

Análisis de ítems de un test de altruismo a partir del modelo logístico de un parámetro

*Facundo Juan Pablo Abal**

Gabriela Susana Lozzia

Diego Blum

María Silvia Galibert

María Ester Aguerri

Horacio Félix Attorresi

Resumen

El objetivo de este trabajo es mostrar los resultados obtenidos a partir de la aplicación del modelo logístico de un parámetro (ML1p) de la Teoría de Respuesta al Ítem (TRI) a un test que evalúa la predisposición a actuar de forma altruista. La prueba se compone de 18 viñetas que recrean situaciones imaginarias con formato de respuesta dicotómico. Participaron 834 estudiantes de Psicología. Previa aplicación del ML1p se verificó el supuesto de unidimensionalidad esencial de los ítems. La estimación de los parámetros se efectuó por máxima verosimilitud marginal. Siete ítems mostraron un inadecuado ajuste del ML1p. Se confirmó que las magnitudes estimadas de los parámetros de localización resultan acordes a los respectivos contenidos de los ítems. La concentración de estos parámetros en los niveles medios-bajos del rasgo y la asimetría positiva de la Función de Información del Test mostraron que la prueba carece de ítems que discriminen en la parte superior del espectro de la variable. Se discute sobre la escasa aplicación de los modelos de la TRI en los tests de personalidad y la dificultad que presenta el cumplimiento de sus exigentes supuestos. Los resultados dejaron en evidencia las debilidades de la prueba y orientaron acerca de potenciales modificaciones para su mejoramiento.

Palabras clave: Modelo logístico de un parámetro - Altruismo - Teoría de Respuesta al Ítem - Análisis de ítems

Item analysis of an Altruism scale from the One-parameter Logistic Model

Abstract

The goal of this research is to show the results collected by the One-Parameter Logistic Model (1PL) of the Item Response Theory (IRT) applied to a typical performance test. This scale assesses the willingness to behave altruistically. There are 18 items representing imaginary situations with a dichotomous response structure. 834 psychology students were the participants. Analyses based on the IRT were performed by operating the BILOG-MG software program. Previously, the essential uni-dimensionality assumption required by the model was verified. The 1PL parameters estimation was done by marginal maximum probability procedures. Seven items showed an inadequate adjustment to the 1PL. The estimated magnitudes of the location parameters match with the respective contents of the items. The concentration of these parameters in the middle-low feature levels and the positive asymmetry of the Test Information Function revealed the lack of items discriminating at the top of the variable range. The scarce application of IRT models to personality tests is discussed as well as the difficulty in achieving their strict assumptions. The outcomes showed the test weaknesses and led to possible modifications to its improvement.

Key Words: One-Parameter Logistic Model - Altruism - Item Response Theory - Item analysis.

La construcción de tests psicométricos se ha basado tradicionalmente en la Teoría Clásica de Tests (TCT) para formalizar la relación existente entre la cantidad de rasgos que un individuo posee y sus respuestas a un instrumento. La TCT se asienta en el Modelo Lineal de Puntuaciones (Spearman, 1904) que supone que el puntaje obtenido por una persona en un test es el resultado de la suma del puntaje verdadero y un error de medida. Este modelo se sostiene por un conjunto de requerimientos débiles que garantizan su aplicabilidad a la mayoría de los datos empíricos. Como afirmó Muñiz (1994), la baja exigencia de los supuestos es a la vez su fuerza y su flaqueza dado que sacrifica su capacidad predictiva por arrojar conclusiones extremadamente

generales. En la actualidad, el progreso alcanzado en la evolución de la Psicometría permitió identificar las limitaciones de esta perspectiva clásica y desarrollar teorías superadoras. Una de ellas, la Teoría de Respuesta al Ítem (TRI), ha sobresalido por el alcance de sus aportaciones a la calidad de la medición psicológica (de Ayala, 2009; Hambleton, Swaminathan & Rogers, 1991; Lord, 1980; Muñiz, 1996).

La TRI reúne un conjunto de desarrollos psicométricos provenientes de líneas de investigación independientes (e.g. Birnbaum, 1968; Rasch, 1960). El factor común de estos estudios es que parten de la premisa de que la respuesta de un individuo ante un ítem puede predecirse y explicarse a partir de una

* Instituto de Investigaciones, Facultad de Psicología, Universidad de Buenos Aires Zuviría 5691(1439). Ciudad de Buenos Aires, Argentina. Teléfono: (011) 4601-4331. E-mail: fabal@psi.uba.ar

variable inobservable llamada rasgo latente (θ). Los modelos de la TRI formalizan la relación entre el comportamiento de una persona frente a un ítem y el rasgo latente mediante una función matemática. Esta formulación brinda la probabilidad de elegir la opción-clave para cada nivel del rasgo medido. El gráfico de dicha función matemática se denomina *Curva Característica del Ítem (CCI)*.

El objetivo principal de la TRI es elaborar instrumentos psicométricos con propiedades invariantes entre poblaciones y posibilitar mediciones que no varíen aunque cambien los ítems que componen el test. Esta meta se alcanza porque la CCI queda caracterizada por sus propios parámetros independientemente de cómo se distribuya el rasgo latente en la población en la que se administra el reactivo y de cuáles sean los parámetros de los demás ítems. Además la medida de rasgo de una persona no depende ni de los parámetros de los ítems ni de la medida de los demás individuos. Estas características distintivas de la TRI son las que ayudan a encontrar respuestas a las principales limitaciones observadas en la TCT. Pero para obtener estas propiedades en la medición se requiere cumplir con un conjunto de supuestos exigentes a los que la mayoría de los datos empíricos difícilmente se acomodan, y que por ende, condicionan su aplicabilidad.

El modelo logístico de un parámetro o modelo de Rasch

En el modelo logístico de un parámetro (ML1p) el nivel θ del individuo y el grado de adhesión (b) que despierte el ítem determinan la probabilidad de que se elija la opción clave. En un ítem de personalidad se considera como clave a aquella respuesta que indica la presencia de un nivel mayor de rasgo en el individuo.

Genéricamente el nivel de adhesión es denominado parámetro de localización. Este nombre se debe a que b se mide en la misma métrica que el rasgo latente y coincide con el valor de θ necesario para tener una probabilidad de .50 de seleccionar la clave. En consecuencia, el parámetro de localización puede ser interpretado como un punto de transición (dentro de la escala del rasgo) entre la probabilidad de tomar al enunciado del ítem como no-descriptivo del evaluado y la de considerarlo como descriptivo (Panter, Swygert & Dahlstrom, 1997).

La formulación del ML1p es la siguiente:

$$P_i(\theta) = \frac{1}{1 + e^{-1.7a(\theta - b_i)}}$$

Donde:

θ es el rasgo latente que se desea medir con el ítem i ,

$P_i(\theta)$ es la probabilidad de responder la opción clave del ítem i para un nivel dado de θ .

b_i es el parámetro de adhesión o localización del ítem i .

a es el parámetro de discriminación que se estima

libremente y que, para este modelo, es el mismo en todos los ítems.

1.7 es una constante de escalamiento que se fija para que la función logística sea de mejor aproximación a la normal acumulada cuando $a=1$.

La aplicación del ML1p a tests de personalidad

A pesar de que el desarrollo formal de los modelos de la TRI se originó en el campo de la educación, en las últimas décadas se ha registrado un creciente interés de los investigadores por demostrar su utilidad para analizar ítems de tests de ejecución típica (Ozer & Reise, 1994). Son numerosas las investigaciones que han confirmado la utilidad de estudiar la respuesta a ítems de personalidad a partir de las premisas del modelo de Rasch. Carter y Wilkinson (1984) usaron el ML1p para analizar por separado cada una de las escalas clínicas de la versión original del MMPI. Los investigadores de Jong-Gierveld y Kamphius (1985) estudiaron el ajuste de una escala de Soledad y Lange y Honran (1999) modelizaron la escala AT-20 de Tolerancia a la Ambigüedad de MacDonald (1970).

El constructo Depresión recibió un particular interés por parte de los psicómetras especialistas en TRI. Hammond (1995) ajustó este modelo a los ítems del Inventario de Depresión de Beck (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) en tanto que Chang (1996) lo utilizó con la escala de Depresión del MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989). En un estudio reciente, Cadavid, Delgado y Prieto (2007) usaron la sintomatología depresiva indagada por diferentes inventarios existentes para construir una nueva escala. El ajuste de los ítems al ML1p permitió diferenciar los indicadores menos relevantes de aquellos síntomas específicos que ayudan a conseguir un diagnóstico más preciso.

Objetivo

El objetivo de este trabajo es mostrar los resultados obtenidos a partir de la aplicación del modelo logístico de un parámetro a un test de ejecución típica que evalúa la predisposición a actuar de forma altruista (Attorresi et al, 2008). En una investigación previa, Abal, Lozzia y Galibert (2008) estudiaron el ajuste de los ítems de esta escala al modelo logístico de dos parámetros. Estos autores concluyeron que, dada la baja dispersión de los índices de discriminación obtenidos, podría resultar conveniente ensayar el ajuste de un modelo más parsimonioso como el ML1p.

Método

Participantes

Colaboraron 834 estudiantes de segundo año de la carrera de Psicología de la Universidad de Buenos Aires. La edad osciló entre 18 y 69 años siendo la mediana de 20 años y la amplitud semi-intercuartil de 1.5 años. El 86% eran del sexo femenino.

Instrumento

Escala de Altruismo. Siguiendo a Batson (1991), Attorresi et al (2008) definieron la tendencia a actuar de forma altruista como un estado motivacional que tiene como objetivo principal buscar el bienestar de los otros. La escala que permite medir este constructo se compone de 18 viñetas que recrean situaciones imaginarias contextualizadas en las que los protagonistas pueden elegir actuar o no de forma altruista. Las opciones de cada uno de los reactivos son dicotómicas y contrastan la generosidad en dimensiones económicas (sacrificio económico) y morales (capacidad de perdón) versus la legítima autoconservación; el bien común versus el interés personal inmediato y el sentido de justicia o solidaridad versus la mezquindad o la comodidad.

Procedimiento de recolección de datos

Los alumnos colaboraron en administraciones grupales de forma voluntaria y anónima con el fin de favorecer la sinceridad en las respuestas. La administración del protocolo se realizó durante el horario de clase. Previamente se explicó a los participantes la finalidad de la actividad y la futura utilización de los resultados.

Análisis estadístico

El ML1p presenta un conjunto de supuestos que condicionan su aplicabilidad. Los más importantes son la unidimensionalidad de la escala y la independencia local de los ítems. El primero se verificó a partir del test de Stout que ofrece el programa DIMTEST (Stout, Nandakumar, Junker, Chang & Steindinger, 1991). Este procedimiento somete a contraste la hipótesis de que el conjunto de respuestas a ítems binarios pueden ser explicadas esencialmente por un único factor dominante. Con respecto al supuesto de independencia local, se dará por satisfecho si se comprueba la unidimensionalidad de la escala. Como demostraron Lord y Novick (1968), es posible derivar la existencia de independencia local si se cumple la unidimensionalidad del espacio latente.

La estimación del nivel θ de cada individuo y del parámetro b para cada ítem se realizó utilizando el algoritmo de máxima verosimilitud marginal que ofrece el programa informático BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996). Dado que un análisis previo de la distribución de los puntajes del test en la muestra no corroboró la normalidad de los mismos, la estimación se efectuó a partir de la distribución empírica del rasgo latente. También se estimó el valor de discriminación a constante para todos los ítems.

El mismo programa también ofreció indicadores para evaluar el ajuste del modelo a los datos. La bondad de ajuste se examinó ítem a ítem mediante un estadístico

basado en la distribución de chi-cuadrado que contrasta las frecuencias observadas con las esperadas por el modelo analizando las discrepancias en distintos niveles del rasgo latente. Asimismo, BILOG-MG permitió obtener la Función de Información de los ítems y del test, las cuales constituyen herramientas de análisis útiles para estudiar la precisión de la medida para cada nivel del rasgo latente.

Resultados

Verificación del supuesto de Unidimensionalidad

El valor del estadístico T de Stout tanto en su versión conservadora ($T=0.11$; $p=.46$) como en la más potente ($T=0.13$; $p=.45$) no permiten rechazar al 5% la hipótesis de que el conjunto de datos analizados cumple el supuesto de unidimensionalidad esencial.

Estimación de los parámetros y ajuste del modelo

La estimación de los parámetros fue realizada en 5 ciclos mediante los cuales se alcanzó, considerando un criterio de convergencia de .01, un valor

$-2\text{Log Verosimilitud} = 16646.31$. Los valores de θ estimados para los participantes quedaron distribuidos con media de 0.07 y desvío estándar de 1.47. Respecto de la capacidad discriminativa de los ítems, el ML1p estimó un valor de $a = 0.36$ ($Se = 0.017$) constante para todos los reactivos.

En la Tabla 1 se resumen los resultados obtenidos en relación con la estimación de los parámetros b de cada uno de los ítems y los valores de los estadísticos de ajuste de cada ítem al ML1p con su correspondiente valor p . Si se considera un nivel de significación de .01, siete ítems de un total de 18 ensayados no mostraron un adecuado ajuste al ML1p. Dado que esta prueba es más sensible a rechazar ante tamaños muestrales elevados como los que se utilizan en la TRI, se decidió salvar este inconveniente adoptando un nivel de significación pequeño. Aún así, si se utilizara un nivel de significación de .05 quedaría afuera, además de los ya rechazados, solamente el ítem 3.

Tabla 1. Estimación de los parámetros y ajuste al modelo de los ítems de la escala de Altruismo

Ítems	Parámetros de localización b (Se)	Ajuste al modelo	
		χ^2	g.l.
Ítem 01	-2.88 (0.16)	10.6	7
Ítem 02	-0.69 (0.13)	9.7	8
Ítem 03	-1.81 (0.14)	15.8*	6
Ítem 04	-1.78 (0.14)	6.3	7
Ítem 05	-0.23 (0.13)	11.6	8
Ítem 06	-2.16 (0.15)	18.9**	7
Ítem 07	-4.77 (0.25)	4.7	5
Ítem 08	-0.85 (0.13)	29.1**	8
Ítem 09	-0.95 (0.13)	33.9**	7
Ítem 10	-0.70 (0.13)	8	8
Ítem 11	-1.84 (0.14)	43.4**	7
Ítem 12	-3.36 (0.18)	9	7
Ítem 13	2.22 (0.15)	3.7	8
Ítem 14	-0.41 (0.13)	48**	8
Ítem 15	-0.04 (0.13)	7.3	8
Ítem 16	-0.18 (0.13)	28.1**	7
Ítem 17	-0.66 (0.13)	34.6**	8
Ítem 18	-4.92 (0.26)	6.8	5

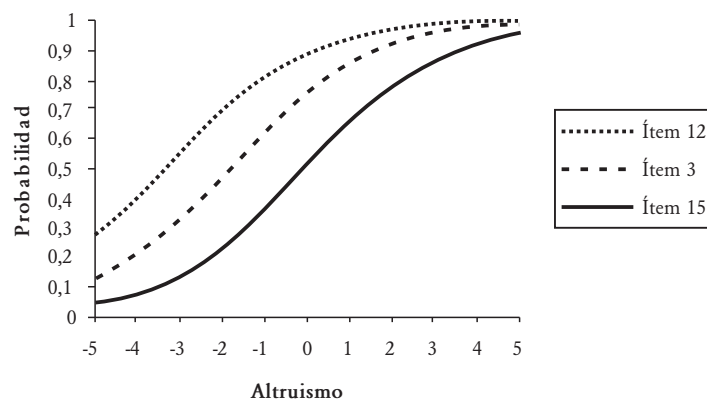
* $p < .05$ ** $p < .01$

A partir de los gráficos de ajuste que proporciona BILOG-MG se pudieron observar las discrepancias entre las CCI teóricas y empíricas. Esto mostró una tendencia prácticamente general en los ítems desajustados: para niveles elevados de Altruismo, el modelo tendió a subestimar, según se observa por los valores empíricos y viceversa para los niveles más bajos de altruismo.

El valor promedio de los parámetros b de los ítems ajustados fue de -1.72 ($SD = 2.15$). A excepción del ítem

13, cuyo valor de b fue de 2.22, el resto de los ítems obtuvo valores negativos que oscilaron entre -4.92 (ítem 18) y -0.04 (ítem 15). La interpretación de los parámetros b calibrados es equivalente a la utilizada para los ítems de habilidad en el marco de la TRI. Cuanto más elevado resulte el parámetro de localización mayor nivel de rasgo latente será necesario para optar por la respuesta clave.

Figura 1. Comparación de la Curvas Características de tres Ítems de Altruismo



A modo ilustrativo, la Figura 1 exhibe las Curvas Características de los ítems 12, 3 y 15. Como es de esperar, las tres curvas son paralelas porque el valor de discriminación estimado por el ML1p es homogéneo para todos los ítems. Si se observa el parámetro de localización del ítem 12, se puede deducir que los individuos con un nivel bajo de rasgo ($\theta = -3.36$) tendrán una probabilidad de .50 de dar la respuesta altruista. Sin embargo, se requerirá de un nivel de altruismo mayor ($\theta = -1.81$) para tener igual probabilidad de responde la clave en el ítem 3 ($b_{12} < b_3$). Finalmente, la opción altruista del ítem 15 demandará un nivel de altruismo medio ($\theta = -0.04$), superior a los otros dos ítems ($b_{12} < b_3 < b_{15}$).

Ítem 12) Me gusta el café. El gobierno está pensando en cargar con un impuesto el consumo de café para satisfacer las demandas salariales de los maestros que, de otro modo, irán a paro por tiempo indeterminado. Suponiendo que la medida sea realmente eficaz (soluciona el problema docente), soy partidario de que

- a) No se grave el café.
- b) Se grave el café.

Ítem 3) Tengo 20 años. Suelo ir a bailar los sábados por la noche y dormir hasta la tarde del domingo. Unos amigos me proponen integrarme con ellos a un grupo que se dedica a la promoción social de niños carecientes; lo cual comprometería mis domingos de 9 a 12 hs. Reconozco que la actividad es interesante.

- a) Acepto integrarme al grupo.
- b) No acepto integrarme al grupo.

Ítem 15) Soy minero en una empresa que nos explota. Estamos presionando con paro sostenido para conseguir condiciones dignas de trabajo. La situación está muy tensa, la empresa amenaza con despidos y ya ha comenzado a hacerlo pero si nos mantenemos firmes y unidos unos días más, probablemente consigamos que nos reconozcan nuestros derechos. Soy el único sostén de mi familia, con 6 hijos.

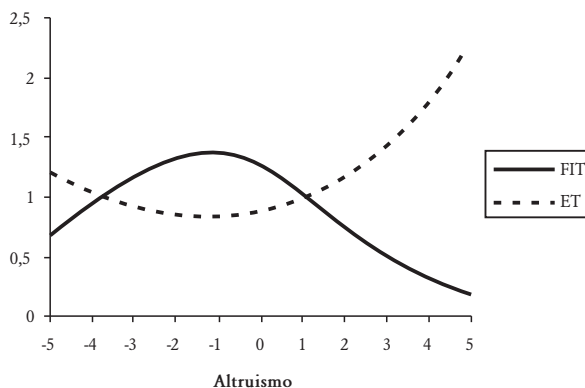
- a) Continúo adhiriendo al paro.
- b) Vuelvo a trabajar a pesar del paro.

Este ordenamiento de los ítems según el nivel de altruismo tiene un correlato conceptual si se analiza el contenido de los ítems. En la viñeta del ítem 12, un individuo que consume habitualmente café debe decidir si acepta o no un gravamen en este producto para contribuir a satisfacer las demandas salariales de los docentes. En cambio, la respuesta altruista del protagonista del ítem 3 requiere de mayor compromiso en tanto que se trata de un adolescente que decide sacrificar sus salidas de fin de semana para integrarse a un grupo de voluntarios que ayuda a niños con carencias económicas. Incluso un mayor nivel del rasgo es necesario para optar por la respuesta altruista del ítem 15. En este reactivo el protagonista es un minero con seis hijos y debe decidir si se pliega a una huelga para conseguir condiciones dignas de trabajo, a riesgo de ser despedido de la empresa. La posibilidad de optar por la respuesta altruista bajo estas circunstancias demanda un nivel mayor de θ que en los ítems anteriores dado que aquí se opone la búsqueda de un bien común a la legítima autoconservación.

Funciones de Información de los ítems y del Test

Dado que se aplicó el ML1p, la información suministrada por cada ítem es máxima para $\theta = b_i$. La cuantía de información en estas localizaciones es $(1.7a)^2/4$ y alcanzó un valor de 0.094 para todos los reactivos. El valor máximo de Función de Información del Test (FIT) fue de 1.4 y se obtuvo en un $\theta = -1.25$. Consecuentemente, el error típico de medida alcanzó un valor mínimo de 0.83 en el mismo punto (Figura 2).

Figura 2. Función de Información del Test (FIT) y Error típico (ET) de la escala de Altruismo



Discusión

El uso de la TRI en la medición de la personalidad

El estudio de la calidad de los ítems en el marco de la TCT presenta las limitaciones propias que se desprenden de utilizar indicadores globales de validez y confiabilidad. Este análisis resulta de gran utilidad pero está lejos de ser exhaustivo por lo que siempre es fructífero profundizar en el estudio del funcionamiento de los ítems. En la actualidad, han cobrado importancia las evidencias de validez interna-estructural y medidas de precisión provenientes de fuentes que toman a los ítems como unidad de análisis (Elosua, 2003). En este sentido, el uso de los modelos de la TRI se ha tornado cada vez más frecuente como soporte metodológico de los estudios instrumentales.

A pesar de tener un reconocido prestigio entre los psicómetras del mundo, resulta notorio que las aplicaciones de la TRI se concentran en un campo específico vinculado a la medición de variables de habilidades, aptitudes y rendimiento (Jones & Thissen, 2007; Reise & Henson, 2003). Una revisión reciente mostró que la TRI fue escasamente utilizada para la modelización de tests de comportamiento típico, lo cual refleja cierta desatención por parte de los teóricos de la personalidad y las actitudes sobre estos procedimientos (Morizot, Ainsworth & Reise, 2007). Las razones de estas exigüas aplicaciones son múltiples y responden tanto a factores históricos como a dificultades inherentes a la medición de este tipo de constructos (Abal, Lozzia, Aguerri, Galibert & Attorresi, 2010).

Aplicación del ML1p

El objetivo del presente artículo se inscribe en una línea de investigación psicométrica que persigue la modelización de tests de ejecución típica desde el marco de la TRI. No existe un argumento a priori que permita decidir cuál es el modelo dicotómico más adecuado para describir un conjunto específicos de datos de tests de personalidad. Algunos investigadores recomiendan el modelo logístico de dos parámetros por la dificultad de conseguir los parámetros de discriminación homogéneos necesarios para aplicar del modelo de un parámetro y por la ausencia de una respuesta azarosa que sustente la aplicación del modelo de tres parámetros (Ferrando, 1994; Reise & Waller, 1990). Sin embargo, el uso de cualquier modelo está justificado siempre que los resultados empíricos lo corroboren (Lord, 1980).

Al comparar los resultados del presente estudio con los obtenidos tras la aplicación del ML2p realizado en una investigación previa (Abal et al, 2008) no se hallaron diferencias importantes. Los mismos reactivos presentaron problemas en el ajuste para ambos modelos, salvo el ítem 3 que sólo resultó inadecuado para el ML2p. El ML1p consiguió estimaciones de los parámetros de localización similares al ML2p pero con un menor nivel de error.

La aplicación de los modelos de la TRI está

condicionada por el cumplimiento de un conjunto de fuertes supuestos que dificultan el ajuste a los datos. El análisis de los ítems del test de Altruismo mediante el ML1p mostró que siete de ellos (39% del total de reactivos) no tienen un comportamiento acorde con lo esperable como para predecir el nivel del rasgo del individuo que los responde. Esto muestra cómo la rigurosidad de los requisitos de la TRI puede dificultar la construcción de una escala dado que puede llevar a descartar una cantidad considerable de los ítems ensayados.

La importancia de examinar este ajuste se basa en que este aspecto marca una diferencia sustancial entre la TCT y la TRI. Sólo cuando el modelo se ajusta a los datos es posible suponer que la Curva Característica representa de forma apropiada la relación entre el rasgo latente y la probabilidad de dar una determinada respuesta al ítem.

Pero la decisión de descartar un ítem debe ser tomada con cierta prudencia y más aún cuando se cuenta, como en este test, con una escasa cantidad de reactivos. La evaluación de la concordancia entre los datos observados y el modelo a partir de pruebas de bondad de ajuste enfrenta el problema de la sensibilidad de los estadísticos de bondad de ajuste a los grandes tamaños muestrales requeridos desde la TRI. Asimismo, la eliminación de los siete ítems con pobre ajuste trae aparejado un fuerte déficit en la validez de contenido de la prueba. En consecuencia, resultará indispensable considerar la reformulación de los mismos o, en su defecto, se debería eliminarlos si se corroboran los mismos resultados tras sucesivas administraciones.

Interpretación de los parámetros de localización

Un resultado que es importante destacar es que la mayoría de los parámetros de localización estimados se concentraron en los niveles medios-bajos del rasgo latente. Esto es, todos los reactivos arrojaron valores de b con signos negativos (exceptuando el ítem 13 cuya interpretación será analizada más adelante). Las implicancias de este hallazgo son relevantes en tanto que revelan una debilidad en las características psicométricas de la escala de Altruismo. La misma carece de ítems que discriminen en la parte superior del espectro del rasgo, lo que hace notoria la necesidad de incorporar nuevos reactivos en futuras administraciones.

Desde la perspectiva de la TRI se requiere que los parámetros de localización de los reactivos que componen una prueba se distribuyan en todos los niveles del rasgo incluidos los extremos. Sin embargo, construir un ítem para un inventario de personalidad intentando anticipar en qué nivel del rasgo latente discriminará es una tarea bastante compleja (Morizot et al., 2007; Reise & Waller, 1990). Al respecto, los ítems calibrados en el presente estudio ofrecen una información valiosa para comprender las características del rasgo medido y para la elaboración de nuevos reactivos. Los resultados de esta administración muestran que el nivel de altruismo que se requiere para

responder la clave en el ítem 15 es uno de los más elevados. Por ende, el contenido de los futuros reactivos debe tomarse como referencia y plantear situaciones más extremas.

Una interpretación particular debe darse al resultado obtenido con el ítem 13, cuyo parámetro de localización se destaca por ser el único positivo ($b = 2.22$). En la viñeta, un adolescente es reprendido por su madre por algo que en realidad había provocado su hermano. El individuo debe elegir entre: a) callar y aceptar con resignación un castigo leve o, b) delatar la culpabilidad de su hermano. En principio, la primera alternativa fue considerada como clave dado que refleja un comportamiento abnegado (Attorresi et al., 2008). Como se mencionó anteriormente, la TRI estudia la relación que existe entre los indicadores propuestos por los ítems y la capacidad de éstos para discriminar en un valor específico del espectro del rasgo. De esta manera, permitió ordenar los ítems en función del nivel de altruismo que refleja la conducta que realiza el protagonista de la viñeta. Sin embargo, comparado con el resto de los reactivos, el contenido del ítem 13 no debería asociarse con un valor de localización tan elevado. Se consideró entonces que este resultado anómalo podría estar originado por cierta ambigüedad en el contenido del reactivo. En efecto, el hecho de que se decida por la alternativa que considera hacer manifiesta la responsabilidad del hermano puede ser entendido a la luz de una búsqueda de justicia, otro aspecto contemplado en la definición conceptual del constructo. Por ende, a pesar de haber ajustado, este ítem debe ser sometido a una reformulación que clarifique la dirección de la respuesta altruista.

La precisión de la medida

La ausencia de ítems que discriminen en los niveles más elevados del rasgo también se reflejó en la FIT. Para tests que miden constructos en población general se espera que la FIT sea relativamente simétrica respecto de $\theta = 0$ y que adopte valores de precisión aceptables en un rango razonable de la variable que oscila entre -3 y 3. Sin embargo, la marcada asimetría positiva de la FIT reveló que la prueba de Altruismo no puede ser considerada confiable para evaluar en valores medio-altos de la variable. La precisión es mayor sólo para discriminar individuos con niveles de rasgo entre -3 y 0 aproximadamente. Esto significa que la escala sólo podría utilizarse para estimar con precisión los niveles bajos de altruismo en la población considerada en tanto que para los niveles altos sería indispensable agregar

reactivos.

Aún con esta aclaración, es conveniente resaltar que la FIT obtenida adopta valores relativamente bajos, lo que implica que, incluso en su punto máximo, el error típico es considerable. Esto se explica principalmente por dos problemas que tiene esta versión preliminar de la escala. El primero responde a la baja capacidad discriminatoria de los ítems. Como ya se mencionó, la máxima cantidad de información aportada por un ítem depende del valor a en la ecuación $(1.7a)^2/4$. No obstante, incluso teniendo ítems con un valor estimado de a bajo como en el presente caso, se podría alcanzar un nivel de información más elevado tratando de agregar más ítems al proceso de medición que ajusten al modelo y que lo logren efectivamente elevar. En efecto, la FIT es el resultado de la suma de las funciones de información de todos los ítems que componen la prueba. Pero esto conduce al segundo problema de la prueba ensayada que es su corta longitud. Nuevamente, se observa aquí la necesidad de incorporar nuevos reactivos en la próxima administración.

Consideraciones finales

Las limitaciones del presente estudio responden a la etapa de construcción en la que se encuentra el test de Altruismo. Los resultados de esta aplicación están lejos de ser concluyentes. Por el contrario, los mismos sirvieron para mostrar las debilidades de la prueba y orientar acerca de las potenciales modificaciones. Es prioritario efectuar una significativa ampliación de la cantidad de ítems con contenidos que permitan discriminar en los niveles más elevados del rasgo. La incorporación de nuevos reactivos y un aumento del tamaño de muestra permitirían obtener indicadores de la invariancia de los parámetros estimados y detectar el Funcionamiento Diferencial de los Ítems. Debido a las limitaciones indicadas, ninguno de los dos estudios se realizó porque no arrojarían resultados confiables e interpretables en esta instancia preliminar.

Otra limitación importante remite a las características particulares de la muestra utilizada en el presente estudio. La homogeneidad de la muestra afecta la posibilidad de generalizar los resultados obtenidos a otras poblaciones de estudio en detrimento de la validez externa. En este sentido, es necesario también un aumento en la heterogeneidad de la muestra; lo que también permitiría obtener indicadores de la invariancia de los parámetros estimados en otras poblaciones dianas.

Referencias

- Abal, F., Lozzia, G. & Galibert, M. S. (2008). Aplicación del modelo logístico de dos parámetros en una escala de altruismo. *Memorias de las XV Jornadas de Investigación y Cuarto Encuentro de Investigadores en Psicología del Mercosur*, 2, 453 - 454.
- Abal, F., Lozzia, G., Aguerrí, M. E., Galibert, M. S. & Attorresi, H. (2010). La escasa aplicación de la Teoría de Respuesta al Ítem en Tests de Ejecución Típica. *Revista Colombiana de Psicología*, 19(1), 111 - 122.
- Attorresi, H., Abal, F., Lozzia, G., Picón Janeiro, J., Aguerrí, M. E. & Galibert, M. S. (2008). Construcción de una escala para la medición del Altruismo. *Investigando en Psicología*, 10, 7 - 24.
- Batson, D. C. (1991). *The Altruism Question: Toward A Social-Psychological Answer*. New Jersey: Lawrence Erlbaum Associates.

- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-569.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord & M. R. Novick (Eds.). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Cadavid, N., Delgado, A. R. & Prieto, G. (2007). Construcción de una escala de depresión con el modelo de Rasch. *Psicothema*, 19, 515-521.
- Carter, J. E., & Wilkinson, L. (1984). A latent trait analysis of the MMPI. *Multivariate Behavioral Research*, 19, 385-407.
- Chang, C. H. (1996). Finding two dimensions in MMPI-2 depression. *Structural Equation Modeling*, 3, 41-49.
- de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: Guilford Press.
- de Jong-Gierveld, J., & Kamphuis, F. (1985). The development of a Rasch-type loneliness scale. *Applied Psychological Measurement*, 9, 289-299.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321.
- Ferrando, P. J. (1994). Fitting response models to the EPI-A Impulsivity scale. *Educational and Psychological Measurement*, 54, 118-127.
- Hambleton R. K, Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Londres: Sage.
- Hammond, S. M. (1995). An IRT investigation of the validity of non-patient analogue research using the Beck Depression Inventory. *European Journal of Psychological Assessment*, 11, 14 - 20.
- Jones, L.V. & Thissen, D. (2007). A History and Overview of Psychometrics. En C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics, 26: Psychometrics* (pp. 1 - 27). Amsterdam: North Holland.
- Lange, R. & Honran, J. (1999). Scaling MacDonald's AT-20 using item-response theory. *Personality and Individual Differences*, 26, 467-475.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental tests scores*. Reading, MA: Addison-Wesley.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N. J.: Lawrence Erlbaum.
- MacDonald, A. P., Jr. (1970). Revised scale for ambiguity tolerance: Reliability and validity. *Psychological Reports*, 26, 791-798.
- Morizot, J., Ainsworth, A. T. & Reise, S. P. (2007). Toward Modern Psychometrics. Application of Item Response Theory Models in Personality Research. En R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology*, (pp. 407 - 423). New York: Guilford Press.
- Muñiz, J. (1994). *Teoría Clásica de Test*. Madrid: Pirámide.
- Muñiz, J. (1996). *Psicometría*. Madrid: Universitat.
- Ozer, D. J., & Reise, S. P. (1994). Personality assessment. *Annual Review of Psychology*, 45, 357-388.
- Panther, A. T., Swygert, K. A. & Dahlstrom, G. W. (1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment*, 68, 561-589.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute for Educational Research.
- Reise, S. P. & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14, 45-58.
- Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81, 93-103.
- Spearman, C. E. (1904). General Intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201 - 293.
- Stout, W., Nandakumar, R., Junker, B., Chang, H. y Steidinger, D. (1991). *DIMTEST [Computer Program]*. Champaign, IL: Department of Statistics, University of Illinois.
- Zimowski, M., Muraki, E., Mislavy, R. y Bock, R. (1996). *BILOG-MGTM: Multiple-group IRT analysis and test maintenance for binary items* [Computer program]. Chicago, IL: Scientific Software International.

Fecha de Recepción: 04-05-10

Fecha de Aceptación: 05-10-10