

# Relación entre la Deseabilidad Social y los parámetros del Modelo Logístico de tres parámetros

## The relationship among social desirability and the parameters of the three-parameter logistic model

Facundo Abal, Sofía Auné, Horacio Attorresi  
Autor referente: sofiaaune177@hotmail.com

Universidad de Buenos Aires (Argentina)

### Historia editorial

Recibido: 24/09/2014

Aceptado: 25/05/2015

### RESUMEN

El objetivo de este trabajo es estudiar la relación de los parámetros  $a$ ,  $b$  y  $c$  del modelo logístico de tres parámetros y la tendencia de los individuos a atribuirse a sí mismo cualidades socialmente deseables. Participaron 1592 personas que respondieron las escalas Perseverancia y Distorsión del Big Five Questionnaire (BFQ). Se aplicó el ML3p a los ítems de perseverancia mediante el programa MULTILOG. Previamente se debieron eliminar 3 reactivos para cumplir con la condición de unidimensionalidad requerida por el modelo. Como la escala Likert de los ítems de perseverancia tiene 5 de alternativas se contemplaron diferentes procedimientos para establecer 4

condiciones de dicotomización de los datos. Los parámetros de los ítems se estimaron el procedimiento de Máxima Verosimilitud Marginal a Posteriori. El ajuste del ML3p fue adecuado para los 9 ítems analizados. Considerando las diferentes condiciones de dicotomización se identificaron entre 3 y 5 ítems de cuya respuesta clave se asoció a mayores niveles de Distorsión. Sin embargo, los resultados obtenidos no permiten concluir que la respuesta influida por la deseabilidad social repercute sobre alguno de los parámetros estudiados. Se discuten las limitaciones del presente estudio y se comentan otras líneas de investigación que pueden guiar a futuros estudios.

**Palabras clave:** Deseabilidad Social; Distorsión de Respuesta; Teoría de Respuesta al Ítem; Modelo Logístico de tres parámetros.

## ABSTRACT

The aim of this study is to examine the association of the parameters  $a$ ,  $b$  and  $c$  of the three-parameter logistic model (3pLM) and the tendency of to self-attribute personality qualities socially desirable. The incidental sample was made up of 1592 participants which responds the Perseverance and Lie Scale of the Big Five Questionnaire (BFQ). The 3pLM to items of Perseverance by MULTILOG program was applied. Previously 3 items were deleted to meet the unidimensionality assumption of the model. Since the Likert scale of Perseverance items has 5-points, 4 procedures

of dichotomization of the data were considered. The parameters estimation was carried out through marginal maximum likelihood a posteriori procedure. The 3pLM fitted to the data satisfactorily for 9 items. Between 3 and 5 items, depending procedures of dichotomization, showed keys response associated with higher levels of Lie. However, the results do not support the conclusion that the social desirability response to have an impact on the 3pLM parameters. The limitations of this study and potential future directions for the research are discussed.

**Keywords:** Social Desirability; Response Distortion; Item Response Theory; Three-parameters Logistic Model.

La Deseabilidad Social (DS) es un tema espinoso en el campo de la medición de los atributos de la personalidad. Es fácil reconocer que las respuestas a los ítems pueden estar influidas por un deseo más o menos consciente de dar una buena imagen. Por ende, se hace evidente que los tests de comportamiento típico no miden sólo las diferencias individuales en el rasgo de interés, sino que además existe otra fuente de variación que afecta insoslayablemente las puntuaciones de los tests. Sin embargo, resulta complejo determinar qué proporción de varianza de las puntuaciones es explicada por el rasgo medido y cuánta puede ser explicada por la falta de sinceridad y objetividad en las respuestas de un individuo (Morales, 2006).

En sus orígenes, el estudio de la DS se enfocó en el análisis de las propiedades de los ítems que despertaban una mayor tendencia a dar una falsa imagen. El supuesto central de este enfoque sostenía que algunas opciones de respuesta de los ítems serían más valoradas socialmente y que los individuos podrían elegir las

independientemente de su nivel de rasgo (Edwards, 1957). Durante los años 50 y 60, el abordaje de la DS mostró un cambio de perspectiva. La construcción de numerosos instrumentos específicos para su medición demostró que podía ser concebida como un rasgo estable y consistente de la personalidad de los individuos. La investigación empírica pasó de entender a la DS como un atributo propio del ítem a identificarla como una característica inherente al evaluado que puede aparecer en mayor o menor medida (Ferrando & Chico, 2000).

El desarrollo de la Teoría de Respuesta al Ítem (TRI) y su aplicación a la modelización de tests de comportamiento típico proporcionó nuevas herramientas para el estudio de la DS (Zickar & Sliter, 2012). La TRI propone un conjunto de modelos que postulan la existencia de una relación directa entre el comportamiento de un individuo frente a un ítem y el rasgo que genera esta conducta. La formalización de se expresa mediante una función matemática que vincula la probabilidad de dar la respuesta *clave* a un ítem para cada nivel del rasgo latente que se pretende medir. Los modelos de la TRI se componen de parámetros que caracterizan las propiedades psicométricas de los ítems (parámetros estructurales) y parámetros que describen los niveles de rasgo de los individuos (parámetros incidentales). Si bien se han desarrollado múltiples líneas de investigación en el marco de la TRI, una de las más importantes tiene como objetivo encontrar correlatos empíricos que vinculen la DS con los parámetros estructurales de los modelos. De esta manera, la TRI permitió volver a considerar al ítem como unidad de análisis para el estudio de la DS.

Los modelos de la TRI más utilizados para el estudio de la DS consideran hasta tres parámetros para caracterizar el comportamiento de un ítem: la localización del ítem en la escala del rasgo (parámetro *b*), la capacidad discriminativa (parámetro *a*) y la probabilidad de que un individuo con muy bajo nivel de rasgo esté de acuerdo con el

ítem (parámetro  $c$ ). Zumbo, Pope, Watson y Hubley (1997) fueron los primeros en hipotetizar que el parámetro  $c$  del podría ser interpretado como indicador de la incidencia de la DS en tests de personalidad. Según estos autores, el enunciado de un ítem podía estar redactado de tal manera que personas con muy bajo nivel del rasgo adhieran al mismo por adoptar una postura socialmente valorada. Rouse, Finger y Butcher (1999) lograron obtener evidencias que apoyan esta hipótesis en algunas escalas del MMPI-2, pero los resultados fueron bastante acotados como para generalizar dicha interpretación. Reise y Waller (2003) disintieron con esta hipótesis considerando que es más razonable suponer que la DS afecta principalmente al parámetro  $b$ . Si un grupo de individuos responde a los ítems de una escala usando un estilo de respuesta socialmente deseable la estimación de sus niveles de rasgo debería resultar artificialmente más elevada. Esto podría llevar a sesgar la estimación de los parámetros tanto de la distribución del rasgo latente como a la localización de los ítems. En la misma línea, Zickar y Ury (2002) encontraron que la DS puede afectar la estimación de  $b$  pero sólo en interacción con la transparencia del ítem (grado en que puede reconocerse el constructo que pretende medir el reactivo).

Asimismo, se ha intentado buscar una asociación entre la DS y la capacidad discriminativa de los ítems reflejada por el parámetro  $a$ , pero los resultados están lejos de ser concluyentes. Zickar y Ury (2002) no encontraron una relación entre DS y  $a$ , mientras que Hubley, Wu y Zumbo (2009) reportaron una asociación significativa aunque de baja intensidad. En consonancia con los estudios presentados, el objetivo general de este trabajo es estudiar la relación de los parámetros  $a$ ,  $b$  y  $c$  pertenecientes al modelo logístico de tres parámetros y la tendencia de los individuos a atribuirse a sí mismo cualidades socialmente deseables.

Por tratarse de un estudio empírico se trabaja a partir de la modelización de las respuestas observadas a los ítems de la escala Perseverancia perteneciente al Big Five Questionnaire (BFQ; Caprara, Barbarnelli, & Borgogni, 1993, adaptación de Bermúdez, 1995).

## Método

### Participantes

Se contó con la colaboración voluntaria de una muestra incidental compuesta por 1592 personas de población general residentes en Ciudad de Buenos Aires y Gran Buenos Aires (55% fueron mujeres). La edad de los participantes osciló entre 18 y 74 años, con una media de 35.8 años ( $DE = 13.2$ ). El 62.7% alcanzó un nivel de estudio terciario/universitario completo o incompleto. Las ocupaciones más frecuentemente reportadas fueron: empleados (40.6%), docentes (11.1%) y estudiantes (10.9%).

### Instrumentos

*Escalas Perseverancia y Distorsión del Big Five Questionnaire* (Caprara et al., 1993, adaptación de Bermúdez, 1995). Cada una de ellas está compuesta de 12 ítems con cinco opciones de respuesta (de *Completamente verdadero para mí* a *Completamente falso para mí*). La Escala Perseverancia mide aspectos como la persistencia y tenacidad con que se llevan a cabo las tareas y actividades emprendidas así como también la puntualidad con que se cumple lo prometido. La Escala Distorsión mide la tendencia del evaluado a adjudicarse cualidades socialmente deseables en un grado mayor al nivel en que tales características se presentan en los individuos de la población. Es decir, esta escala permite evaluar la tendencia a alterar las respuestas con el propósito más o menos intencionado de ofrecer una impresión distorsionada de sí mismo.

En el manual de la adaptación del BFQ, Bermúdez (1995) refiere un índice alfa de Cronbach de .77 para Distorsión aunque para la presente investigación se halló índice ligeramente más bajo de .74.

### Procedimiento y Análisis de datos

Los individuos respondieron el cuestionario de forma voluntaria y sin tiempo límite. Previa administración se les explicó a los participantes que la prueba tenía como objetivo la medición de aspectos de su personalidad y que no había respuestas correctas o incorrectas. Con el fin de propiciar una baja motivación para la disimulación se dieron garantías del anonimato de las respuestas.

Se corroboró la unidimensionalidad del espacio latente mediante un análisis exploratorio de la estructura factorial de los datos politómicos (Muñiz, Fidalgo, García Cueto, Martínez, & Moreno, 2005). Antes se verificó la factibilidad de este estudio mediante el índice de adecuación muestral de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett. Se empleó el método factorización de ejes principales partiendo de un criterio que a priori supuso la unidimensionalidad del constructo analizado (Hair, Anderson, Tatham, & Black, 1999). Siguiendo la demostración de Lord y Novick (1968), la existencia de independencia local se dio por satisfecha al confirmarse la unidimensionalidad de la escala.

Se aplicó el Modelo Logístico de tres parámetros (ML3p) operando el programa MULTILOG 7.03 (Thissen, 2003). Este modelo describe la probabilidad de dar la respuesta clave a un ítem dicotómico siguiendo la formulación:

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

Donde:

$\theta$  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  $\theta$ .

$b_i$  es el parámetro de localización que coincide con el valor  $\theta$  necesario para tener probabilidad  $(1 + c_i)/2$  de contestar la clave en el ítem  $i$ .

$a_i$  es el índice de discriminación del ítem  $i$ .

$c_i$  es el parámetro de pseudo azar en el contexto de los tests de rendimiento máximo pero aún no ha recibido una denominación que describa su relevancia en el contexto de los tests de comportamiento típico.

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

Dado que la escala de Perseverancia cuenta con un formato de respuesta politómica los datos debieron recodificarse en dos categorías para posibilitar la aplicación del ML3p. Como la escala Likert presenta una cantidad impar de alternativas se contemplaron tres condiciones para la dicotomización: a) *Recodificación 23* que agrupa las dos categorías inferiores y las tres superiores, b) *Recodificación 32* que fusiona las tres inferiores y las dos superiores y c) *Recodificación Azar* que colapsa las dos primeras categorías, las dos últimas y distribuye en éstas de manera aleatoria a los individuos que eligieron la categoría central. También se agregó una cuarta condición (*Recodificación Missing*) en la que se consideró que como perdido el dato del individuo que eligió la categoría central.

Para las estimaciones de los parámetros del ML3p se aplicaron las opciones que MULTILOG ofrece por default. No obstante, en una primera corrida del programa no se alcanzó el criterio de convergencia de 0.001 habiendo efectuado las 100 iteraciones. La lectura de la salida computacional de MULTILOG mostró que la dificultad aparecía

para estimar el valor del parámetro  $c$ . El programa no estima el parámetro  $c$  directamente por conveniencia sino que lo estima a partir de la inversa de la función logística, el  $\text{logit}(c)$ , siendo éste igual al  $\ln[c/(1 - c)]$  y teniendo un recorrido entre  $\pm \infty$ . Para el evitar el problema de estimación del parámetro  $c$  se procedió a efectuar una especificación sobre la distribución previa del  $\text{logit}(c)$  (Thissen, 2003).

Se utilizó una distribución previa *logit-normal* con media -1.4 y desvío estándar de 1; equivalente a suponer una distribución con media  $c = .20$  (proporción de respuesta por azar en las cinco categorías del ítem). Este cambio en la distribución previa del  $\text{logit}(c)$  permitió alcanzar el criterio de convergencia en la estimación. La incorporación de esta especificación implicó realizar una estimación bayesiana de los parámetros de los ítems. El procedimiento utilizado en este caso fue de Máxima Verosimilitud Marginal a Posteriori (Abad, Olea, Ponsoda, & García, 2011).

Como indicador del ajuste de los datos a los modelos MULTILOG brinda las proporciones observadas y las esperadas de elección para cada una de las categorías de respuesta (Thissen, 2003). En virtud de que esta información es pobre como única evidencia de ajuste, se adicionaron otros indicadores indirectos recomendados por Gray-Little, Williams y Hancock (1997): a) la convergencia en la estimación de los parámetros; b) la estimación de parámetros con valores razonables y c) errores de estimación de los parámetros relativamente bajos. En el caso de los parámetros  $c$  se controlaron los errores de estimación de los  $\text{logit}(c)$  ofrecidos por el programa.

El estudio de la asociación entre los parámetros del ML3p y la DS en las respuestas se realizó empleando dos pasos. Se siguió igual procedimiento para cada una de las condiciones de recodificación. En el primer paso se aplicó para cada ítem dicotomizado una prueba de diferencia de medias (muestras independientes) contemplando a la variable Distorsión como dependiente. Esto es, se buscó identificar

cuáles fueron los ítems de Perseverancia que mostraron una diferencia estadísticamente significativa en la variable Distorsión entre quienes puntuaron 1 (o sea, se perciben como más perseverantes) y quienes puntuaron 0. El segundo paso consistió en comprobar si los ítems que presentaron niveles promedio de Distorsión significativamente más elevados para los individuos que puntuaron 1 en Perseverancia también presentaron diferencias significativas también en los valores de los parámetros  $b$ ,  $a$  y  $c$ . Para esto se aplicó una prueba  $U$  de Mann-Whitney comparando cada uno de los parámetros obtenidos por el conjunto de ítems identificados en el primer paso versus el resto de los ítems. Según las hipótesis recogidas de la literatura, se espera que los ítems que despertaron mayor tendencia a la DS presenten menores valores de  $b$  o mayores valores de  $c$  comparados con los demás reactivos. En relación con el parámetro  $a$  no se ha reportado ninguna dirección específica para definir esta diferencia.

## Resultados

### *Comprobación del supuesto de unidimensionalidad*

Tanto la medida de adecuación muestral  $KMO = .83$  como la prueba de esfericidad de Bartlett con un valor de 2268.3 ( $p < .0001$ ) también sugirieron la posibilidad de aplicar un análisis factorial a la matriz de datos. El análisis factorial mostró que 3 de los 12 ítems presentaban saturaciones inferiores a  $.30$  en el primer factor por lo que se procedió a su eliminación.

Luego de esta depuración, se alcanzó una estructura factorial acorde con el requerimiento de unidimensionalidad del ML3p. Así se obtuvo una solución con un único factor dominante cuyo autovalor fue de 2.23 y que describió el 24.8% de la varianza ( $KMO = .83$ ; Prueba de esfericidad de Bartlett,  $\chi^2 = 1876,9$ ;  $p < .0001$ ). Los pesajes de los ítems fueron adecuados y oscilaron entre  $.35$  y  $.58$ . Este análisis

permite corroborar la unidimensionalidad de los datos considerando el criterio propuesto por Reckase (1979) que requiere de un porcentaje de varianza descripta para el primer factor superior al 20%.

### **Estimación y ajuste**

La tabla 1 resume la cantidad de ciclos necesarios en cada una de las condiciones para alcanzar el criterio de convergencia de 0.001 y el máximo cambio de parámetros entre ciclos registrado. Allí puede observarse que, comparativamente, la *Recodificación 23* requirió de una mayor cantidad de ciclos que el resto de las condiciones mientras que la *Recodificación Missing* registró el valor más elevado entre los máximos cambios de un parámetro con respecto al ciclo anterior de la estimación. Ambos indicadores señalan cierta dificultad del modelo para ajustarse a los datos bajo estas condiciones de recodificación. No obstante, los residuos obtenidos entre las proporciones empíricas y esperadas (según la probabilidad de respuesta dada por ML3p) fueron iguales o inferiores a .01. Esto implica que, según la información suministrada por MULTILOG, el ajuste de los datos al ML3p fue aceptable para los nueve reactivos analizados y bajo las cuatro condiciones.

*Tabla 1.* Características de los procesos de estimación.

	Recodificación 23	Recodificación 32	Recodificación Azar	Recodificación Missing
Cantidad de ciclos	45	19	29	25
Máximo cambio interciclo	0.00086	0.00065	0.00080	0.01935

La estimación alcanzó el criterio de convergencia empleando una cantidad de iteraciones razonable, los parámetros presentaron valores dentro de un rango esperable (tabla 2) y los errores de estimación resultaron relativamente bajos. Estos tres indicadores aportan evidencia indirecta que apunta a sostener un ajuste satisfactorio de los datos (Gray-Little et al., 1997). Todas las condiciones mostraron un parámetro *a* promedio que refleja una capacidad discriminativa moderada según la clasificación de Baker (2001). Los parámetros *b* resultaron negativos indicando que se requirieron niveles bajos de Perseverancia para acordar con los ítems. La *Recodificación 32* mostró los valores de *c* más bajos incluso por debajo de la probabilidad de respuesta por azar (es decir, .20 dado que el ítem presenta cinco opciones de respuesta). La *Recodificación 23*, en cambio, presentó el valor promedio de *c* más elevado.

**Tabla 2.** Parámetros estimados y estadísticos descriptivos para las cuatro condiciones.

	Recodificación			Recodificación			Recodificación			Recodificación		
	23			32			Azar			Missing		
	<i>a</i>	<i>b</i>	<i>c</i>									
1	1.03	-1.36	.22	1.15	-2.28	.16	1.34	-2.01	.19	0.80	-1.99	.19
2	0.57	-0.91	.23	0.47	-2.76	.27	0.53	-2.29	.26	0.44	-1.87	.20
3	0.75	-1.05	.22	0.84	-2.58	.16	1.01	-2.15	.18	0.54	-2.12	.17
4	0.68	-0.76	.21	0.72	-2.16	.19	0.75	-1.87	.19	0.49	-1.77	.19
5	0.97	-0.71	.26	0.81	-1.72	.11	0.97	-1.42	.21	0.96	-1.01	.17
6	1.11	0.35	.32	0.75	-0.76	.19	1.60	-0.07	.20	0.62	-0.29	.36
7	0.84	0.10	.37	1.29	-0.55	.13	1.60	-0.43	.18	0.88	-0.35	.22
8	0.74	-0.74	.22	0.91	-1.96	.09	1.09	-1.60	.17	0.55	-1.64	.15
9	0.77	-0.57	.39	1.13	-1.17	.11	1.26	-1.05	.44	1.17	-0.34	.26
M	0.90	-1.77	.27	0.83	-0.63	.17	0.72	-1.26	.22	1.13	-1.43	.21
DE	0.25	0.79	.07	0.18	0.54	.06	0.25	0.77	.09	0.36	0.78	.06

### ***Asociación de los parámetros b, a y c con Distorsión***

Para considerar el impacto de la DS en la respuesta a los ítems se realizó una prueba de diferencia de medias de la variable Distorsión entre quienes fueron codificados con 1 y 0 para cada una de las condiciones. En la tabla 3 se muestran los estadísticos de contraste *t* correspondientes a cada una de las pruebas efectuadas. Además de la significación estadística de la diferencia, es necesario rescatar también el signo del estadístico de contraste para analizar cuál es el grupo que presentó mayores niveles promedio de Distorsión. Los valores positivos y significativos de *t* suponen diferencias a favor de quienes respondieron la opción clave en Perseverancia, por lo que podría

aceptarse una mayor tendencia de los individuos a distorsionar la respuesta a estos ítems. Este análisis permitió identificar entre tres (*Recodificación 23*) y cinco ítems (*Recodificaciones 32 y Azar*) cuyas respuestas clave se asocian niveles de Distorsión más elevados según la condición.

*Tabla 3.* Estadísticos de contraste de la diferencia de medias de Distorsión entre quienes respondieron o no la clave en cada ítem de Perseverancia.

	Recodificación 23	Recodificación 32	Recodificación Azar	Recodificación Missing
1	-0.89	-1.43	1.05	-0.74
2	-4.84**	-3.99**	-4.03**	-4.36**
3	1.25	<b>6.43**</b>	<b>4.85**</b>	<b>2.27*</b>
4	0.92	<b>2.84**</b>	<b>2.40*</b>	1.30
5	-4.24**	-3.19**	-3.43**	-4.32**
6	<b>5.29**</b>	<b>7.56**</b>	<b>6.63**</b>	<b>7.17**</b>
7	<b>2.71**</b>	<b>4.97**</b>	<b>3.99**</b>	<b>3.79**</b>
8	<b>2.45*</b>	<b>5.16**</b>	<b>4.15**</b>	<b>3.08**</b>
9	-1.97*	-2.43*	-2.12*	-2.24*

*Nota:* En negrita se destacan los estadísticos *t* significativos con signo positivo.  
\**p*<.05; \*\**p*<.01

A partir de los estadísticos *t* se agruparon los ítems de cada condición en dos subconjuntos: a) los ítems con valores de *t* positivos y significativos (cuya respuesta positiva se asoció a mayores niveles de Distorsión) y b) los ítems con valores *t* negativos y significativos o no significativos. Se aplicó la prueba *U* de Mann-Whitney para confirmar si los parámetros estimados de los ítems pertenecientes al primer grupo presentaban diferencias estadísticamente significativas con respecto a los del segundo grupo. Las pruebas *U* mostraron que no hay evidencias que permitan afirmar que el rango promedio de los parámetros *b* y *a* difiera entre los dos subconjuntos de

ítems. Este hallazgo fue replicado bajo las cuatro condiciones (tabla 4). Sólo se halló una diferencia significativa en los parámetros *c* para la condición de *Recodificación Azar* ( $z = -2.01$ ;  $p = .04$ ).

Tabla 4. Estadísticos de contraste de las pruebas U de Mann-Whitney.

	Recodificación 23	Recodificación 32	Recodificación Azar	Recodificación Missing
a	-0.76	-0.25	-0.98	-1.48
b	-1.81	-0.73	-0.25	-0.98
c	-0.79	-0.12	-2.1*	-0.12

Nota: \*  $p < .05$

## Discusión

Los resultados obtenidos en el presente estudio demuestran que no es posible concluir una relación entre la tendencia a dar una imagen favorable de sí mismo y ninguno de los parámetros del ML3p. El único resultado estadísticamente significativo logrado para el parámetro *c* sólo fue observado en una de las condiciones de recodificación de los datos y presenta una dirección contraria a la hipótesis propuesta por Zumbo et al. (1997). Esto es, los ítems en los que se observó una mayor tendencia de los individuos a distorsionar la respuesta presentaron valores de *c* más bajos. Sin embargo, y a pesar del resultado estadístico, esta interpretación no parece justificada conceptualmente.

Aceptar la hipótesis de Zumbo et al. (1997) implica suponer que los individuos sólo tendieron a responder influenciados por la deseabilidad social a algunos pocos ítems

de una escala. En efecto, si hubieran respondido a la mayoría de ellos, entonces ya no hubiesen tenido un bajo nivel de rasgo. Reise y Waller (2003) consideran que es poco probable que un examinado distorsione su respuesta en sólo unos pocos ítems; se trata más bien de un estilo de respuesta que presenta cierta sistematicidad en el comportamiento de un individuo. Si un porcentaje considerable de la muestra responde sistemáticamente a toda la prueba dando una imagen positiva de sí mismo, los ítems que se hubiesen localizado en niveles más altos del rasgo aparecerían, como consecuencia de este sesgo, con parámetros  $b$  correspondientes a niveles más bajos. Ahora bien, dado que las administraciones realizadas en este estudio se llevaron a cabo en un contexto con baja motivación para la disimulación es lógico suponer que las personas que respondieron influidos por la DS no son mayoría y, por esta razón, no es posible registrar una asociación con el parámetro  $b$ .

La DS tampoco mostró asociación con el parámetro  $a$ , por lo que no parece ser un factor perjudique particularmente la capacidad discriminativa de los ítems. Otros investigadores han reportado resultados similares utilizando distintos instrumentos (Hubley et al., 2009; Zickar & Ury, 2002). Dado que los ítems de personalidad suelen resultar menos discriminativos que los ítems que miden habilidades, la investigación sobre el parámetro  $a$  ha sido abundante pero también, en la mayoría de los casos, fue infructuosa. Diferentes investigadores encontraron que aspectos tales como la complejidad sintáctica del ítem, el grado en que el indicador es concreto o abstracto y la ambigüedad-vaguedad de un enunciado no provocan variaciones significativas en los valores de  $a$  (Ferrando & Demestre, 2008; Zickar & Ury, 2002).

Otra consideración para analizar es el grado de relevancia atribuido a la DS en la determinación de la respuesta. Según Morales (2006), no se ha demostrado de manera consistente que la DS explique una parte considerable de la variabilidad de las

puntuaciones cuando la administración se efectúa bajo condiciones neutrales. El ML3p, como los demás modelos de la TRI habitualmente utilizados para este tipo de estudios, es coherente con este planteo dado que requiere la unidimensionalidad del constructo. En una situación real son múltiples los factores que pueden afectar a las respuestas a un ítem y, en consecuencia, difícilmente pueda satisfacerse totalmente este supuesto (Muñiz et al., 2005). Por este motivo, para los modelos unidimensionales exigen la delimitación de un factor dominante que se diferencie sustancialmente de otras dimensiones menores que puedan acompañarlo. Esto significa que la aplicación de estos modelos supone implícitamente que la DS tiene una baja incidencia en la determinación de la respuesta.

En un estudio realizado con las escalas del EPQ-R, Ferrando (2008) evaluó la posible influencia de la DS basándose en un modelo bidimensional de la TRI. Esto permitió un análisis pormerizado a nivel del ítem respecto de cuáles fueron los contenidos más afectados. Entre las limitaciones del presente trabajo debe destacarse que las posibilidades de generalización de estos resultados dependen de los instrumentos utilizados. Futuras investigaciones buscarán explorar estas hipótesis en otras escalas y empleando otros instrumentos para la evaluación de la DS.

La metodología aplicada en este estudio sólo ha considerado a los parámetros estructurales del ML3p. Pero dado los resultados obtenidos, parece conveniente contemplar en ulteriores estudios otras de las líneas de investigación llevadas adelante en el marco de la TRI basadas en los parámetros incidentales de los modelos. Se han desarrollado índices para la detección de patrones de respuesta deshonestos a partir de indicadores de ajuste de los parámetros de los individuos. Aun cuando se verifica un ajuste global de los datos es posible identificar un conjunto de individuos para los que el modelo no resulta adecuado. Sin embargo, no todo desajuste de estos índices

representa necesariamente un patrón de respuesta influenciado por DS por lo que debe controlarse la aparición de falsos positivos (Zickar & Drasgow, 1996; Zickar & Sliter, 2012).

A pesar de la extensa bibliografía que se ha ocupado de la DS, es un problema que nunca ha sido resuelto de forma totalmente satisfactoria (Morales, 2006). La manera en que la deseabilidad social afecta a la validez de las escalas sigue siendo una pregunta abierta para la investigación psicométrica del área (Ziegler, MacCann, & Roberts, 2012). La modelización con TRI de tests de comportamiento típico es un área que se encuentra atravesando un momento de gran desarrollo. Su utilización en test de ejecución típica también ha ganado terreno en aplicaciones que hasta no hace muchos años parecían reservadas a las variables de aptitud (Morizot, Ainsworth, & Reise, 2007; Reise & Waller, 2009). De esta manera, el paulatino crecimiento de las aplicaciones permitirá continuar ampliando un campo de conocimiento delimitado hasta ahora para captar con mayor especificidad la influencia de la DS en la respuesta a tests de personalidad.

## Referencias

- Abad, F., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Baker, F. B. (2001). *The basics of Item Response theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Bermúdez, J. (1995). *Manual del Cuestionario "Big five" (BFQ)*. Madrid, ES: TEA.
- Caprara, G. V., Barbarnelli, C., & Borgogni, L. (1993). *Big Five Questionnaire (BFQ): Manuale*. Florencia, IT : Organizzazioni Speciali.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.

- Ferrando, P. (2008). The impact of social desirability bias on the EPQ-R item scores: An item response theory análisis. *Personality and Individual Differences*, 44(8), 1784-1794.
- Ferrando, P., & 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.
- Ferrando, P., & Chico, E. (2000). Adaptación y análisis psicométrico de la escala de deseabilidad social de Marlowe y Crowne. *Psicothema*, 12, 383-389.
- Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An Item Response Theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 23, 443-451.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1999). *Análisis multivariante*. Madrid, ES: Prentice Hall.
- Hubley, A. M., Wu, A. D., & Zumbo, B. D. (2009). *Interpreting IRT parameters: Putting psychological meat on the psychometric bone*. Trabajo presentado en 117th Annual Meeting of the American Psychological Association (APA), Toronto, Canada.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental tests scores*. Reading, MA: Addison-Wesley.
- Morales, P. (2006). *Medición de actitudes en psicología y educación*. Madrid, ES: Universidad Pontificia Comillas.
- Morizot, J., Ainsworth, A., & 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, NY: Guilford.
- Muñiz, J., Fidalgo, A. M., García Cueto, E., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems*. Madrid, ES: La Muralla.

- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor test: Results and implications. *Journal of Educational Statistics, 4*, 207-230.
- Reise, S., & Waller, N. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods, 8*, 164-184.
- Reise, S., & Waller, N. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27-48.
- Rouse, S. V., Finger, M. S., & 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, IL: Scientific Software International.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement, 20*, 71-87.
- Zickar, M. J., & Sliter, K. A. (2012). Searching for unicorns: Item Response Theory-based solutions to faking problem. En M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp.113-130). NY: Oxford University.
- Zickar, M. J., & Ury, K. L. (2002). Developing an interpretation of item parameters for personality items: content correlates of parameter estimates. *Educational and Psychological Measurement, 62*, 19-31.
- Ziegler, M., MacCann, C., & Roberts, R. (2012). *New perspectives on faking in personality assessment*. NY: Oxford University.
- Zumbo, B., Pope, G. A., Watson, J. E., & Hubley, A. M. (1997). An empirical test of Roskam's conjecture about the interpretation of an ICC parameter in personality inventories. *Educational and Psychological Measurement, 57*, 963-969.

### Formato de citación

---

Abal, F., Auné, S., y Attorresi, H. (2015). Relación entre la Deseabilidad Social y los parámetros del Modelo Logístico de tres parámetros. *Psicología, Conocimiento y Sociedad*. 5 (1) 5 - 24. Disponible en: <http://revista.psico.edu.uy/>

---