie de dominios bien diferenciados que pueden evaluarse 'a priori' mediante un análisis de contenido (Marsh y Richards, 1987; Rotter, 1990). Como es de esperar, los numerosos AFs a que ha sido sometida consiguen separar habitualmente estos dominios como factores diferenciados (véanse e.g., Furnham, 1993; Lefcourt, 1982, 1991; Marsh y Richards, 1987). Sin embargo, la utilidad de separarlos es discutible, ya que las subescalas resultantes tienen un excesivo error de medida, al ser muy cortas y con baja consistencia interna (Lefcourt, 1991). En nuestra opinión, el problema no es el de decidir si la E I-E es unidimensional o multidimensional, sino si es lo suficiente-

mente unidimensional como para que las puntuaciones totales sean interpretables.

En conjunto, y de acuerdo con la discusión hecha hasta ahora, los tres puntos a evaluar en nuestro estudio son: (a) el impacto de los sesgos de respuesta (particularmente RA); (b) la dimensionalidad; y (c) la precisión de la medida. El estudio es secuencial y está organizado en tres etapas.

El aspecto inicial más básico es la elección del modelo psicométrico. Ferrando (2006; Ferrando, Anguiano-Carrasco y Chico, 2010) ha propuesto para los ítems normativos de elección forzosa un mecanismo de respuesta que da lugar a un modelo de Teoría de Respuesta al Ítem (TRI). Dicho modelo es una versión del modelo de ojiva normal de dos parámetros (MON2P) en el que el parámetro de posición del ítem se interpreta como el punto medio entre las posiciones de los dos enunciados. En este estudio, el MON2P se ajusta mediante procedimientos de AF de ítems basados en la matriz de correlaciones tetracóricas (AFI, véase e.g., Muthén, 1978). Como procedimiento específico de AF se emplea el de rango mínimo (AFRM; Ten Berge y Kiers, 1991), que permite una separación rigurosa entre varianza de error (error de medida) y varianza común.

La primera etapa empieza con una evaluación preliminar del posible impacto de la DS sobre los ítems y las puntuaciones totales de la E I-E. El estudio principal, el que pretende evaluar el impacto de la RA, se lleva a cabo mediante el procedimiento propuesto por Ferrando, Lorenzo-Seva y Chico (2009) para escalas parcialmente balanceadas, en el cual la tendencia a la RA se modela como un factor dentro del MON2P. Conceptualmente el procedimiento puede verse como un AF secuencial en dos tiempos. En el primero se estima el factor de RA y su influencia se elimina de la matriz de datos/correlación. En el segundo se evalúa la unidimensionalidad de la escala a partir de la matriz depurada.

La evaluación de la unidimensionalidad esencial se basa en dos procedimientos. En el primero se evalúa el factor principal en una solución canónica. En una escala esencialmente unidimensional deberían obtenerse 4 resultados básicos: (a) un ajuste aceptable; (b) los signos de los pesos deberían reflejar la dirección de medida de los ítems; (c) la magnitud de los pesos debería ser consistentemente alta para todos los ítems; y (d) el factor debería explicar una elevada proporción de la varianza común.

En el segundo procedimiento, el factor general se modela como un factor de segundo orden que se estima a dos niveles. En el primer nivel se lleva a cabo un AF semiconfirmatorio de primer orden para evaluar hasta qué punto los datos apoyan la estructura 'a priori' en 5 dominios propuesta por Marsh y Richards (1987). Si el ajuste es aceptable, la matriz de correlación entre factores se analiza en el segundo nivel para obtener un factor de segundo orden. En caso de unidimensionalidad esencial la matriz de correlación entre factores debería ajustarse bien al modelo de un solo factor común (el factor de segundo orden).

No tenemos noticias de que, hasta la fecha, se haya evaluado la unidimensionalidad de la E I-E mediante la doble estrategia propuesta. Pero además, aportamos dos nuevas contribuciones: (a) se parte de una matriz 'limpia' de sesgos que permite llegar a estimaciones más claras; y (b) se utiliza un modelo no lineal que evita posibles estimaciones espurias de multidimensionalidad.

En la tercera etapa, y suponiendo aceptable la hipótesis de unidimensionalidad esencial, se evalúa la precisión de la medida. Los estudios previos se han basado en la teoría clásica, que proporciona un solo índice de precisión 'marginal' (usualmente, el coefi-

ciente de fiabilidad) y que no permite evaluar la precisión con que el test mide en diferentes niveles del rasgo. En contraste, la TRI evalúa la precisión 'local' (véase Mellenbergh, 1996) con la que se mide el rasgo en diferentes niveles del mismo y que se representa mediante la función de información. La raíz cuadrada inversa del valor de información es el error típico de medida y permite evaluar la precisión con que estimamos el nivel de un individuo.

La E I-E es un instrumento de tipo general que pretende obtener medidas a través de todos los niveles de un rasgo. Si, además, se pretende medir con precisión al mayor número posible de individuos, su función de información «ideal» (e.g., Hambleton y Swaminathan, 1985) debería ser simétrica, acampanada, centrada en los valores medios del rasgo (es decir, que mida con más precisión en la zona central donde se espera encontrar más sujetos) y relativamente plana (para abarcar también las colas). Evaluaremos hasta qué punto la función de información estimada de la E I-E se aproxima a este perfil ideal, así como la precisión con que pueden hacerse estimaciones individuales a diferentes niveles.

Método

Instrumentos y medidas

Existen en nuestro país diversas adaptaciones de la E I-E que son traducciones de la escala original (e.g., Pérez-García, 1984). En este estudio se ha utilizado la adaptación llevada a cabo por Tous (1984) en la Universidad de Barcelona (tabla 1).

Participantes y procedimiento

Los análisis se basaron en un banco de datos compuesto por 1.299 estudiantes universitarios procedentes de diversas facultades. La media de edad era de 21 años y el 70% eran mujeres. En todos los casos, los cuestionarios se administraron en grupos de unos 60 alumnos. La administración fue voluntaria, anónima (solo se pedía edad y sexo) y realizada siempre por los mismos psicólogos. De esta muestra, un subgrupo de 264 alumnos completó además la escala L del cuestionario EPQ-R de Eysenck (Eysenck, Eysenck y Barret, 1985). Esta submuestra se utiliza para evaluar el impacto de la DS. Los análisis se realizaron mediante SPSS 17.0, Matlab 5, Tarrot (para obtener una solución Procusteana) y Mplus 5.2.

Resultados

Análisis preliminares

Los procedimientos utilizados para evaluar el impacto de la RA y la dimensionalidad requieren puntuar todos los ítems según la respuesta dada (sea cual sea la dirección de medida). En las dos primeras etapas se asignaba 1 punto cuando se escogía el primer enunciado y 0 cuando se escogía el segundo. Obtenidas las puntuaciones, se estimó la matriz de correlaciones tetracóricas y se comprobó que era apropiada para el AF.

Primera etapa: estudio del impacto de los sesgos de respuesta. El potencial impacto de la DS se determinó utilizando como indicador la escala Lie del EPQ-R, que, de acuerdo con Paulhus (1991), es una medida casi pura del «Manejo de Impresiones», el componente consciente de manipulación de la información ofrecida con el fin de presentar una imagen favorable. En la submuestra de 264 sujetos se evaluaron las correlaciones entre cada uno de los

ítems de la E I-E y la escala Lie, así como entre las puntuaciones directas de ambas escalas. Ninguna de las correlaciones ítem-Lie excedía de 0.20 en valor absoluto y la correlación entre las puntuaciones totales era de -0.05. Los datos son consistentes con el proceso de selección de ítems descrito antes y apoyan la hipótesis de que, administrada en condiciones neutrales, la E I-E está libre de DS.

La tabla 2 muestra los resultados del AF de los ítems (patrón factorial). Aunque la lógica de la estimación es secuencial, como se ha explicado antes, la tabla debe interpretarse como una solución única en dos factores. El primer factor, estimado mediante el procedimiento de Ferrando et al. (2009), sería el hipotético factor de aquiescencia, y se comentará aquí. El segundo sería el factor de contenido que se comenta más abajo. El ajuste del modelo en dos factores era aceptable: la raíz media cuadrática residual (RMCR) valía 0.065 y el índice gamma-GFI 0.94.

Para interpretar el primer factor se utilizaron tres criterios: el signo de los pesos, la magnitud de los mismos y la proporción de varianza común explicada por el factor. En el segundo criterio se utilizó la propuesta de McDonald (1985): hacen falta un mínimo de 3-4 variables con cargas superiores a 0.30 para definir adecuadamente un factor.

Tabla 2
Resultados de la calibración de los ítems

Item	F1(Aq.)	F2(Cont.)	Direc.	δ	α
1	0,31	0,42	E	0,16	0,46
2	0,10	-0,33	I	-1,96	0,35
3	0,30	-0,37	I	-0,75	0,40
4	0,06	-0,42	I	-0,10	0,47
5	0,32	0,34	E	-0,36	0,36
6	0,10	0,30	E	-1,20	0,31
7	0,07	0,31	E	-0,57	0,33
8	0,13	-0,44	I	1,33	0,49
9	0,15	-0,52	I	-0,29	0,61
10	0,39	-0,47	I	-1,65	0,54
11	0,32	-0,54	I	-0,67	0,64
12	-0,08	-0,53	I	1,45	0,62
13	0,15	0,62	E	0,57	0,79
14	-0,17	0,52	E	-0,99	0,61
15	0,34	0,58	E	-1,17	0,72
16	0,01	0,22	E	-0,40	0,23
17	0,30	0,14	E	-2,88	0,14
18	0,33	-0,47	I	-1,56	0,53
19	-0,18	0,52	E	0,68	0,61
20	0,21	0,66	E	-0,11	0,88
21	0,18	-0,32	I	-0,41	0,34
22	0,19	-0,48	I	0,63	0,55
23	0,02	0,35	E	-0,43	0,37
VCE	9%	38%			

Nota: F1(Aq.): factor hipotético de aquiescencia; F2(Cont.): factor general de contenido; Direc.: dirección de medida del ítem; VCE: porcentaje de varianza común explicada; δ : índice de posición TRI; α : índice de discriminación TRI

De acuerdo con Ray (1990), la tendencia a la RA en una escala de elección forzosa se manifiesta como la tendencia a escoger el primer enunciado. De ser así, los pesos se esperarían todos de signo positivo. Como se observa en la tabla 2, la mayor parte lo son, especialmente todos los considerados relevantes (por encima de 0.30). Con respecto al segundo y tercer criterios, hay 8 pesos que alcanzan o superan el valor 0.30 y la media es de 0.19. La proporción de varianza común explicada por este factor (parte inferior de la tabla) se estima en un 9%. En conjunto, se trata de un factor relativamente débil (como cabría esperar), pero no residual. Cumple con los criterios de identificación, y la magnitud de los pesos y proporción de varianza común explicada son similares a los que se obtienen habitualmente con escalas balanceadas de personalidad (Ferrando et al., 2009; Hofstee, Ten Berge y Hendricks, 1998).

Segunda etapa: Evaluación de la dimensionalidad. Como se muestra en la tabla 2, la proporción de varianza común explicada por el segundo factor se estimó en un 38%. Los pesos se muestran en la segunda columna de la tabla. Para ayudar en la interpretación, la tercera columna indica la dirección en que mide el primer enunciado del ítem.

Resulta claro que el segundo factor exhibe la perfecta bipolaridad en el patrón de signos que cabría esperar en un factor general de contenido. Además, excepto por dos ítems (16 y 17), todos los pesos son superiores a 0.30. Finalmente, la proporción de varianza común explicada por este factor es respetable. Estos resultados, junto al razonable ajuste del modelo descrito arriba, dan apoyo a la hipótesis de que, una vez parcializada la influencia de la RA, la escala mide un factor general. No es un factor «fuerte» con una elevada consistencia, pero está bien definido y se identifica con claridad.

Para el AF de segundo orden se determinaron, en primer lugar, los 5 dominios propuestos por Marsh y Richards (1987): suerte en general (SG), control político (CP), éxito vía iniciativa personal (EIP), control interpersonal (CI) y control académico (CA). Tal como discuten Marsh y Richards (1987), es poco realista proponer a este nivel una estructura totalmente pura formada por conglomerados independientes. Lo que se hizo en este primer nivel fue obtener una solución Procusteana basada en una matriz diana semiespecificada (Browne, 1972) formada mediante los ítems que 'a priori' definen cada dimensión. Como se ha dicho antes, este enfoque puede considerarse como semiconfirmatorio.

El modelo en 5 factores proporcionaba un ajuste excelente a los datos: RMCR= 0.036, GFI= 0.98. La solución Procusteana se muestra en la parte superior de la tabla 3 junto a la matriz diana. Se ajusta bien a la propuesta a priori y coincide estrechamente con la obtenida por Marsh y Richards (1987). Así pues, se procedió en el segundo nivel a estimar el modelo de un factor común (factor de segundo orden) a partir de la matriz de correlación entre los factores de primer orden.

Los pesos estimados en el factor de segundo orden se muestran en la parte inferior de la tabla 3. El ajuste era excelente: RMCR= 0.030, GFI= 0.99, lo que concuerda también con los resultados reportados por Marsh y Richards (1987). Los pesos de los factores de primer orden están todos por encima de 0.30, por lo que cabe considerar que todos ellos son indicadores del factor general. Ahora bien, una notable proporción de la varianza de los factores de primer orden es residual. De nuevo, pues, cabe concluir que el factor general está bien definido y es identificable, pero no es un factor «fuerte».

Tercera etapa: Evaluación de la precisión de la medida. Las dos últimas columnas de la tabla 2 muestran los índices de posición (δ) y de discriminación (a) de los ítems de acuerdo con el MON2P. Las respuestas se puntúan ahora en la forma convencional (puntuaciones más altas indican mayor externalidad). Así pues, los índices de posición son esencialmente transformaciones probit de las medias de los ítems y los de discriminación son todos de signo positivo.

Los índices de posición (δ) cubren de forma homogénea el rango efectivo de niveles del rasgo, como cabría esperar en un test de este tipo. Cabe notar, sin embargo, una tendencia a situarse más hacia el polo negativo (i.e., interno).

Los índices de discriminación se distribuyen homogéneamente en un rango aproximado de 0.15 a 0.85 (media 0.49). En la métrica utilizada aquí, la capacidad discriminativa típica de un ítem de personalidad está alrededor de 0.7 y se considera ya buena a partir de 1.0 (e.g., Zickar, 2002). Si se utilizan estas referencias, cabe considerar que la calidad de los ítems va de mala a mediocre, y, en particular, los ítems 16 y 17 son, con diferencia, los que peor miden. Con respecto al 17, este resultado es consistente con la literatura. Gurin et al. (1969) indicaban que las alternativas no estaban equiparadas en contenido, y proponían eliminarlo.

Tabla 3
Solución factorial de segundo orden

Ítem	Matriz rotada oblicua					Matriz diana				
	F1	F2	F3	F4	F5	SG	CP	EIP	CI	CA
1	0,06	0,10	0,42	0,05	0,23	*				
2	0,11	-0,55	-0,18	-0,28	0,22		*			
3	0,08	0,03	-0,18	-0,15	-0,24				*	
4	0,17	0,05	-0,08	-0,11	-0,34					*
5	-0,09	0,07	-0,01	0,49	0,00			*		
6	0,02	0,08	-0,05	0,49	0,09				*	
7	-0,16	0,16	0,14	-0,02	0,03			*		
8	0,02	0,02	0,21	-0,04	-0,80					*
9	0,45	-0,12	0,06	-0,03	-0,19		*	*		
10	0,06	-0,72	0,06	-0,03	-0,18					
11	0,25	-0,05	-0,57	-0,05	0,03			*		
12	0,66	0,15	-0,09	-0,03	-0,07	*				
13	-0,77	0,03	-0,14	0,05	0,11	*				
14	-0,01	0,84	-0,10	-0,11	0,17		*			
15	-0,40	0,21	0,29	-0,11	0,07	*				
16	0,15	-0,04	0,09	0,46	0,14				*	
17	0,31	0,10	0,29	0,18	0,09	*				
18	0,01	-0,64	-0,17	-0,08	0,10		*			
19	-0,02	0,03	0,01	0,08	0,66					*
20	-0,44	0,12	0,45	-0,07	0,05	*				
21	0,01	-0,14	-0,04	-0,39	-0,05				*	
22	0,26	0,05	-0,32	-0,23	-0,04			*		
23	0,12	0,42	0,17	-0,02	0,15		*			
Pesos de segundo orden										
	0,65	0,34	0,43	0,36	0,63					

La función de información en la figura 1 refleja los resultados anteriores. La forma es más que aceptable para este tipo de test. Proporciona información en el rango efectivo de valores del rasgo (de -2 a + 2), y está centrada en los valores medios del mismo, donde se sitúa el 'grueso' de la población. Aún y así, el mayor peso de la cola izquierda indica que el test es más preciso a niveles 'internos', lo cual refleja el hecho de que hay más ítems con posiciones en este polo.

La altura de la curva indica que la cantidad de información es baja, incluso en la zona donde mejor funciona el test, un resultado que refleja las contribuciones sistemáticamente bajas de todos los ítems. Para ver con detalle las implicaciones, observemos que en la zona de máxima información (alrededor de 0) ésta vale aproximadamente 3.7. El correspondiente error típico es: $1/\sqrt{3.7}=0.52$. Así pues, para un sujeto al que se estimara un nivel de 0, el intervalo de confianza al 95% sería $0 \pm 1.96 \times 0.52$ y la diferencia entre los límites superior e inferior (-1.01; 1.01) serían (en una distribución normal) 69 puntos percentilares. Incluso en la región donde mejor funciona el test, la estimación del nivel de un individuo es muy imprecisa.

Promediando el valor de información a través de los niveles del rasgo se obtiene una medida marginal de precisión interpretable como un coeficiente de fiabilidad (Lord y Novick, 1968). En nuestro caso, el valor estimado fue de 0.77. En contraste, la fiabilidad de las puntuaciones directas, estimada mediante el coeficiente alfa, fue de 0.72, y el correspondiente intervalo de confianza al 95% (0.70; 0.74). Los resultados sugieren, pues, que la precisión basada en la estimación TRI es sensiblemente superior a la estimación clásica.

Para evaluar la conjetura de Gurin-Tyler con respecto a la baja calidad de los ítems, se categorizaron los ítems según dos criterios: discriminación y grado de referencia personal. En el primer criterio se distinguieron dos niveles: altos en discriminación ($a > 0.60$) y bajos en discriminación ($a < 0.40$). En el segundo se distinguieron también dos niveles, en este caso de referencia personal: alto (redacción en primera o segunda persona) y bajo (tercera persona o sujeto elíptico). No se observaron diferencias significativas en la dirección esperada ($\chi^2 = 0.63$, $p = 0.43$).

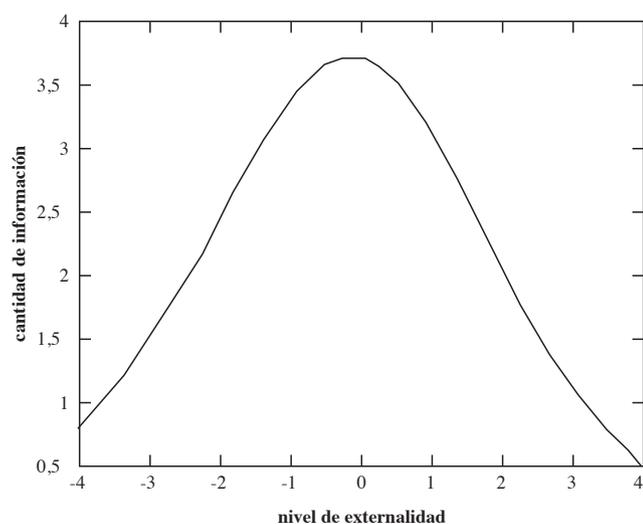


Figura 1. Función de información de la E I-E

Discusión y conclusiones

En este trabajo hemos utilizado un nuevo enfoque para evaluar las propiedades internas de la E I-E. Por una parte, se han tenido en cuenta los fundamentos y finalidades con que se desarrolló originalmente la escala, así como sus peculiaridades. Por otra, se han utilizado nuevas metodologías, teóricamente más apropiadas. Los resultados principales no chocan con los que generalmente se han reportado, pero permiten matizar ciertas críticas. Otros resultados son novedosos.

El formato de elección forzosa de la E I-E no la hace totalmente inmune a los sesgos de respuesta. Con respecto a la DS, los resultados apoyan las conclusiones de Rotter (1975) y sugieren que, cuando se administra en condiciones neutrales (las condiciones habituales en una investigación), el impacto de la DS es prácticamente inexistente. Sin embargo, el impacto de la RA parece ser similar al que se observa en cuestionarios convencionales de personalidad.

Una vez parcializada la influencia de la RA, los resultados de los AF apoyan la idea de que la escala mide un factor general. Se trata de un factor de banda ancha y relativamente débil en términos de consistencia interna. Sin embargo, puede identificarse con claridad. Conceptualmente, este resultado apoya la idea de que la escala mide una expectativa generalizada que tendería a manifestarse, en mayor o menor grado, a través de diferentes situaciones. Desde el punto de vista aplicado, justifica el uso de una puntuación total basada en el conjunto entero de ítems. Esta medida global tiene ciertas propiedades interesantes, pero también fuertes limitaciones.

Se ha encontrado que la función de información se aproxima mucho al perfil ideal que debería tener un test de este tipo. Sin embargo, la cantidad de información es sistemáticamente baja en todos los niveles del rasgo. La contribución de los ítems a la información es baja de forma bastante uniforme, y esta baja calidad no parece depender del grado de referencia personal del enunciado como se había conjeturado previamente.

Los resultados hasta ahora discutidos están necesariamente relacionados con las características de la población (i.e., estudiantes universitarios) de la que se han obtenido. Este hecho, en principio, podría considerarse como una limitación del estudio, sobre todo en cuanto a las posibilidades de generalización. A este respecto, sin embargo, debemos decir que la población 'natural' donde más se ha utilizado la E I-E es precisamente la universitaria (Lefcourt, 1982, 1991). Nótese, además, que 3 ítems son específicamente académicos. En conjunto, utilizar la misma población que la mayor parte de los estudios existentes hace más clara la comparación de resultados. Sin embargo, deben admitirse, también, las limitaciones, así como la necesidad de evaluar la escala en otros colectivos.

Por último, discutiremos las aportaciones en cuanto a recomendaciones de uso y utilidades. La E I-E no sirve para hacer evaluaciones individuales precisas, tales como las que se requieren en los ámbitos clínico o de selección. La cantidad de información es pequeña en todos los niveles y el error típico de estima individual es excesivamente grande. Dado que el contenido de los ítems es diverso y que el factor general es débil, puede anticiparse que tampoco servirá para hacer predicciones precisas en dominios específicos.

Las mismas características de la E I-E que la inhabilitan para ser un buen predictor específico pueden hacerla útil como predictor general. Se ha dicho que la E I-E es una escala extensiva pero no intensiva (Lefcourt, 1982). Muestra de forma consistente relacio-

nes con un buen número de variables pertenecientes a diferentes dominios, pero la capacidad predictiva en dominios específicos es baja. La escala será posiblemente útil en estudios de tipo exploratorio en los que interesa ver las potenciales relaciones entre el locus de control y nuevas variables. Los resultados obtenidos aquí refuerzan esta utilidad. En primer lugar, en situaciones neutrales no cabe esperar que la DS distorsione la estructura y magnitud de las relaciones. Tampoco cabe esperar que lo haga la RA dado el balanceo casi completo de la escala. En segundo lugar, la unidimensionalidad esencial justifica el empleo de una puntuación única, de

tipo general. Por último, la función de información plana indica que se podrá discriminar en un amplio rango de niveles del rasgo, y esto es una condición necesaria para poder obtener correlaciones no distorsionadas con las potenciales variables de interés. Los presentes resultados permiten hacer, además, dos recomendaciones. En primer lugar, si se calibra la escala mediante TRI, podría ser ventajoso en términos de precisión utilizar puntuaciones derivadas de la TRI en lugar de las puntuaciones directas. En segundo lugar, dado el considerable error de medida de la escala, sería conveniente corregir los coeficientes de validez por atenuación.

Referencias

- Beretvas, S.N., Suizzo, M., Durham, J.A., y Yarnell, L.M. (2008). A reliability generalization study of Scores on Rotter's and Nowicki-Strickland's Locus of Control Scales. *Educational and Psychological Measurement*, 68, 97-119.
- Browne, M.W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 207-212.
- Condon, L., Ferrando, P.J., y Demestre, J. (2006). A note on some item characteristics related to acquiescent responding. *Personality and Individual Differences*, 40, 403-407.
- Collins, B.E. (1974). Four separate components of the Rotter I-E Scale: Belief in a difficult world, a just world, a predictable world and a politically responsive world. *Journal of Personality*, 41, 471-492.
- Cone, J.D. (1971). Locus of control and social desirability. *Journal of Consulting and Clinical Psychology*, 36, 449.
- Crowne, D., y Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Eysenck, S.B.G., Eysenck, H.J., y Barrett, P.T. (1985). A revised version of the Psychoticism scale. *Personality and Individual Difference*, 6, 21-29.
- Ferrando, J.P. (2006). Two item response theory models for analysing normative forced-choice personality items. *British Journal of Mathematical and Statistical Psychology*, 59, 379-395.
- Ferrando, P.J., Lorenzo-Seva, U., y Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling*, 16, 364-381.
- Ferrando, P.J., Anguiano-Carrasco, C., y Chico, E. (2010). The impact of acquiescence on forced-choice responses: A model-based analysis. *Psicológica* (en prensa).
- Furnham, A., y Steele, H. (1993). Measuring locus of control: A critique of general, children's, health- and work-related locus of control questionnaires. *British Journal of Psychology*, 84, 443-479.
- Gurin, P., Gurin, G., Lao, R.C., y Beattie, M. (1969). Internal-external control in motivational dynamics of Negro youth. *Journal of Social Issues*, 25, 29-53.
- Hambleton, R.K., y Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.
- Hicks, L.E. (1970). Some properties of ipsative, normative and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hofstee, W.K.B., Ten Berge, J.M.F., y Hendricks, A.A.J. (1998). How to score questionnaires. *Personality and Individual Differences*, 25, 897-910.
- Holland, C., Geraghty, J., y Shah, K. (2010). Differential moderating effect of locus of control on effect of driving experience in young male and female drivers. *Personality and Individual Differences*, 48, 821-826.
- Kestebaum, J.M. (1976). Social desirability scale values of Locus of Control Scale Items. *Journal of Personality Assessment*, 40, 306-309.
- Lefcourt, H.M. (1982). *Locus of Control: Current trends in theory and research*. Hillsdale: LEA.
- Lefcourt, H. M. (1991). Locus of control. En Robinson, J.P., Shaver, P. R., y Wrightsman, L.S. (Eds.), *Measures of personality and social psychological attitudes* (pp. 413-499). New York: Academic Press.
- Lord, F.M., y Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.
- Marsh, H.W., y Richards, G.E. (1987). The multidimensionality of the Rotter I-E Scale and its Higher-order structure: An application of confirmatory factor analysis. *Multivariate Behavioral Research*, 22, 39-69.
- McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale: LEA.
- Mellenbergh, G.J. (1996). Measurement precision in test score and item response models. *Psychological Methods*, 1, 293-299.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.
- Nowicki, S., y Duke, M.P. (1978). Sex differences in locus of control and performance under competitive and cooperative conditions. *Journal of Educational Psychology*, 70, 482-486.
- Paulhus, D.L. (1991). Measurement and control of response bias. En J.P. Robinson, P.R. Shaver y L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.
- Pérez-García, A.M. (1984). Dimensionalidad del constructo «Locus de Control». *Revista de Psicología General y Aplicada*, 39, 471-788.
- Pérez-García, A.M., Sanjuán, P., y Bermúdez, J. (2002). Control personal y situacional y reactividad cardiovascular en tareas de estrés físico. *Psicothema*, 13, 583-590.
- Ray, J.J. (1990). Acquiescence and problems with forced-choice Scales. *The Journal of Social Psychology*, 130, 397-399.
- Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(609).
- Rotter, J.B. (1975). Some problems and misconceptions related to the construct of internal vs external control of reinforcement. *Journal of Consulting and Clinical Psychology*, 48, 56-67.
- Rotter, J.B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist*, 45, 489-493.
- Schuman, H., y Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question forms, wording and context*. London: Academic Press.
- Spector, P.E. (1988). Development of the Work Locus of Control Scale. *Journal of Occupational Psychology*, 61, 335-340.
- Ten Berge, J.M.F., y Kiers, H.A.L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, 56, 309-315.
- Tyler, F.B., y Gatz, M. (1977). A constructivist analysis of the Rotter I-E scale. *Journal of Personality*, 47(1), 11-35.
- Tous, J.M. (1984). *Cuestionarios del «lugar de control»*. Documento de trabajo. U. de Barcelona.
- Zickar, M. (2002). Modeling data with Polytomous Item Response Theory. En F. Drasgow y N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations*. San Francisco: Jossey-Bass.