

## **Análisis psicométrico mediante la Teoría de la Respuesta al Ítem: modelización paso a paso de una Escala de Soledad**

### **A psychometric analysis from the Item Response Theory: step-by-step modelling of a Loneliness Scale**

### **Análise psicométrica a partir da Teoria da Resposta ao Item: modelagem passo a passo de uma Escala de Solidão**

*Sofía Esmeralda Auné*<sup>1</sup>, ORCID 0000-0002-0620-0199  
*Facundo Juan Pablo Abal*<sup>2</sup>, ORCID 0000-0001-7023-5380  
*Horacio Félix Attorresi*<sup>3</sup>, ORCID 0000-0002-3027-1069

<sup>1 2 3</sup> *Instituto de Investigaciones, Facultad de Psicología, Universidad de Buenos Aires. Argentina*  
<sup>1 2</sup> *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Argentina*

**Resumen:** La teoría de la respuesta al ítem (TRI) es un conjunto de modelos psicométricos utilizados en el desarrollo, evaluación, mejora y puntuación de escalas de evaluación. Este artículo pedagógico provee un recorrido inicial acerca de cómo conducir, interpretar y exponer resultados de aplicación de modelos de la TRI aptos para opciones de respuesta politómica ordenada. Los datos utilizados como ejemplo para la modelización con TRI corresponden a la administración de la Escala de Soledad de Buenos Aires, un nuevo instrumento para evaluar la autopercepción de la soledad, a una muestra no probabilística e incidental de 509 participantes residentes en el Área Metropolitana de Buenos Aires (53% mujeres). El objetivo de este artículo es mostrar un recorrido por los pasos generales y los componentes necesarios para realizar un análisis con TRI para incrementar el acceso a esta poderosa técnica psicométrica.

**Palabras Clave:** teoría de la respuesta al ítem, pedagógico, soledad, modelo de respuesta graduada, psicometría avanzada

**Abstract:** The Item Response Theory (IRT) is a set of psychometric models used in the development, assessment, improvement, and scoring of evaluating scales. This pedagogical article provides an initial overview of how to conduct, as well as interpret and present the results from, the application of IRT models suitable for ordered polytomous response options. The data used as an example for IRT modelling corresponds to the administration of the Buenos Aires Loneliness Scale (BALS), a new instrument for the assessment of loneliness self-perception. This data also corresponds to a non-probabilistic and incidental sample of 509 participants residing in the Buenos Aires Metropolitan Area (53 % women). The objective of this article is to present an overview of the general steps and components needed to perform an IRT analysis, as a way to increase access to this powerful psychometric technique.

**Keywords:** item response theory, pedagogical, loneliness, graduated response model, advanced psychometrics



**Resumo:** A teoria da resposta ao item (TRI) é um conjunto de modelos psicométricos utilizado no desenvolvimento, avaliação, melhoria e pontuação das escalas de avaliação. Este artigo pedagógico fornece uma visão geral inicial de como conduzir, interpretar e apresentar resultados da aplicação de modelos TRI adequados para opções de respostas politômicas ordenadas. Os dados utilizados como exemplo para modelagem com TRI correspondem à aplicação da Escala de Solidão de Buenos Aires, um novo instrumento para avaliar a autopercepção da solidão, à amostra não probabilística e incidental de 509 participantes residentes na área metropolitana de Buenos Aires (53% mulheres). O objetivo deste artigo é apresentar as etapas gerais e os componentes necessários para realizar uma análise com TRI para aumentar o acesso a essa poderosa técnica psicométrica.

**Palavras-chave:** teoria da resposta ao item, pedagógico, solidão, modelo de resposta graduada, psicometria avançada

Recibido: 06/11/2018

Aceptado: 27/04/2020

Cómo citar este artículo:

Auné, S. E., Abal, F.J.P., & Attorresi, H.F. (2020). Análisis psicométrico mediante la Teoría de la Respuesta al Ítem: modelización paso a paso de una Escala de Soledad. *Ciencias Psicológicas*, 14(1), e-2179. doi: <https://doi.org/10.22235/cp.v14i1.2179>

*Correspondencia:* Sofía Esmeralda Auné, UBA; CONICET. E-mail: [sofiaaune177@hotmail.com](mailto:sofiaaune177@hotmail.com)

## La Teoría de la Respuesta al Ítem (TRI)

Es consensuado actualmente entre los psicometristas que la teoría clásica de los tests (TCT) tiene ciertas limitaciones (Attorresi, Lozzia, Abal, Galibert, & Aguerri, 2009). Entre ellas: (a) todas las medidas que se obtienen (e.g. el Alpha de Cronbach) dependen de la muestra particular de individuos que respondieron al instrumento, (b) instrumentos con índices de dificultad y discriminación diferentes generan resultados diversos en los mismos individuos, y (c) si el mismo constructo es medido por dos o más test diferentes, los resultados no son medidos en la misma escala. Por otra parte, la relación lineal ítems-constructo que supone la TCT podría no ser demasiado realista en muchos casos.

Se ha afirmado frecuentemente (e.g. Paek & Cole, 2019) que la teoría de la respuesta al ítem (TRI) resuelve muchas de las limitaciones de la TCT, aunque al costo de una mayor demanda matemática y computacional, el requerimiento de una muestra grande y supuestos más exigentes. Sin embargo, con los avances en capacidad computacional y programación, los expertos de muchos campos accedieron a los beneficios de aplicar la TRI.

La TRI es un conjunto de modelos orientados a explicar la relación entre las respuestas observadas a un ítem -que forme parte de una escala- y un constructo subyacente (Cappelleri, Lundy, & Hays, 2014). Para ello, los modelos de la TRI utilizan funciones matemáticas no lineales, frecuentemente la función logística, que describen la asociación entre el nivel del participante en un rasgo latente  $\theta$  y la probabilidad de seleccionar una determinada respuesta -o categoría de respuesta- a un ítem. En el ejemplo que se brinda en este estudio, el rasgo latente  $\theta$  es el nivel de soledad.

La primera cuestión a tener en cuenta para seleccionar un modelo de la TRI es la categorización de las opciones de respuesta de los ítems. Si esta categorización resulta en una dicotomía, los modelos más utilizados son los Modelos Logísticos de Uno (ML1P), Dos (ML2P) o Tres Parámetros (ML3P). Si se trata de tres o más categorías de respuesta, serán apropiados los modelos de la TRI para ítems con respuesta politómica. Si la respuesta politómica no es ordenada, se utiliza el Modelo de Respuesta Nominal (Bock, 1997). Los modelos más utilizados en la actualidad para respuesta politómica ordenada son: el Modelo de Respuesta Graduada (MRG; Samejima, 1969; 2016) y su versión restringida (*Reduced GR Model*, MRG<sub>R</sub>; Toland, 2013), el Modelo Generalizado de Crédito Parcial (MGCP; Muraki, 1992) y el Modelo de Crédito Parcial (MCP; Masters, 1982, 2016). Aunque se ha planteado la elección entre los modelos politómicos de la TRI como una cuestión de preferencia del investigador (e.g. Edelen & Reeve, 2007), actualmente existen métodos objetivos para comparar el ajuste relativo entre modelos con respecto a un determinado conjunto de datos (e.g. DeAyala, 2009; Toland, 2013) para poder determinar cuál es en cada caso el más adecuado.

Además, una cuestión muy importante para ofrecer garantías de validez de la escala es el análisis del funcionamiento diferencial del ítem (*differential item functioning*, DIF). La existencia de ítems con DIF atenta contra la unidimensionalidad cuando se pretende medir un único rasgo y pone en riesgo a la validez. Usualmente los estudios de DIF comparan entre dos grupos, denominados De Referencia y Focal. Si un ítem tiene DIF, en este caso implica que iguales puntuaciones en el ítem están representando diferentes niveles de soledad entre los dos grupos, lo que definitivamente no es una característica deseable en una técnica psicométrica.

### **La Escala de Soledad de Buenos Aires**

El propósito fundamental de este artículo es exponer los pasos necesarios para conducir, interpretar y presentar los resultados de la modelización con TRI en un formato accesible. Los pasos generales para realizar un análisis con TRI incluyen: (a) clarificar el objetivo del estudio, (b) considerar modelos relevantes, (c) testear los supuestos de los modelos y comparar su ajuste relativo y (d) aplicar el modelo seleccionado e interpretar los resultados. Para ejemplificar este análisis con TRI en forma didáctica, se ha elegido la Escala de Soledad de Buenos Aires ([ESBA]; Auné, Abal & Attorresi, 2019) como técnica psicométrica a modelizar.

Se trata de un nuevo instrumento para evaluar la autopercepción de la soledad. Es un test breve, unidimensional, conformado por siete ítems de respuesta politómica, los cuales fueron redactados en base a entrevistas grupales a población general adulta y de la tercera edad residentes en el Área Metropolitana de Buenos Aires (AMBA). Aunque existía una escala adaptada al país para medir la soledad, se realizó un estudio de la misma que evidenció que las respuestas a los ítems se veían influenciadas por la dirección de respuesta (Auné, Abal & Attorresi, 2020).

Inicialmente se obtuvieron evidencias de validez de contenido mediante el acuerdo interjueces utilizando el índice V de Aiken (Aiken, 1980, 1985) y se realizó un estudio piloto. Se realizó un Análisis Factorial Exploratorio, descartando los ítems que poseyeran uno o más de los siguientes criterios: a) asimetría y curtosis exageradas, b) elevados residuos estandarizados (> 2.58, criterio de Hair, Anderson, Tatham, & Black, 1999) y c) carga factorial menor a .40. Una vez depurados los ítems, se obtuvieron evidencias de validez convergente con la versión argentina de la UCLA (Sacchi & Richaud de Minzi, 1997) y con la autopercepción del nivel de soledad, así como evidencias de validez discriminante con respecto a la discapacidad social. La consistencia interna fue muy adecuada (alfa de Cronbach = .80, alfa ordinal = .87). Se realizaron, además, estudios de funcionamiento diferencial del ítem con respecto al género, hallándose que los ítems se hallaban libres de funcionamiento diferencial en este aspecto.

## Objetivos

El objetivo general de este trabajo es ejemplificar la modelización de una escala psicométrica de respuesta politómica con TRI mediante el análisis de los ítems que componen la ESBA.

Los objetivos específicos son los siguientes:

- a.- Comprobar los supuestos de los modelos de la TRI de unidimensionalidad e independencia local.
- b.- Comparar entre los modelos MRG, MRG<sub>R</sub>, MGCP y MCP y determinar cuál es el más adecuado para calibrar las respuestas a los ítems que componen la ESBA.
- c.- Explorar la existencia de Funcionamiento Diferencial del Ítem según el estado civil de los participantes utilizando el modelo de la TRI seleccionado.
- d.- Calibrar los ítems de la ESBA con los parámetros del modelo seleccionado.
- e.- Analizar para qué niveles de soledad es la ESBA resulta más precisa.

## Método

### *Participantes*

Se recolectó de una muestra no probabilística e incidental de 509 participantes. Los mismos fueron en un 53% mujeres y residían en el AMBA. Su promedio de edad fue de 44.3 ( $DE = 13$ ); 47.2% manifestó que se encontraba casado o en unión de hecho, 25% soltero/a, 15.3% divorciado/a, 4.7% viudo/a, mientras que 7.9% seleccionó la opción “otro”.

### *Procedimiento*

Se recolectaron los datos mediante un diseño muestral no probabilístico por accesibilidad. El protocolo fue administrado en formato de encuesta online, donde se incluía un consentimiento informado anónimo. Además, se aclaraba que el uso era exclusivamente para fines de investigación y la participación enteramente voluntaria, pudiendo cesar en cualquier momento.

### *Instrumentos*

*Cuestionario Sociodemográfico.* Consta de una serie de preguntas realizadas ad hoc para la presente investigación que indagaron las variables de género, edad, estado civil, nacionalidad y lugar de residencia.

*Escala de Soledad de Buenos Aires* ([ESBA]; Auné, et al., 2019). Se trata de un instrumento de siete ítems, donde la modalidad de respuesta se especifica mediante una escala Likert de cuatro opciones (1 = *Nada de Acuerdo*, 2 = *Poco de Acuerdo*, 3 = *Algo de Acuerdo*, 4 = *Totalmente de Acuerdo*).

### *Cumplimiento de los Supuestos de los Modelos de la TRI*

La verificación del supuesto de unidimensionalidad requerido por los modelos MRG, MRG<sub>R</sub>, MGCP y MCP a comparar se realizó mediante la implementación óptima del análisis paralelo (Timmerman, & Lorenzo-Seva, 2011) y el porcentaje de varianza explicada por el primer factor. Ambos índices se obtienen mediante un Análisis Factorial Exploratorio (Ferrando & Lorenzo-Seva, 2017a).

Además del supuesto de unidimensionalidad, los modelos MRG, MRG<sub>R</sub>, MGCP y MCP también asumen que, dado un nivel fijo de  $\theta$ , las respuestas a los ítems son independientes entre

sí. El índice  $X^2_{LD}$  (Chen & Thissen, 1997) se calcula para cada par de ítems e indica el no cumplimiento del supuesto cuando es mayor que 10.

### ***Comparación entre Modelos de la TRI***

Para comparar el ajuste relativo de los modelos MRG, MRG<sub>R</sub>, MGCP y MCP se utilizaron múltiples métodos, tal como lo describen De Ayala (2009) y Toland (2013). Por una parte, se implementó el Test de la Razón de Verosimilitudes (*Likelihood Ratio Test*; LRT) que compara dos modelos anidados, complementado con el estadístico  $R^2_{\Delta}$  (Haberman, 1978). En este caso el MRG<sub>R</sub> es una restricción del MRG y el MCP del MGCP. El LRT informa acerca de si la complejidad del modelo completo, sin restricciones en el parámetro  $a$ , es necesaria para mejorar el ajuste del modelo. Adopta una distribución  $\chi^2$ , donde un estadístico  $\chi^2_{\Delta}$  no significativo implica que la complejidad adicional del modelo no restringido es innecesaria para mejorar el ajuste de los datos. El estadístico  $R^2_{\Delta}$  mide en qué porcentaje el modelo completo aumenta la explicación de las respuestas a los ítems con respecto al modelo reducido. El  $R^2_{\Delta}$  se calcula como: (valor del logaritmo de la verosimilitud del modelo restringido - valor del logaritmo de la verosimilitud del modelo completo) / valor del logaritmo de la verosimilitud del modelo restringido (Toland, 2013).

Por otra parte, se calcularon el Criterio de Información de Akaike (*Akaike Information Criterion*, AIC) y el Criterio de Información Bayesiano (*Bayesian Information Criterion*, BIC) para cada uno de los modelos, donde valores más pequeños de AIC y BIC indican mejor ajuste relativo. Por último, se observó el ajuste global y la existencia de desajuste de ítems de cada uno de los modelos. El ajuste global se calculó con el estadístico  $M_2$  (Maydeu Olivares & Joe, 2005, 2006) y el índice RMSEA asociado, considerándose que el modelo ajusta si  $RMSEA \leq .05$ . Valores más pequeños del estadístico  $M_2$  señalan un mejor ajuste. Para determinar si cada ítem es explicado por un modelo se calculó el índice  $S-\chi^2$  (Orlando & Thissen, 2000, 2003). Si el valor  $p$  asociado al  $S-\chi^2$  es mayor que .01 indica ajuste (Toland, 2013).

### ***Análisis del Funcionamiento Diferencial del Ítem***

Se exploró la existencia de DIF por estado civil, dividiendo a la muestra entre quienes estaban casados o en unión de hecho y los participantes que seleccionaron otros estados civiles. Se realizó este análisis mediante la serie de pasos detallados por Woods (2009). Como primer paso, cada uno de los reactivos de la ESBA se verificó con el test de Wald modificado (Cai, 2012; Cai, Thissen, & du Toit, 2011; Langer, 2008) considerando el resto como anclaje. Posteriormente, se realizó un segundo paso, donde se testeó a un ítem con potencial DIF con anclaje en las respuestas al ítem más certeramente libre de DIF, evitando así la contaminación.

### ***Evidencias en el Marco de la TRI***

Se obtuvieron evidencias de confiabilidad en el marco de la TRI, mediante la Función de Información de los Ítems (FII) y la Función de Información del Test (FIT). La FII indica la precisión de determinado ítem en la medición de cada nivel de  $\theta$ . La sumatoria de las FII conforma la FIT, que informa la confiabilidad de la escala según el nivel en el rasgo  $\theta$ .

### ***Softwares Utilizados***

Los índices de unidimensionalidad se obtuvieron con el programa FACTOR versión 10.5 (Ferrando & Lorenzo-Seva, 2017b). Los análisis de independencia local, DIF y modelización con TRI se implementaron mediante el programa IRTPRO 4.2 (Cai et. al, 2011).

## Resultados

### *Cumplimiento de los Supuestos de los Modelos de la TRI*

La implementación óptima del análisis paralelo indicó que la cantidad sugerida de factores es uno, mientras que el porcentaje de varianza explicada por el primer factor fue de 57.48%. Por lo tanto, los datos pueden considerarse esencialmente unidimensionales. En las salidas de cada uno de los modelos MRG, MRG<sub>R</sub>, MGCP y MCP el índice  $X^2_{LD}$  resultó menor que 10 para cada par de ítems. Por lo tanto, para los cuatro modelos se pueden dar por satisfechos los dos supuestos, unidimensionalidad e independencia local.

### *Comparación entre Modelos de la TRI*

Los índices de ajuste para MRG, MRG<sub>R</sub>, MGCP y MCP se pueden observar en la Tabla 1. Aunque todos los modelos ajustaron globalmente, el MRG obtuvo valores más pequeños en estadístico  $M_2$ , logaritmo de la verosimilitud, AIC y BIC. El LRT entre el MRG y el MRG<sub>R</sub> indicó que la complejidad adicional del modelo completo es necesaria para mejorar el ajuste a los datos ya que  $\chi^2_{\Delta}(6) = 7108.19 - 7011.88 = 96.31$ ,  $p = 7.07 \times 10^{-19}$ . El cambio relativo entre estos modelos fue de  $R^2_{\Delta} = .0135$ , es decir que el MRG mejora la explicación de los datos por sobre el MRG<sub>R</sub> en un 1.35%. Muy similares resultados se obtienen en la comparación del MGCP y el MCP. El LRT entre el MGCP y el MCP resultó de  $\chi^2_{\Delta}(6) = 5939.43 - 5866.85 = 72.58$ ,  $p = 1.21 \times 10^{-13}$ . En este caso el  $R^2_{\Delta} = .0122$ , señalando una mejora en el modelo completo de 1.22%. Para todos los modelos, ningún ítem desajustó ya que el  $p$  asociado al  $S-\chi^2$  fue mayor que .01, para todos los reactivos.

Dado que un modelo con parámetro  $a$  libre es necesario para mejorar tanto el ajuste como la explicación de estos datos y que el MRG es el modelo con mejor ajuste relativo, se lo selecciona para la modelización con TRI de las respuestas a los ítems de la escala.

Tabla 1

### *Comparación de Ajuste a la ESBA entre Modelos de la TRI*

Modelo	$M_2$	$gl$	$p$ valor	RMSEA	-2lnL	AIC	BIC	DesIt
MRG	391.86	182	.0001	.05	7011.88	7067.88	7186.39	No
MRG <sub>R</sub>	489.08	188	.0001	.06	7108.19	7067.88	7245.30	No
MGCP	413.51	182	.0001	.05	8407.11	5922.85	6036.44	No
MCP	503.87	188	.0001	.06	5939.43	5983.43	6072.68	No

*Nota.* ESBA = Escala de Soledad de Buenos Aires; TRI = Teoría de la Respuesta al Ítem;  $M_2$  = estadístico  $M_2$ ;  $gl$  = grados de libertad; RMSEA = error medio cuadrático de aproximación; -2lnL = valor del logaritmo de la verosimilitud; AIC = Criterio de Información de Akaike; BIC = Criterio de Información bayesiano; DesIt = Existencia de desajuste de ítems al modelo según el índice  $S-\chi^2$ ; MRG = Modelo de Respuesta Graduada; MRG<sub>R</sub> = MRG con restricción de igualdad en el parámetro  $a$ ; MGCP = Modelo Generalizado de Crédito Parcial; MCP = Modelo de Crédito Parcial.

### *Modelización de la ESBA con el MRG*

El MRG fue, de los cuatro modelos, el que resultó más adecuado. Supone que un único  $\theta$  -en este caso el nivel de soledad- explica no linealmente las respuestas a los ítems. Para cada uno de los reactivos fueron calculados un parámetro de inclinación  $a$  y tres parámetros de umbral de categoría  $b_m$ , dado que las opciones de respuesta son cuatro. El parámetro  $a$  informa el grado en que las categorías de respuesta distinguen entre los niveles de  $\theta$ . Además, se lo ha comparado

con las cargas factoriales, en tanto refleja la magnitud de la relación de cada ítem de la escala con el rasgo latente  $\theta$ . Cada parámetro  $b_m$  informa el nivel de soledad  $-\theta$  que se necesita para tener igual probabilidad (.50) de seleccionar la categoría de respuesta  $m$  o superior con respecto a las categorías inferiores. Así, las opciones de respuesta de los ítems son separadas en una serie de dicotomías en cada una de las cuales se aplica el ML2P. Las CCCRI representan la probabilidad de seleccionar cada categoría de respuesta según el nivel de  $\theta$ .

### **Análisis del Funcionamiento Diferencial de los Ítems**

Los resultados de explorar la presencia de DIF utilizando para cada ítem al resto como anclaje se encuentran en la Tabla 2. El ítem 5 - *Estoy completamente fuera de todo grupo social* - fue señalado por los análisis como candidato a tener DIF no uniforme por género, aunque el valor  $p$  estaba muy cercano al límite de .05. Se implementó entonces el segundo paso del método de Woods (2009), tomando al ítem 2 como anclaje entre los dos grupos. Este ítem tiene el menor valor de  $\chi^2_{\text{Total}}$  por lo cual se asume que es el más libre de DIF. El resultado estadísticamente no significativo de este segundo paso indicó que el ítem 5 no exhibe DIF ( $\chi^2_a = 2.6$ ,  $gl = 1$ ,  $p = .1048$ ). Por lo tanto, la totalidad de los ítems de la escala se pueden considerar libres de DIF según el estado civil.

Tabla 2

*Análisis de Funcionamiento Diferencial del Ítem según Estado Civil.  $\chi^2$  de Wald*

Ítem	$\chi^2_{\text{Total}}$	(gl 4)	$p$	$\chi^2_a$ (gl 1)	$p$	$\chi^2_b$ (gl 3)	$p$
1		2.9	.5812	1.5	.2225	1.4	.7123
2		0.1	.9992	0.1	.7937	0.0	.9996
3		5.4	.2477	2.2	.1349	3.2	.3657
4		0.8	.9431	0.3	.6099	0.5	.9180
5		4.2	.3819	3.9	.0491*	0.3	.9563
6		4.6	.3311	0.1	.7216	4.5	.2153
7		1.1	.8947	0.0	.8780	1.1	.7835

Nota. El  $\chi^2_{\text{Total}}$  se refiere al test global, el  $\chi^2_a$  al test para DIF no uniforme y el  $\chi^2_b$  al test para DIF uniforme.

\* $p < .05$ .

### **Calibración de los Ítems con el MRG**

Los resultados de la aplicación del MRG a la escala indicaron que el modelo ajustó tanto globalmente ( $M_2 = 391.86$ ;  $gl = 182$ ;  $p = 0.0001$ ;  $RMSEA = 0.05$ ) como al nivel de los ítems ( $p$  asociado al  $S-\chi^2 > .01$ ). Se estimaron 28 parámetros, cuyos valores se exponen en la Tabla 3.

Los parámetros de umbral de los ítems de encuentran repartidos en un espectro relativamente amplio del rasgo latente, desde -1.43 ( $b_1$  ítem 3) a 3.15 ( $b_3$  ítem 7). Se puede observar una relativa heterogeneidad de los  $b_1$ , mientras que los  $b_2$  se encuentran en niveles medios o altos del rasgo, y los  $b_3$  en niveles aún más altos.

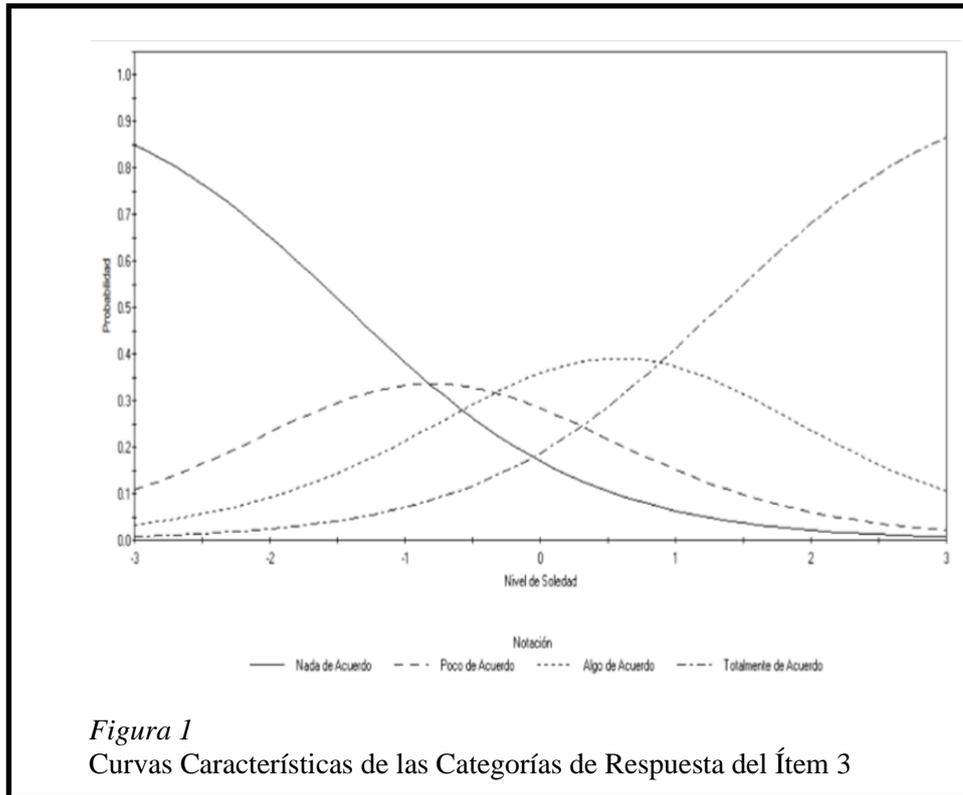
En cuando a los parámetros de discriminación  $a$ , los mismos tomaron valores desde 1.11 a 3.01. Esto indica que las categorías de respuesta son potentes para distinguir entre participantes con diferentes niveles de soledad, siendo su capacidad de discriminación moderada en el caso de los ítems 3 y 7, alta en el caso del ítem 6 y muy alta para los ítems 1, 2, 4 y 5 (Baker & Kim, 2017).

Tabla 3  
*Parámetros estimados mediante MRG para los ítems de la ESBA*

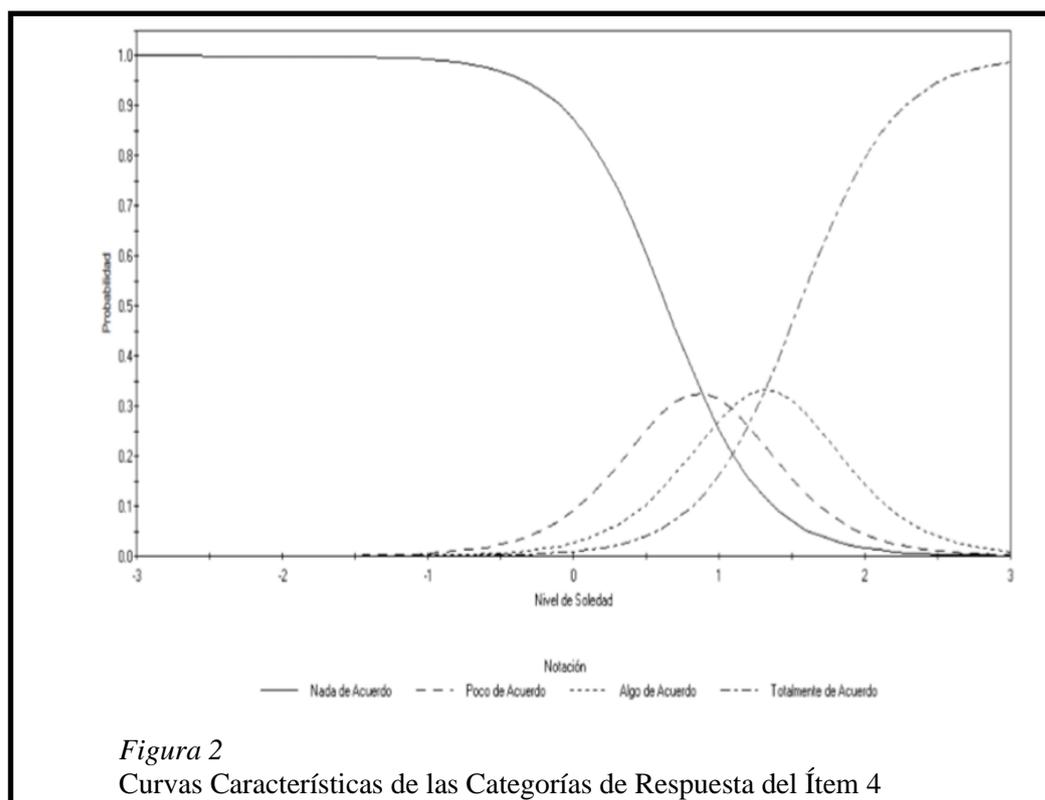
Ítem	$a(s.e)$	$b_1(s.e)$	$b_2(s.e)$	$b_3(s.e)$	$b_{prom}$
1	2.83(0.31)	0.43(0.06)	1.07(0.08)	1.80(0.12)	1.10
2	1.96(0.20)	0.10(0.07)	0.56(0.08)	1.12(0.10)	0.59
3	1.11(0.12)	-1.43(0.17)	-0.17(0.10)	1.32(0.15)	-0.09
4	3.01(0.39)	0.64(0.06)	1.09(0.08)	1.55(0.10)	1.09
5	2.41(0.28)	0.73(0.07)	1.31(0.10)	1.72(0.13)	1.25
6	1.43(0.15)	-0.20(0.09)	1.00(0.11)	2.10(0.19)	0.97
7	1.15(0.13)	0.29(0.10)	1.46(0.16)	3.15(0.34)	1.63
Media	1.99	0.08	0.90	1.82	
DE	0.79	0.74	0.55	0.67	
Mínimo	1.11	-1.43	-0.17	1.12	
Máximo	3.01	0.73	1.46	3.15	

*Nota.* MRG: Modelo de Respuesta Graduada; ESBA: Escala de Soledad de Buenos Aires;  $a$ : parámetro de inclinación calculado con el MRG;  $b_1$ ,  $b_2$  y  $b_3$ : parámetros de umbral calculados con el MRG. *s.e.*: error estándar;  $b_{prom}$ : promedio de los parámetros de umbral  $b$ .

En la Figura 1 se muestran las CCCRI del ítem 3, aquél con menor parámetro  $a$ . Como se puede observar, las CCCRI correspondientes a las categorías de respuesta centrales muestran una forma achatada. Para este ítem si bien todas las opciones de respuesta son máximamente probables en algún nivel del rasgo, la categoría *Poco de Acuerdo* lo es en un rango acotado. Dado el parámetro  $b$  promedio de -0.09 se lo puede considerar un ítem de dificultad media. Es suficiente un nivel muy bajo en el rasgo para seleccionar la categoría *Poco de Acuerdo* o una superior, un nivel medio para seleccionar las categorías *Algo o de Acuerdo* o *Totalmente de Acuerdo* por sobre las dos anteriores y es necesario un nivel muy alto en el rasgo para seleccionar la categoría superior *Totalmente de Acuerdo*. Por tanto, es factible afirmar que las categorías de respuesta se comportan de una manera esperable.



En la Figura 2 se muestran las CCCRI del ítem 4, aquél con parámetro  $a$  más grande. En este caso, las CCCRI correspondientes a las categorías de respuesta extremas muestran una forma elevada. Aunque todas las opciones de respuesta son máximamente probables en algún nivel del rasgo, las opciones intermedias lo son en un lapso muy pequeño. Dado el parámetro  $b$  promedio de 1.09 se lo puede considerar un ítem de dificultad alta. Aunque las categorías de respuesta se comportan de una manera esperable, este ítem sería apto para integrar una prueba dicotómica.



### ***Evidencias de Confiabilidad***

En la Tabla 4 se observan valores puntuales de las FIIs y FIT para determinados niveles de soledad distribuidos a lo largo del continuo del rasgo. La mayoría de los ítems, así como el test completo, evidencian brindar un mayor nivel de información para niveles medios y altos de soledad, donde a su vez disminuye el error estándar (*standard error*, s.e.). Los ítems poseen cierta paridad en el nivel de información máxima que brindan.

Tabla 4  
*Funciones de Información de los Ítems y del Test obtenidas mediante MRG*

Ítem	$\theta$						
	$\theta = -2.4$	$\theta = -1.6$	$\theta = -0.8$	$\theta = 0$	$\theta = 0.8$	$\theta = 1.6$	$\theta = 2.4$
1	0.02	0.08	0.27	0.64	0.87	0.87	0.77
2	0.04	0.15	0.43	0.80	0.89	0.67	0.29
3	0.25	0.61	0.85	0.86	0.82	0.57	0.23
4	0.01	0.05	0.18	0.50	0.84	0.89	0.66
5	0.01	0.05	0.16	0.47	0.81	0.89	0.69
6	0.06	0.22	0.56	0.82	0.87	0.84	0.62
7	0.03	0.12	0.36	0.73	0.87	0.83	0.78
Test	1.44	2.28	3.82	5.82	6.96	6.56	5.05
s.e.	0.83	0.66	0.51	0.41	0.38	0.39	0.45

Nota. MRG = modelo de respuesta graduada;  $\theta$  = nivel de soledad; Test = nivel de información de la Escala de Soledad de Buenos Aires para cada  $\theta$ ; s.e. = error estándar.

La Figura 3 muestra la FIT. Esta alcanzó para el MRG su valor máximo de 6.9983 en  $\theta = 1.00$  con un valor mínimo del s.e. en ese punto de 0.378. El nivel de información fue mayor en los niveles medios y altos del rasgo, decreciendo considerablemente en los niveles bajos del rasgo, así como aquellos altos en extremo.

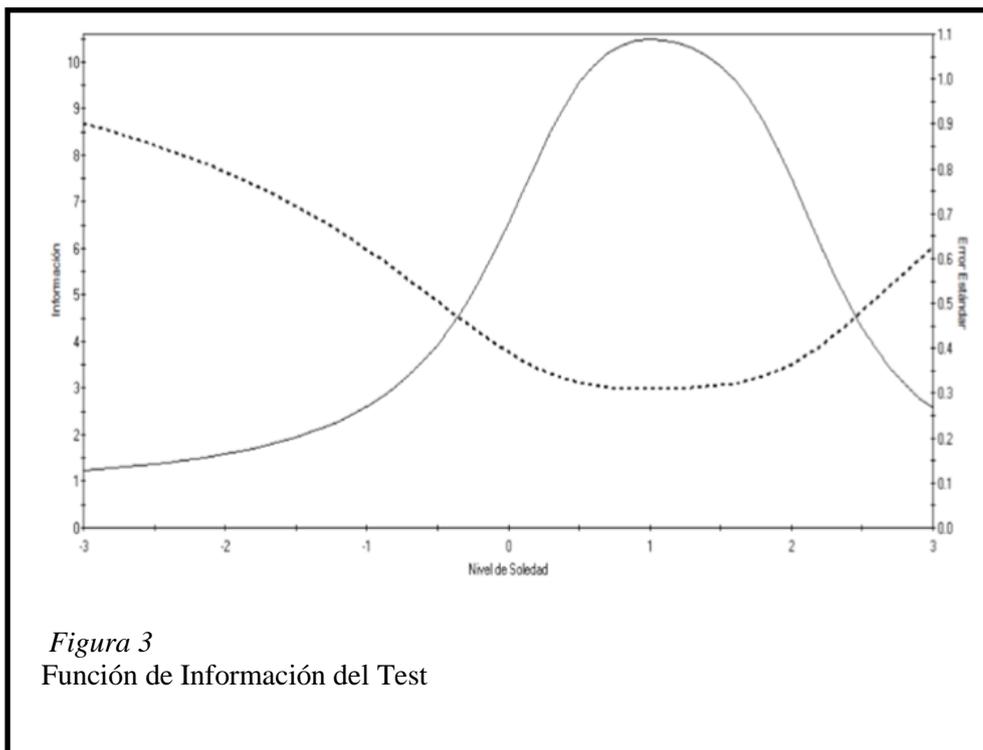


Figura 3  
 Función de Información del Test

## Discusión

Este artículo ha mostrado cómo llevar a cabo un análisis con TRI en sus diferentes aspectos, proveyendo detalles que permiten la replicabilidad a través de la descripción detallada de los pasos necesarios para realizar este tipo de modelización. Es necesario mencionar, para el investigador que desee introducirse en el análisis con TRI, que el programa IRTPRO cuenta con una versión estudiantil, de descarga gratuita, la cual puede ser utilizada para un realizar un acercamiento inicial al análisis con TRI.

En cuanto a los resultados obtenidos, el análisis de la ESBA con el MRG evidenció que la escala tiene un nivel más elevado de precisión en niveles medios y altos del rasgo. El error de medición crece sustancialmente hacia niveles bajos de soledad. La capacidad discriminativa, así como el nivel de información alcanzado tuvieron valores adecuados para la totalidad de los ítems que componen la ESBA. Los nuevos ítems que se incorporen a la ESBA deberían requerir un nivel bajo de soledad para que la escala pueda medir con similar precisión a los distintos niveles del rasgo.

En cuanto al análisis de la adecuación de la cantidad de opciones de respuesta, los valores relativamente alejados de los parámetros  $b_m$  indican su adecuación. Además, resultados empíricos y en base a simulación indican que un diseño de respuesta con cuatro opciones favorece el equilibrio entre la precisión de la medida y el grado de ajuste del modelo de la TRI (e.g. Abal, Auné, Lozzia & Attorresi, 2017; Lozano, García-Cueto & Muñiz, 2008).

Con respecto a un análisis detallado de los reactivos, los ítems 1, 4 y 5 fueron los que brindaron los niveles más altos de información. Además, el parámetro de inclinación  $a$  de estos ítems es muy alto y las distancias entre los parámetros  $b_m$  son amplias, lo que indica que las categorías de respuesta resultan eficaces para discriminar entre participantes con diferentes niveles de soledad. Los restantes ítems, si bien tienen una calidad psicométrica aceptable, son menos informativos que los anteriores.

El análisis del DIF tuvo como resultado la conclusión de que la Escala de Soledad de Buenos Aires se encuentra libre de DIF según Estado Civil, además de por género como ya se había probado en su construcción (Auné et al., 2019). Esto muestra la importancia de los estudios del DIF, tan poco frecuentes sobre todo en el medio latinoamericano. La existencia de ítems con DIF resta validez a la interpretación de los puntajes de la escala, de manera que es necesario testear todos los ítems que se incorporen a la escala en este sentido. Además del DIF estado civil y por género, se puede analizar con respecto a variables sociodemográficas e incluso psicológicas obteniendo así resultados de mucho interés. En futuros estudios se analizará el DIF por edad y por nivel del participante en la Escala de Conducta Empática (Auné et al., 2017a) y en la adaptación argentina de la Escala de Felicidad de Lima (Auné, Abal, & Attorresi, 2017b).

Otra cuestión para destacar es la relevancia de la comparación entre modelos de la TRI. Se pudo observar que no todos los modelos de la TRI ajustaron en el mismo grado a los datos obtenidos empíricamente, las respuestas a la Escala de Soledad de Buenos Aires de esta muestra en particular. Los modelos aplican distintas formas de segmentación del ítem politómico y utilizan diferentes procedimientos para el cálculo de las probabilidades de respuesta de las categorías. Para los tests de comportamiento típico, los resultados de este estudio concuerdan con otros donde el MRG ajustó mejor que los restantes modelos comparados (e.g. Abal, 2013; Asún & Zuñiga, 2008).

## Referencias

- Abal, F. (2013). *Comparación de modelos politómicos y dicotómicos de la Teoría de la Respuesta al Ítem aplicados a un test de Comportamiento Típico*. Tesis de Doctorado, Facultad de Psicología de la Universidad de Buenos Aires.
- Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2017). Funcionamiento de la categoría central en ítems de Confianza para la Matemática. *Evaluar*, 17(2), 18-31.
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955-959. <https://doi.org/10.1177/001316448004000419>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142. <https://doi.org/10.1177/0013164485451012>
- Asún, R. & Zuñiga, C. (2008). Ventaja de los modelos politómicos de teoría de respuesta al ítem en la medición de actitudes sociales. El análisis de un caso. *Psykhé*, 17(2), 103-115.
- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188.
- Auné, S., Abal, F., & Attorresi, H. (2017a). Propiedades psicométricas de una prueba de conducta empática. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 3(45), 47-56. <https://doi.org/10.21865/RIDEP45.3.04>
- Auné, S., Abal, F., & Attorresi, H. (2017b). Versión argentina de la Escala de Felicidad de Lima. *Diversitas*, 13(2), 201-214.
- Auné, S., Abal, F., & Attorresi, H. (2019). Construction and psychometric properties of the Loneliness Scale in adults. *International Journal of Psychological Research*, 12(2), 82-90. <http://dx.doi.org/10.21500/20112084.425782>
- Auné, S., Abal, F., & Attorresi, H. (2020). Modeling of the UCLA Loneliness Scale According to the Multidimensional Item Response Theory. *Current Psychology*, 1-8. <https://doi.org/10.1007/s12144-020-00646-y>
- Baker, F. B., & Kim, S. H. (2017). *The Basics of Item Response Theory Using R*. New York, NY: Springer.
- Bock, R. D. (1997). The nominal categories model. In W. van der Linden & R.K. Hambleton (Eds.), *Handbook of Modern Item response Theory* (pp. 33-50). N.Y.: Springer.
- Cai, L. (2012). flexMIRT: *Flexible multilevel item factor analysis and test scoring* [Computer software]. Seattle, WA: Vector Psychometric Group, LLC.
- Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO user's guide*. Lincolnwood, IL: Scientific Software International.
- Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 36(5), 648-662. <https://doi.org/10.1016/j.clinthera.2014.04.006>
- Chen, W., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- Edelen. M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res.*, 16(1), 5-18. <http://dx.doi.org/10.1007/s11136-007-9198-0>

- Ferrando, P. J., & Loranzo-Seva, U. (2017a). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educ. Psychol. Measur.*, 1-19. <https://doi.org/10.1177/0013164417719308>
- Ferrando, P. J., & Lorenzo-Seva, U. (2017b). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, 29, 236-240.
- Haberman, S. J. (1978). *Analysis of qualitative data: Vol. 1: Introductory topics*. New York, NY: Academic Press.
- Hair, J. F., Anderson, R.E., Tatham, R. L. & Black, W. C. (1999). *Análisis Multivariante*. Madrid, España: Prentice Hall Iberia.
- Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.
- Lozano, L. M., García-Cueto, E. & Muñoz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73-79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Masters, G. N. (2016). Partial Credit Model. En W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume 1: Models (pp. 109-126). Boca Raton: Chapman & Hall/CRC.
- Maydeu Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009-1020. <http://dx.doi.org/10.1198/016214504000002069>
- Maydeu Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71, 713-732. <http://dx.doi.org/10.1007/s11336-005-1295-9>
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm, *Applied Psychological Measurement*, 16, 159-176. <https://doi.org/10.1002/j.2333-8504.1992.tb01436.x>
- Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64. <https://doi.org/10.1177/01466216000241003>
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S- $\chi^2$ : An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27, 289-298. <https://doi.org/10.1177/0146621603027004004>
- Paek, I., & Cole, K. (2019). *Using R for Item Response Theory Model Applications*. New York, NY: Routledge.
- Sacchi, C. & Richaud de Minzi, M. C. (1997). La Escala Revisada de Soledad de UCLA: Una adaptación argentina. *Rev. Argent. Clín. Psicol.*, 6(1), 43-53.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*, 17(4), 2. <http://dx.doi.org/10.1002/j.2333-8504.1968.tb00153.x>
- Samejima, F. (2016). Graded response models. In Wim J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume One* (pp. 123-136). Chapman and Hall/CRC.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16, 209-220. <http://dx.doi.org/10.1037/a0023353>
- Toland, M. (2013). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, 34(1), 120-151. <https://doi.org/10.1177/0272431613511332>
- Woods, C. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33(1), 42-57. <https://doi.org/10.1177/0146621607314044>

Financiamiento: Esta investigación se ha realizado dentro del proyecto UBACyT 2018 con código 20020170100200BA, subvencionado por la Facultad de Psicología de la Universidad de Buenos Aires, el proyecto UBACyT 2018 con código 20020170200001BA subvencionado también por la Facultad de Psicología de la Universidad de Buenos Aires y el proyecto PICT-2017-3226 subvencionado por la Agencia Nacional de Promoción Científica y Tecnológica.

Contribución de los autores: a) Concepción y diseño del trabajo; b) Adquisición de datos; c) Análisis e interpretación de datos; d) Redacción del manuscrito; e) revisión crítica del manuscrito. S.E.A. ha contribuido en a,b,c,d,e; F.J.P.A. en a,b,c,d,e; H.F.A. en a,b,c,d,e.

Editora científica responsable: Dra. Cecilia Cracco