

V Congreso Internacional de Investigación y Práctica Profesional en Psicología  
XX Jornadas de Investigación Noveno Encuentro de Investigadores en  
Psicología del MERCOSUR. Facultad de Psicología - Universidad de Buenos  
Aires, Buenos Aires, 2013.

## **Modelos logísticos de la TRI aplicados a un test que mide el afecto hacia la matemática en estudiantes de psicología.**

Abal, Facundo Juan Pablo, Lozzia, Gabriela y  
Blum, G. Diego.

Cita:

Abal, Facundo Juan Pablo, Lozzia, Gabriela y Blum, G. Diego (2013).  
*Modelos logísticos de la TRI aplicados a un test que mide el afecto hacia  
la matemática en estudiantes de psicología. V Congreso Internacional  
de Investigación y Práctica Profesional en Psicología XX Jornadas de  
Investigación Noveno Encuentro de Investigadores en Psicología del  
MERCOSUR. Facultad de Psicología - Universidad de Buenos Aires,  
Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-054/896>

ARK: <https://n2t.net/ark:/13683/edbf/SXk>

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso  
abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su  
producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite:  
<https://www.aacademica.org>.*

# MODELOS LOGÍSTICOS DE LA TRI APLICADOS A UN TEST QUE MIDE EL AFECTO HACIA LA MATEMÁTICA EN ESTUDIANTES DE PSICOLOGÍA

Abal, Facundo Juan Pablo; Lozzia, Gabriela; Blum, G. Diego

UBACyT, Universidad de Buenos Aires - Agencia Nacional de Promoción Científica y Tecnológica

## Resumen

El objetivo de este trabajo es mostrar los resultados obtenidos al aplicar los modelos logísticos de uno, dos y tres parámetros de la Teoría de Respuesta al Ítem a la modelización psicométrica de un test que mide el Afecto hacia la Matemática en estudiantes de Psicología. Esta variable describe el interés por involucrarse en actividades vinculadas a la matemática y los sentimientos asociados al uso de sus conceptos. Participaron 1875 estudiantes de Psicología (Universidad de Buenos Aires). Previamente se verificó la condición de unidimensionalidad de los ítems requerida por los modelos. La estimación de parámetros de los modelos se realizó con el método de Máxima Verosimilitud Marginal operando el programa MULTILOG. Los parámetros de discriminación fueron elevados y los parámetros de localización se concentraron en los valores medios de la escala del rasgo. La comparación de las aplicaciones permitió concluir que el modelo logístico de 2 parámetros ofrece la estructura paramétrica que mejor representa la respuesta de los individuos a los ítems. Sin embargo, la concentración de los parámetros de localización revela que la prueba carece de ítems que discriminen en niveles bajos y altos. Se plantean las limitaciones de este estudio y propuestas para futuras investigaciones.

## Palabras clave

Teoría de Respuesta al ítem, Modelo Logístico, Afecto hacia la Matemática

## Abstract

LOGISTIC MODELS OF IRT APPLIED TO A TEST THAT MEASURES THE AFFECTION TOWARDS MATHEMATICS IN PSYCHOLOGY STUDENTS  
The aim of this research is to show the results of using of the one-two- and three-parameter logistic models of the Item Response Theory for the modeling of a test that evaluates affection towards mathematics. This variable describes the interest of psychology students to get involved in activities related to mathematics and the feelings associated with the use of its concepts. This test was administered to 1875 Psychology students from the University of Buenos Aires. The item unidimensionality assumption, as required by the models, was previously verified. The parameters estimation was performed through marginal maximum likelihood procedures by operating the MULTILOG software program. The discrimination parameters were high and the location parameters gathered around the middle of the scale of the trait. Comparison of applying the three models show that the 2-parameter logistic model provides the best results. However the merger of the location parameters near the middle levels reveals the lack of items able to discriminate at the low and top trait levels. Limitations and suggestions for future research are discussed.

## Key words

Item Response Theory, Logistic Model, Affection towards mathematics

Los modelos logísticos de la Teoría de Respuesta al Ítem establecen una relación funcional entre los niveles del rasgo latente unidimensional que se pretende medir y la probabilidad de dar la respuesta clave a un ítem dicotómico. Esta relación, denominada Curva Característica del Ítem, está definida a partir de la determinación de hasta tres parámetros ( $a$ ,  $b_i$  y  $c_i$ ) que especifican una función logística genérica y que sirven para describir el comportamiento psicométrico del ítem.

Los diferentes modelos logísticos varían en la cantidad de parámetros que contemplan. El modelo logístico de un parámetro (ML1p) establece que la probabilidad de que un individuo con un nivel de rasgo  $\theta$  responda a la clave está sólo en función del parámetro de localización  $b_i$  (Rasch, 1960). El modelo logístico de dos parámetros (ML2p) agrega al parámetro  $b_i$  un parámetro de discriminación o pendiente  $a_i$ . Por último, el modelo logístico de tres parámetros (ML3p) incluye el parámetro  $c_i$  que describe la probabilidad de que un individuo con un nivel muy bajo de rasgo responda a la respuesta clave (Birnbaum, 1968). Como supuestos adicionales, el ML1p establece que todos los ítems que componen la prueba presentan una discriminación homogénea (i.e. el parámetro  $a$  es constante) mientras que el ML1p y el ML2p establecen que  $c_i = 0$ .

Aunque tradicionalmente los modelos logísticos de la TRI fueron aplicados para la medición ítems de tests de rendimiento máximo, en las últimas décadas se ha podido observar una creciente cantidad de estudios instrumentales que utilizan esta teoría moderna en tests de comportamiento típico (Reise y Waller, 2009). Esto propició un cambio importante en la forma de denominar e interpretar los parámetros que componen los modelos logísticos (Abal, Lozzia, Aguerri, Galibert y Attorresi, 2010).

En ítems de comportamiento típico el parámetro  $b_i$  describe un punto de transición (dentro de la escala del rasgo) entre la probabilidad de tomar al enunciado como no-descriptivo del evaluado y la de considerarlo como descriptivo (Panter, Swygert y Dahlstrom, 1997). El parámetro de discriminación  $a_i$  muestra la capacidad del ítem para diferenciar a personas con un nivel de rasgo más elevado de aquéllos que tienen un nivel más bajo siempre en torno al parámetro de localización. En cambio, el parámetro  $c_i$  puede ser entendido en el marco de los tests de ejecución típica como un indicio de la incidencia de la deseabilidad social en las respuestas (Rouse, Finger y Butcher, 1999). De esta manera en los últimos años se ha comenzado a delimitar un campo de conocimiento que reconoce la especificidad de las respuestas propias de los constructos actitudinales y de la personalidad y que no se conforma con adoptar los resultados heredados de los tests de habilidades (Ferrando y Demestre, 2008).

El objetivo de este trabajo es mostrar los resultados de las aplicaciones de los ML1p, ML2p y ML3p al análisis de ítems de una prueba que mide el nivel de Afecto hacia la Matemática en estudiantes de Psicología (Abal, 2013). Como resultado de esta comparación se intenta determinar cuál es la estructura paramétrica que mejor representa la respuesta de los individuos a los ítems. La ventaja del análisis comparativo de los resultados presenta aristas teóricas y prácticas. Al utilizar los mismos datos con distintos modelos, se facilita un estudio metodológico que permite encontrar puntos de conexión entre las modelizaciones. A su vez, si se alcanzan predicciones similares a partir de formulaciones distintas, se estaría aportando una poderosa evidencia de validez de constructo.

## Metodología

### Participantes

Se contó con la colaboración de 1875 estudiantes del segundo año de la Facultad de Psicología de Universidad de Buenos Aires. El 82% fueron mujeres y el 18% fueron varones. La edad de los participantes osciló entre 18 y 62 años, con una media de 22.7 años (DE = 6.33), una mediana de 20 y una amplitud semi-intercuartil de 2 años.

### Instrumento

*Escala de Afecto hacia la Matemática* (Abal, 2013). Consta de ocho enunciados con formato de respuesta politómica de seis opciones tipo Likert (de *totalmente en desacuerdo* a *totalmente de acuerdo*). La escala mide el interés de los estudiantes por involucrarse en actividades vinculadas a la matemática y sentimientos asociados al uso de términos, símbolos y conceptos matemáticos. El instrumento cuenta con evidencias de validez y confiabilidad obtenidas en el marco de la Teoría Clásica de Tests que garantizan su calidad psicométrica.

### Procedimiento

La administración se realizó en grupos reducidos. Previamente se efectuó una charla motivadora que explicaba la finalidad de la actividad y su futura utilización en investigación. Los alumnos respondieron el cuestionario de manera anónima y voluntaria con el fin de favorecer la sinceridad en las respuestas.

### Análisis de datos

La comprobación del supuesto de unidimensionalidad se realizó con la información de los datos politómicos mediante un Análisis de Componentes Principales por ser éste el estudio más frecuentemente utilizado (Muñiz, 2010). Posteriormente los datos politómicos fueron recodificados en dos categorías para posibilitar la aplicación de los modelos dicotómicos. Con este fin, se reagruparon las tres categorías inferiores y las tres superiores. El valor más alto del código binario fue asignado conforme lo determinó la clave del ítem. La estimación de los parámetros se efectuó con el programa MULTILOG 7.03 (Thissen, 2003) utilizando el método de Máxima Verosimilitud con un criterio de convergencia de .001. Como indicador del ajuste de los datos a los modelos, el programa brinda las proporciones observadas y las esperadas de elección para ambas categorías de respuesta. En virtud de que esta información es limitada como única evidencia (Revuelta, Abad y Ponsoda, 2006), se adicionaron otros indicadores indirectos recomendados por Aguado, Rubio, Hontangas y Hernández (2005): a) la convergencia en la estimación de los parámetros utilizando una cantidad inferior a 100 iteraciones; b) la estimación de parámetros con valores acordes a lo esperable; y c) la existencia de errores estándares relativamente bajos. Si bien los ML1p y ML2p consiguieron estimaciones precisas de los parámetros en la primera corrida del programa, el ML3p no alcanzó el criterio de convergencia habiendo efectuado las 100 iteraciones. Una lectura de la salida computacional ofrecida por el

programa dejó en evidencia la dificultad para estimar el valor del parámetro  $c$ . Conforme lo sugirieron Abad, Olea, Ponsoda y García (2011), se procedió a realizar especificaciones sobre la distribución de este parámetro y esto permitió alcanzar el criterio de convergencia. Salvo estas restricciones para el ML3p, el resto de los comandos del programa MULTILOG fueron utilizados con las opciones que aparecen por defecto.

## Resultados

### Comprobación del supuesto de unidimensionalidad

La medida de adecuación muestral  $KMO$  de .94 y la prueba de esfericidad de Bartlett con un valor de 8488.3 ( $g= 28$ ;  $p < .0001$ ) confirmaron la posibilidad de aplicar un análisis de componentes principales a la matriz de datos. A partir de este estudio se obtuvo una solución con un único factor con autovalor superior a 1 que describió el 61.7% de la varianza. El cociente entre el primer y segundo autovalor resultó de 7.84. Los pesajes de los ítems resultaron adecuados y oscilaron entre .70 y .87. Estos resultados cumplen con los criterios propuestos por Kaiser (1960), Hair, Anderson, Tatham y Black (1999) y Martínez Arias (1995) para tener una aproximación aceptable al supuesto de unidimensionalidad requerido desde los modelos de la TRI.

### Calibración de los ítems.

La Tabla 1 resume la cantidad de ciclos necesarios en cada modelo para alcanzar dicho criterio y el mayor cambio registrado entre la estimación de un ciclo y el siguiente. El ML2p fue el modelo que más ciclos necesitó para alcanzar la convergencia en las estimaciones aun teniendo una cantidad de parámetros considerablemente menor que el ML3p.

Tabla 1. Características de los procesos de estimación.

	ML1p	ML2p	ML3p
Cantidad de parámetros estimados	9	16	24
Cantidad de ciclos	20	23	22
Mayor cambio registrado en la estimación	0.00085	0.00081	0.00051

Los residuos obtenidos entre las proporciones empíricas y esperadas (según la probabilidad de respuesta dada por cada uno de los modelos logísticos) fueron iguales o inferiores a .01. Esto implica que, según la información suministrada por MULTILOG, el ajuste de los datos al modelo fue satisfactorio para los ocho reactivos y para los tres modelos.

En la Tabla 2 pueden observarse los resultados obtenidos en la calibración de los ítems para cada uno de los modelos logísticos. Los errores de estimación resultaron relativamente bajos, lo que aporta evidencias para confiar en la adecuación de los valores de los parámetros alcanzados.

Tabla 2. Parámetros estimados de los modelos dicotómicos.

Ítem	ML1p		ML2p		ML3p	
	$b_i$	$a_i$	$b_i$	$a_i$	$b_i$	$c_i$
1	-0.58	1.72	-0.66	1.34	-0.30	0.21
2	0.02	1.87	0.01	1.29	0.13	0.06
3	-0.08	2.44	-0.10	1.96	0.06	0.09
4	0.69	1.51	0.82	1.05	0.86	0.04
5	-0.16	2.07	-0.18	1.40	-0.06	0.07
6	-0.10	4.41	-0.11	2.97	-0.05	0.03
7	-0.28	4.24	-0.26	3.04	-0.18	0.05
8	-0.71	3.71	-0.65	2.70	-0.52	0.09

La interpretación de los parámetros  $b_i$  calibrados es equivalente a la utilizada para los ítems de habilidad. Cuanto más elevado resulta el parámetro de localización, mayor será el nivel de Afecto hacia la Matemática necesario para optar por la clave. Por ejemplo, los tres modelos convergen en señalar que es necesario tener un nivel de Afecto más elevado para acordar con el ítem 8 (*Me siento contento/a cuando tengo que aplicar algún concepto de Matemática en otra materia*) que para estar de acuerdo con el ítem 1 (*Puedo disfrutar de una buena clase de Matemática*). De esta manera, los  $b_i$  de los modelos logísticos hacen posible ordenar los estímulos de los reactivos en función del nivel del rasgo latente que demandan y así poder corroborar si los contenidos de los indicadores responden al ordenamiento obtenido.

El valor promedio de los parámetros  $b_i$  obtenidos con el ML1p fue de  $-0.15$  ( $DE = 0.40$ ) siendo  $-0.71$  y  $0.69$  los respectivos valores mínimo (ítem 8) y máximo (ítem 4) registrados. Un promedio similar se obtuvo con los  $b_i$  del ML2p ( $-0.14$ ) con un desvío estándar ligeramente mayor ( $DE = 0.43$ ). Los  $b_i$  oscilaron entre valores razonables para una prueba de actitudes entre  $-0.66$  (ítem 1) y  $0.82$  (ítem 4). El valor promedio de los parámetros de localización del ML3p estuvo más próximo al centro de la escala de la variable  $-0.01$  ( $DE = 0.38$ ) y sus valores mínimo y máximo fueron  $-0.52$  (ítem 8) y  $0.86$  (ítem 4).

Un resultado para destacar es que la mayoría de los parámetros de localización estimados para los ML1p y ML2p se concentraron en niveles por debajo de  $\theta = 0$ . Sólo 2 de los 8 reactivos tuvieron valores de  $b$  con signos positivos en cada uno de los modelos. La situación del ML3p no es diferente dado que sólo tiene tres ítems con  $b_i$  positivos. A su vez, los mismos se ubicaron a menos de un desvío por encima de  $\theta = 0$ . Si se toman como ideales las recomendaciones de Reise y Waller (1990), tampoco el espectro inferior se encuentra debidamente representado por los ítems. Estos autores establecieron que los parámetros  $b$  deberían alcanzar valores entre  $-2$  y  $2$  para describir con razonable precisión un intervalo amplio de la variable.

Baker (2001) propuso una categorización general para interpretar la calidad del parámetro de discriminación. Según este autor, un  $a_i$  entre  $0.65$  y  $1.34$  es moderado, entre  $1.35$  y  $1.69$  es alto y a partir de  $1.7$  es muy alto. Reise y Waller (1990) especificaron que los parámetros de discriminación para pruebas de ejecución típica suelen oscilar entre  $0.5$  y  $1.5$ . A partir de los resultados obtenidos se puede afirmar que la prueba presentó una discriminación elevada con todos los modelos.

Mientras que para el ML1p el valor de  $a$  constante para todos los reactivos fue de  $2.37$ , la capacidad discriminativa del conjunto de ítems resultó más alta para el ML2p dado que el promedio de los  $a_i$  fue de  $2.75$  ( $DE = 1.11$ ). El ítem 6 evidenció el valor de  $a$  más elevado ( $a = 4.41$ ) y el ítem 4 presentó el  $a$  más bajo ( $a = 1.51$ ). Para el ML3p el valor promedio de  $a_i$  fue más bajo ( $a = 1.97$ ;  $DE = 0.43$ ), aunque también se considera elevado. Para este modelo, el ítem 4 también mostró el valor de discriminación más pequeño ( $a = 1.05$ ) pero el valor más elevado se encontró en el reactivo 7 ( $a = 3.04$ ). Un aspecto sustancial que cabe destacar es la dispersión que presentan los parámetros de discriminación para los ML2p y ML3p.

Ferrando y Demestre (2008) registraron una relación positiva entre la capacidad discriminativa y la relevancia que la conducta descrita en el ítem tiene para el constructo. Si se observan estos resultados en la Tabla 2 es posible apreciar que los ítems 6, 7 y 8 presentaron una capacidad discriminativa superior al resto, lo que demuestra que los contenidos de estos indicadores guardan una relación de mayor relevancia en la medición del Afecto hacia la Matemática.

Al comparar los valores de  $a_i$  obtenidos en el ML2p y ML3p se puede

apreciar una notable diferencia. La capacidad discriminativa de la prueba se ve afectada de manera sustancial al incorporar el parámetro  $c_i$ . Esto resulta llamativo en virtud de que la mayoría de los ítems mostraron un valor de  $c_i$  relativamente bajo. Sólo se destaca el parámetro  $c_i$  correspondiente al ítem 1. Este fue el único que superó la probabilidad de respuesta al azar de  $.17$  calculada sobre las seis opciones de respuesta Likert originales. Una interpretación posible para explicar este resultado responde a la influencia de la deseabilidad social en la respuesta al ítem. Como ya fue comentado, esta es una importante variable para justificar un aumento en el valor del parámetro  $c_i$  cuando se aplica el ML3p en ítems de ejecución típica. Sin embargo, esta interpretación es, por el momento, una especulación que sólo podrá confirmarse en ulteriores estudios.

## Conclusiones

Si bien los tres modelos logísticos de la TRI aplicados mostraron un comportamiento adecuado, es posible concluir que el ML2p ofrece la estructura paramétrica más conveniente para modelizar las respuestas de los estudiantes a la escala de Afecto hacia la Matemática. Este resultado se encuentra en consonancia con la mayoría de los estudios instrumentales que aplicaron los modelos logísticos en ítems de comportamiento típico (Ferrando, 1994, 2003 Reise y Waller, 1990).

El ML3p presentó resultados comparativamente más pobres. Debe recordarse que el proceso de estimación de los parámetros de este modelo requirió la especificación de la distribución de un parámetro asociado al parámetro  $c_i$ . Esto refleja cierta dificultad del modelo para alcanzar valores razonables en el proceso de estimación de los parámetros y, en definitiva, para alcanzar un ajuste óptimo. Sumado a esto, se ha podido apreciar que, si se compara con los parámetros del ML2p, la incorporación de este tercer parámetro también perjudica a la capacidad discriminativa de la escala porque disminuyen significativamente los parámetros  $a_i$ . Además, en términos generales se podría considerar que los ítems adoptaron valores de  $c_i$  relativamente bajos como para justificar la no incorporación de un tercer parámetro en el análisis de los ítems.

De la comparación de las aplicaciones del ML1p y el ML2p también se desprende que el supuesto de homogeneidad de los  $a_i$  repercutió negativamente en la capacidad discriminativa de la escala. Como se señaló oportunamente, la dispersión de los valores de  $a_i$  fue considerable en los modelos con parámetro de discriminación variable. En consecuencia, la modelización con el ML1p traería aparejada una reducción en la precisión con que cada ítem mide el Afecto hacia la Matemática, más aun en aquellos reactivos que demuestran tener una potencia discriminativa mayor al resto.

Ahora bien, aunque el ML2p resultó la estructura paramétrica más adecuada para representar las respuestas de los estudiantes, también es válido mostrar que presentó limitaciones. Al observar los valores promedio de los parámetros  $b_i$  correspondientes a los tres modelos se puede apreciar que se encontraron próximos al centro de la escala del rasgo latente. Si se complementa esta información con la aportada por los desvíos típicos se advierte una reducida dispersión de estos parámetros  $b_i$ . Esto es, el rango de valores de  $b_i$  estimados fue más estrecho que lo habitual en tests de ejecución típica.

La concentración de los parámetros de localización en torno a los niveles medios de la variable sugiere que la prueba presenta capacidad discriminativa sólo en los valores próximos a la media del rasgo. Esto revela una debilidad en las características psicométricas de la escala en tanto que demuestra que carece de ítems que puedan discriminar a personas con un nivel de Afecto extremo (ya sea positivo o negativo). En consecuencia, la modelización de la

prueba mediante el ML2p obligaría a la incorporación de mayor cantidad de reactivos en los niveles de la variable que han quedado descubiertos por los ítems.

Por otro lado, la alta concentración de los  $b_i$  podría deberse a un efecto de la dicotomización de las opciones de respuestas para el análisis de los datos. En efecto, el mismo comportamiento puede observarse también en las modelizaciones con uno y tres parámetros. Contemplar la información que proporciona la totalidad de las categorías de respuestas de la escala Likert permitiría estudiar con mayor precisión la capacidad discriminativa de la escala en un recorrido más extenso de la variable. En virtud de este resultado se propone en futuras investigaciones ensayar la adecuación de modelos politómicos de la TRI para analizar los ítems de la escala de Afecto hacia la Matemática.

## BIBLIOGRAFIA

Abad, F., Olea, J., Ponsoda, V. y García, C. (2011) *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.

Abal, F. (2013) *Comparación de modelos dicotómicos y politómicos de la Teoría de Respuesta al Ítem aplicados a un test de comportamiento típico*. Tesis Doctoral no publicada. Universidad de Buenos Aires, Argentina.

Abal, F., Lozzia, G., Aguerri, M., Galibert, M. y 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.

Aguado, D., Rubio, V.J., Hontangas, P.M. y Hernández, J.M. (2005) Propiedades psicométricas de un test adaptativo informatizado para la medición del ajuste emocional. *Psicothema*, 17, 484-491.

Baker, F.B. (2001) *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Birnbaum, A. (1968) Some latent trait models and their use in inferring an examinee's ability. En F. Lord & M. Novick (Eds.) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.

Ferrando, P. J. (1994) Fitting response models to the EPI-A Impulsivity scale. *Educational and Psychological Measurement*, 54, 118-127.

Ferrando, P.J. (2003) The accuracy of the E, N and P trait estimates: an empirical study using the EPQ-R. *Personality and Individual Differences*, 34, 665-679.

Ferrando, P. y Demestre, J. (2008) Características de forma y contenido que predicen la capacidad discriminativa en ítems de personalidad: un análisis basado en la Teoría de Respuesta a los Ítems. *Psicothema*, 20, 851-856.

Hair, J.F., Anderson, R.E., Tatham, R.L. y Black, W.C. (1999) *Análisis Multivariante*. Madrid: Prentice Hall.

Kaiser, H.F. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Martínez Arias, M. R. (1995) *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.

Muñiz, J. (2010) Las teorías de los tests: Teoría clásica y Teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31 (1), 57-66.

Panter, A.T., Swygert, K.A. y 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. y Waller, N.G. (1990) Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14, 45-58.

Reise, S. y Waller, N. (2009) Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27-48.

Revuelta, J., Abad, F.J. y Ponsoda, V. (2006) *Modelos Politómicos de respuesta al ítem*. Madrid: La Muralla.

Rouse, S.V., Finger, M.S. y Butcher, J.N. (1999) Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-5 scales. *Journal of Personality Assessment*, 72, 282-307.

Thissen, D. (2003) *MULTILOG*. Chicago: Scientific Software International.