

Análisis Psicométrico de una Escala de Ayuda con el Modelo de Respuesta Graduada

Psychometric Analysis of a Helping Scale with the Graded Response Model

Sofía Esmeralda Auné y Facundo Juan Pablo Abal

Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina

Horacio Félix Attorresi
Universidad de Buenos Aires

Se presentan los resultados obtenidos a partir de la aplicación del modelo de respuesta graduada (MRG) a los ítems de la Escala de Ayuda (EA), una prueba que mide el comportamiento de ayuda, un tipo de conducta prosocial, compuesta por ocho ítems. La muestra fue obtenida por conveniencia y estuvo compuesta por 1383 estudiantes universitarios argentinos (80% mujeres). El 73,5% de los estudiantes cursaba su carrera en la Universidad de Buenos Aires, el 18,9% en otras universidades nacionales públicas y el 7,6% en universidades privadas de la Ciudad de Buenos Aires. Uno de los ítems referido a altruismo presentó funcionamiento diferencial del ítem no uniforme por género y fue eliminado de la escala. Se corroboraron los supuestos del MRG de independencia local con el índice chi cuadrado LD y unidimensionalidad mediante análisis factorial confirmatorio, considerándose satisfechos. Posteriormente, se implementó el MRG. Los parámetros de umbral promedio tendieron a ubicarse en niveles medios y altos de la escala del rasgo, mientras que los parámetros de inclinación evidenciaron una alta capacidad de los ítems para distinguir entre participantes con distinto nivel de ayuda. El análisis con la función de información del test mostró una precisión aceptable en un rango amplio de niveles de ayuda, donde la mayoría de las personas podría encuadrarse. Con respecto a la confiabilidad, la misma resultó adecuada. Un único ítem que se refiere al ámbito estudiantil tuvo un nivel de información bajo y plano y parámetros de umbral fuera de rango, por lo que deberá ser reformulado.

Palabras clave: ayuda, conducta prosocial, modelo de respuesta graduada, teoría de la respuesta al ítem, escala

This article presents the results of the application of the graded response model (GRM) to the items of the Helping Scale, an eight-item test which measures helping behavior, a type of prosocial behavior. A convenience sampling strategy was employed, which made it possible to enroll 1383 Argentine university students (80% women). 73.5% of the participants attended the University of Buenos Aires, 18.9% attended other national public universities, and 7.6% attended private universities in the City of Buenos Aires. One of the items, which referred to altruism, presented gender non-uniform differential functioning and was removed from the scale. The GRM assumptions of local independence and one-dimensionality were successfully corroborated with the chi square LD index and confirmatory factor analysis, respectively. After that, the GRM was implemented. Average threshold parameters tended to be located at middle and high levels of the trait scale, while slope parameters displayed high item capacity to distinguish between participants with different levels of helping behavior. The analysis of the information function of the test showed acceptable accuracy in a wide range of helping levels, encompassing most people. Reliability was also adequate. One item, which refers to the academic domain, exhibited a low and flat level of information and out-of-range threshold parameters; therefore, it will need to be reformulated.

Keywords: helping, prosocial behavior, graded response model, item response theory

La conducta de ayuda forma parte del comportamiento prosocial (Auné, Blum, Abal, Lozzia & Attorresi, 2014; Dunfield, 2014), es decir, se considera una acción positiva y voluntaria que beneficia a otros. Dunfield (2014) puntualiza que ayudar alivia la necesidad instrumental del beneficiario. Estas acciones no solo influyen directamente en quien las recibe, sino que además resultan positivas para quien las realiza, así

Sofía Esmeralda Auné y Facundo Juan Pablo Abal, Instituto de Investigaciones, Facultad de Psicología, Universidad de Buenos Aires, CONICET, Buenos Aires, Argentina; Horacio Félix Attorresi, Instituto de Investigaciones, Facultad de Psicología, Universidad de Buenos Aires, Argentina.

Esta investigación fue financiada con los subsidios de la Universidad de Buenos Aires, UBACyT 2018 Código 20020170100200BA y UBACyT 2018 Código 20020170200001BA, y de la Agencia Nacional de Promoción Científica y Tecnológica, PICT-2017-3226.

La correspondencia relativa a este artículo debe ser dirigida a Sofía Esmeralda Auné, Virrey Liniers 577 piso 15 departamento B, Ciudad Autónoma de Buenos Aires, Código Postal 1220, Argentina. E-mail: sofiaaune177@hotmail.com

como para la sociedad entera (Aknin, Broesch, Hamlin & Van de Vondervoort, 2015; Klein, 2017). Cobran importancia en el actor de la conducta de ayuda los valores internalizados y la autoimagen prosocial (Eisenberg, Hofer, Sulik & Liew, 2014).

La ayuda se entiende como un comportamiento de asistencia, cuidado y compromiso con los otros, implicando necesariamente un benefactor y un beneficiario claramente diferenciados (Auné, 2018; Dunfield, 2014; Warneken & Tomasello, 2009). Los comportamientos de ayuda incluyen una variedad de acciones, tales como rescate, donación, asistencia, voluntariado y apoyo social, incluyendo cada una de estas categorías una amplia gama de comportamientos específicos (Auné, Blum et al., 2014), solicitados o no explícitamente (Chou & Stauffer, 2015). Spitzmuller y Van Dyne (2013) clasificaron globalmente al comportamiento de ayuda en acciones proactivas y reactivas. Mientras que las primeras se focalizan en satisfacer las necesidades del actor, las segundas se centran en beneficiar al receptor. Aunque el comportamiento de ayuda es el prototipo de conducta prosocial, este último es un concepto más amplio (Drummond, Hammond, Satlof- Bedrick, Waugh & Brownell, 2017; González Portal, 2000). Por ejemplo, confortar a un amigo en un momento difícil puede ser considerada una acción prosocial, pero no se incluye en la categoría de ayuda.

Actualmente, el comportamiento de ayuda se considera complejo y con una determinación múltiple, siendo abordado desde diversas perspectivas (Auné, Blum et al., 2014; Chou & Stauffer, 2015). Se considera que una variedad de factores causales, circunstancias moderadoras y mecanismos mediadores se conjugan para la concreción de este tipo de acciones (Penner, Dovidio, Piliavin & Schroeder, 2005). Por ejemplo, la ayuda humanitaria en adultos se asocia con la preocupación por los derechos y necesidades humanas, el bajo nivel de etnocentrismo, la igual valoración de las vidas de los miembros del endo y exogrupo y el conocimiento de los temas relevantes del humanitarismo (McFarland, Brown & Webb, 2013). A un nivel más general, la preocupación empática y la responsabilidad social son consideradas el núcleo de la personalidad prosocial, que, a su vez, predispone a la acción de ayudar (Ruci, van Allen & Zelenski, 2018).

Uno de los factores que influyen en el comportamiento de ayuda es el género, obteniendo las mujeres, en general, mayores niveles (Auné, Abal & Attorresi, 2015). Los estereotipos tradicionales suponen una realización diferenciada de la conducta de ayuda. Mientras que la mujer realiza tareas de cuidado y ayuda relacional al interior del propio hogar o dentro del círculo social cercano, el hombre actúa en situaciones de urgencia e interviene colectivamente (Eagly, 2009). Actualmente, estos estereotipos se encuentran en un importante proceso de transformación, pero aun así mantienen cierta vigencia.

Con respecto a la medición de la ayuda en adultos, se destaca en el plano internacional la Escala para la Medición de la Prosocialidad en Adultos (*Prosocialness Scale for Adults* [PSA]) de Caprara, Steca, Zelli y Capanna (2005), por su brevedad, sencillez y alcance de los estudios psicométricos realizados. Estos permiten concluir no solo acerca de su consistencia interna, sino también acerca de su capacidad de discriminación entre sujetos. Los autores utilizan para la modelización de la PSA la teoría de la respuesta al ítem (TRI). Como resultado de este análisis, se desprende que los parámetros de umbral de los ítems que componen la escala se ubican mayoritariamente en los niveles bajos de ayuda, donde se presenta la mayor capacidad discriminativa. Esto supone que incluso quienes se encuentran por debajo de la media de ayuda tenderán a responder en forma favorable al rasgo. En definitiva, los contenidos de los ítems presentan una falencia, al no medir por igual en todos los niveles del rasgo. Es decir, como puntualizan Caprara et al. (2005), los ítems de su escala resultan "fáciles" y esta facilidad se da particularmente en ítems de ayuda, así como en aquellos que implican compartir y mostrar sensibilidad hacia los sentimientos de otros. Esta facilidad implica que la escala presenta dificultades para distinguir entre personas con niveles medios y altos de ayuda. Es justamente este déficit en las propiedades psicométricas de la PSA lo que motivó la construcción de un nuevo instrumento psicométrico: la Escala de Ayuda (EA).

Este no es el único motivo que establecería la necesidad de construir una nueva escala. Conceptualmente, las conductas de ayuda son una parte de las conductas prosociales de asistencia, cuidado y compromiso con un otro, suscitando particular interés. A continuación, se expone un recorrido por las principales escalas existentes para evaluar ayuda.

Medición de Comportamientos de Ayuda

Habitualmente los comportamientos de ayuda se evalúan subsumidos dentro de escalas globales que miden conducta prosocial, como es el caso de la mencionada PSA de Caprara et al. (2005). Los siguientes

instrumentos de medición psicológica incluyen ítems referidos a comportamientos de ayuda: (a) Inventario de Habilidades Sociales para Adolescentes (*Teenage Inventory of Social Skills* de Inderbitzen & Foster, 1992), (b) Escala de Conducta Prosocial (*Prosocial Behavior Scale* de Caprara & Pastorelli, 1993), (c) Escala de Tendencias Prosociales (*Prosocial Tendencies Measure* de Carlo & Randall, 2002) y (d) Cuestionario de Conducta Prosocial (Sánchez-Queija, Oliva Delgado & Parra Jiménez, 2006).

También se ha medido la ayuda construyendo una subescala específica: (a) en la Escala de Habilidades Prosociales para Adolescentes (Morales Rodríguez & Suárez Pérez, 2011, Noviembre) existen las subescalas Respuesta de Ayuda y la de Asistencia; (b) en la Batería de Personalidad Prosocial (*Prosocial Personality Battery* de Penner, Fritzsche, Craiger & Freifeld, 1995), la Tendencia a Ayudar, que describe la participación en actividades de ayuda, es una de las dos subescalas que la compone y (c) en la escala de Conducta Prosocial de Nielson, Padilla-Walker y Holmes (2017), una de sus subescalas es Ayuda Física.

En cuanto a las escalas que midan exclusivamente ayuda, su número es bastante reducido. La Escala de Actitudes de Ayuda (*Helping Attitude Scale* de Nickell, 1998, Agosto) es una medida multidimensional orientada a medir creencias, sentimientos y acciones relacionadas con la ayuda. Anteriormente, se había desarrollado el Cuestionario de Orientación hacia la Ayuda (*Helping-Orientation Questionnaire* [HOQ] de Romer, Gruder & Lizzadro, 1986). El HOQ mide cuatro orientaciones: altruista, ayuda receptiva, sostenimiento propio y egoísta. Posteriormente, se construyeron escalas para grupos específicos, como la Escala de Altruismo en los Maestros (*Teacher Altruism Scale* de Yavuzer et al., 2006) y la Escala de Motivos para la Ayuda en Enfermeros (*Nursing Motives for Helping Scale* de González Gutiérrez, Peñacoba Puente, Moreno Rodríguez, López López & Velasco Furlong, 2006).

Resultaría útil, entonces, contar con una escala de evaluación psicológica que permita medir específicamente el comportamiento de ayuda en forma general, breve, actualizada y con reactivos de fácil comprensión. Por supuesto, al tratarse de un autorreporte, se está evaluando en realidad la percepción de la propia conducta de ayuda a otros. Los modelos reflectivos que se utilizarán para modelizar la escala son adecuados, ya que existe una variable latente no observable directamente (Borsboom, Cramer, Kievit, Scholten, & Franić, 2009) que subyace a la conducta de ayuda.

La Escala de Ayuda

La EA es un instrumento creado, diseñado y validado por Auné (2018). Los ítems fueron construidos intencionalmente de forma tal de que resulten exigentes con respecto al nivel de ayuda requerido, para así cubrir el vacío que se manifiesta en la PSA (Caprara et al., 2005), como se ha mencionado anteriormente. Es decir, se procuró incluir comportamientos que potencialmente requieran niveles medios o elevados de ayuda para ser realizados y, más concretamente, acciones de ayuda altruista. La EA se compone de ocho reactivos que se responden en una escala Likert (1 = *Nunca*, 2 = *Casi nunca*, 3 = *A veces*, 4 = *Con frecuencia*, 5 = *Casi siempre*, 6 = *Siempre*). Estos reactivos conforman una estructura unidimensional (Auné, 2018). En cuanto a las correlaciones ítem-total corregidas y las comunales, las mismas fueron adecuadas. La consistencia interna, evaluada por alfa de Cronbach, fue buena ($\alpha = 0,77$). Se aportaron evidencias de validez convergente de los puntajes de la EA con las dimensiones Asistencia y Altruismo de la Escala de Habilidades Prosociales de Morales Rodríguez y Suárez Pérez (2011, Noviembre), versión argentina (Auné, Abal & Attorresi, 2014), y con la autopercepción del participante en conducta prosocial.

Auné (2018) realizó un análisis preliminar de la EA, modelizándola con la TRI. Se utilizó el programa MULTILOG (Thissen, Chen & Bock, 2003), aunque este tiene ciertas limitaciones, por ejemplo, en cuanto a los indicadores de ajuste de los modelos. Asimismo, Auné (2018) analizó el funcionamiento diferencial del ítem (*differential item functioning*, DIF) de los ítems con el procedimiento de Mantel-Haenszel (MH), una prueba de tabla de contingencia muy potente para la detección de DIF uniforme, pero inadecuada para detectar DIF no uniforme (Osterlind & Everson, 2009). Mediante MH, se detectó un ítem con DIF uniforme por género, *Hago donaciones a organizaciones benéficas*. El mismo fue eliminado de la EA. Sin embargo, es necesario implementar un procedimiento más exigente en este sentido, apto para detectar además DIF no uniforme en los ítems que componen la EA.

La ayuda es un comportamiento producto de rasgos y tendencias individuales y sociales, que se ha relacionado con facetas de la personalidad y con la predisposición a realizar acciones prosociales. Esta predisposición, que subyace a la conducta de ayuda, se ha denominado prosocialidad (Caprara, Alessandri & Eisenberg, 2012) o personalidad prosocial (Penner et al., 1995). El análisis de la EA mediante la TRI permite corroborar que los ítems reflejan el constructo latente, que el mismo es unidimensional —ya que,

más allá de que se haya corroborado la unidimensionalidad separadamente, su falta se reflejaría en un desajuste del modelo TRI—, que los ítems son igualmente representativos del constructo para varones y mujeres y los niveles del constructo para los que es más preciso el test.

El presente estudio. El objetivo general de este trabajo fue aplicar el modelo de respuesta graduada (MRG) de Samejima (1969, 2016) al análisis de los ítems que componen la EA. Los objetivos específicos fueron los siguientes: (a) realizar estudios de detección del DIF uniforme y no uniforme según el género de los participantes, aplicando TRI; (b) calibrar los ítems de la EA mediante los parámetros de las curvas características de las categorías de respuesta del ítem (CCCRI) del MRG y (c) analizar, mediante la función de información de los ítems y de la escala completa, para qué niveles del rasgo ayuda es elevada la precisión de la EA.

Método

El estudio formó parte de dos proyectos de investigación UBACyT: Aislamiento Social Autopercebido y Razonamiento: su Relación en Diferentes Contextos Socioeconómicos y Educativos (Director: Prof. Horacio Attorresi, Código 20020170100200BA) y Manifestaciones del Neuroticismo Asociadas a la Actividad de Personas Adultas en las Redes Sociales Virtuales (Director: Dr. Facundo Juan Pablo Abal, Código 20020170200001BA).

Participantes

La muestra fue por conveniencia. Se solicitó la colaboración voluntaria de estudiantes universitarios de Argentina. Los participantes fueron contactados mediante redes sociales, donde se invitaba a participar en la investigación mediante publicaciones de solicitudes generales en grupos de estudiantes. Luego de rechazar los protocolos con más de 10% de respuestas incompletas, la muestra estuvo compuesta por 1434 sujetos, de los cuales 80% fueron mujeres y 20% varones. La edad promedio fue de 22,4 años ($DE = 5,8$), con un mínimo de 18 años y un máximo de 80. Las carreras con mayor peso en la muestra fueron Psicología (29%), Medicina (10%), Abogacía (7%), Contador Público (7%), Licenciado en Administración de Empresas (4,2%) y Licenciado en Artes Audiovisuales (3,6%). El resto de los sujetos se repartieron en más de 50 carreras diferentes. El 73,5% de los estudiantes cursaba su carrera en la Universidad de Buenos Aires (universidad pública de la Ciudad de Buenos Aires), el 18,9% en otras universidades nacionales públicas como la Universidad Nacional de La Plata, la Universidad Nacional de Rosario y la Universidad Nacional de Lomas de Zamora y el 7,6% en universidades privadas de la Ciudad de Buenos Aires. Generalmente, en la Argentina los estudiantes de universidades públicas tienen un estatus socioeconómico medio-bajo o medio y los estudiantes de universidades privadas un estatus socioeconómico alto. Por otra parte, si bien la muestra se encuentra desbalanceada por género, se contó con un número de participantes masculinos que posibilitaba realizar los análisis propuestos.

Instrumentos

Cuestionario sociodemográfico. Realizado ad hoc para el presente estudio. Se indagaron variables tales como edad, género, carrera universitaria que cursa y grado de avance en la misma.

Escala de Ayuda ([EA] de Auné, 2018). Ya se han expuesto previamente sus propiedades psicométricas. Los reactivos que la componen son: 1) Dedico tiempo para facilitar la cursada a mis compañeros, 2) Si tuviera un único plato de comida, lo compartiría con un desconocido, 3) Me comprometo con causas nobles, 4) Dedico un tiempo importante de mi vida a mejorar el mundo, 5) Expreso abiertamente mi apoyo a minorías discriminadas aunque no pertenezca a las mismas, 6) Doy dinero a cualquiera que lo necesite, 7) Me ocupo del bienestar de cualquier individuo, grupo o comunidad y 8) Me quedo con lo justo y necesario para vivir y reparto todo lo demás.

Procedimiento

El estudio fue aprobado en sus aspectos académicos y éticos por la Facultad de Psicología de la Universidad de Buenos Aires.

Se recolectaron los datos en el período de enero a mayo de 2017, donde se incluía un consentimiento informado. Los participantes respondieron on-line, incluyendo un consentimiento informado anónimo. Se aclaraba que el uso era exclusivamente académico y la participación enteramente voluntaria, pudiendo cesar en cualquier momento. Los participantes no recibieron recompensas económicas. Se garantizó la confidencialidad de la información entregada por los participantes y el resguardo de cualquier elemento vinculado con su identidad.

Análisis de Datos

En primer lugar, se analizaron los datos anómalos multivariados. Para ello se calculó la distancia de Mahalanobis, utilizando un punto de corte de $p \leq 0,001$ (Tabachnick & Fidell, 2001). Los datos anómalos u *outliers* multivariantes son observaciones que se consideran extrañas, no por el valor que toman en una determinada variable, sino en un conjunto de variables (Muñoz García & Amón Uribe, 2013). En este caso, se trata de un patrón de respuestas a los ítems muy poco probable. En forma progresiva se eliminaron 51 casos de la muestra que cumplieron el criterio para ser considerados *outliers*. La cantidad total de individuos analizados, entonces, fue de 1383, siendo el tamaño de la muestra original, como ya se indicó, de 1434 participantes.

Para modelizar los ítems de la EA, se utilizó el MRG. Este es una extensión al caso politómico del modelo logístico de dos parámetros para ítems de respuesta dicotómica. El modelo es apropiado cuando las respuestas a un ítem pueden ser clasificadas en más de dos categorías ordenadas y su aplicación se basa en tres supuestos básicos: unidimensionalidad, independencia local y ausencia de DIF.

Dado que la existencia de un único factor preponderante es un supuesto del MRG, las respuestas de los 1383 participantes se utilizaron para realizar un análisis factorial confirmatorio (AFC) del modelo unidimensional. El AFC se implementó mediante el programa Mplus (Muthén & Muthén, 2010). Para respetar el carácter ordinal de los datos, el análisis se realizó en base a las matrices de correlaciones policóricas. Se estimaron los parámetros con el método de mínimos cuadrados ponderados robustos. Siguiendo las recomendaciones de Abad, Olea, Ponsoda y García (2011), el ajuste del modelo se examinó mediante el índice de ajuste comparativo de Bentler-Bonett (CFI), el índice de Tucker-Lewis (TLI) y la raíz del error medio cuadrático de aproximación (RMSEA). Estos son los índices de bondad de ajuste más utilizados al aplicar AFC con datos categóricos (Byrne, 2012). Si bien no existe un consenso acabado sobre los puntos de cortes de estos índices de ajuste, es posible aceptar razonablemente el ajuste del modelo para CFI y $TLI \geq 0,90$ y $RMSEA \leq 0,08$ (Abad et al., 2011; Brown, 2015; Byrne, 2012). También se obtuvo el índice de ajuste χ^2 , aunque su significación estadística resulta sensible al tamaño muestral y las cargas factoriales. Además, se calcularon los índices de modificación sobre el modelo.

El supuesto de independencia local se evaluó para cada par de ítems a través del índice χ^2_{LD} propuesto por Chen y Thissen (1997). A mayor valor de este índice, más dependencia local (*local dependence*, LD), donde si excede el valor de 10, se considera que habría LD (Cai, Thissen & du Toit, 2011). El índice χ^2_{LD} puede tomar valores negativos, porque adiciona al estadístico χ^2 diferencias estandarizadas de *odds ratio* que permiten determinar la dirección de la asociación entre los dos ítems. Un valor positivo indica mayor LD de las frecuencias observadas que la que el modelo predice, mientras que un valor negativo indica que las frecuencias observadas del par de ítems tienen menor LD que la esperada por el modelo (Chen & Thissen, 1997).

Se realizó un análisis de DIF para determinar si el género podría influir en las respuestas a los ítems de la EA. Si las mujeres responden diferencialmente que los varones, esto significa que los mismos puntajes en un ítem representan diferentes niveles de ayuda entre ambos grupos. La presencia de DIF es contraria a la validez e implica falta de unidimensionalidad. Cuando se estudia el DIF en el marco de la TRI, se considera que un ítem tiene funcionamiento diferencial cuando las curvas características del ítem —en el caso de ítems de respuesta dicotómica— o las CCCRI —en el caso de ítems de respuesta politómica— no coinciden entre los grupos (Lord, 1977, 1980). Existen básicamente dos tipos de DIF. Mientras que el DIF uniforme afecta al parámetro b_m , el DIF no uniforme afecta al menos al parámetro a (Harpole et al., 2015).

Se analizó el DIF mediante el test de Wald modificado, un enfoque basado en la TRI apto para detectar DIF uniforme y no uniforme. En primer lugar, el test de Wald (1943) fue modificado por Lord (1980) y posteriormente la versión de Lord fue mejorada por Cai (2012), Cai et al. (2011) y Langer (2008). Las mejoras incluyeron una mayor exactitud en la estimación de la matriz de covarianzas, que el rasgo latente se mantenga constante entre grupos y que se estimen parámetros de ítems en lugar de ecuaciones ad hoc

(Woods, Cai & Wang, 2013). El test de Wald modificado se calculó utilizando el programa estadístico IRTPRO 4.2 (Cai et al., 2011).

El test de Wald modificado se utilizó en el marco del procedimiento recomendado por Woods (2009) para la detección del DIF. El mismo consiste en, primeramente, testear cada uno de los ítems de la escala, utilizando el resto como anclaje para detectar los posibles ítems que pueden tener DIF. Luego, se realiza un nuevo análisis individual para cada ítem estudiado candidato a presentar DIF. Se utilizan como anclaje uno o más ítems, aquel o aquellos con menor valor en el estadístico (en este caso, el χ^2 de Wald) que se presumen libres de DIF. Woods (2009) indica que el uso de un único ítem de anclaje disminuye sustancialmente las chances de contaminación con ítems de anclaje que presenten DIF y resulta adecuado cuando el tamaño de la muestra es grande, como en el caso del presente estudio.

Complementariamente al análisis de DIF, se realizó un análisis de invarianza por género, consistente, en primer lugar, en aplicar el modelo confirmatorio a mujeres y varones por separado y evaluar su ajuste. En un segundo modelo, se añadió la constante de que las cargas factoriales fueran iguales para los dos grupos y, en un tercero, se agregó la restricción de igualdad en los errores de las varianzas.

Una vez corroborados sus supuestos, se implementó el MRG (Samejima, 1969, 2016), adecuado para los ítems politómicos de la EA. Para los análisis de la TRI se utilizó el programa IRTPRO 4.2 (Cai et al., 2011). En el MRG, cada CCCRI describe la probabilidad de elegir una opción de respuesta al ítem para cada nivel del rasgo medido por este. Asimismo, para cada ítem son estimados un parámetro de inclinación α y un conjunto de parámetros de umbral de categoría b_m . Los b_m son los puntos del rasgo latente, donde la probabilidad de elegir la categoría m o una superior es igual a 0,50. El rango esperable de valores para los b_m es de -3 a 3. El parámetro de inclinación α es constante para todas las CCCRI de cada ítem e informa el grado en que sus categorías de respuesta distinguen entre los niveles del rasgo. En este caso, el MRG consideró un único rasgo latente representando el nivel de ayuda. Los parámetros de umbral b_m fueron cinco, por existir seis opciones de respuesta. Mediante el parámetro de inclinación α y la distancia entre los parámetros b_m de un mismo ítem, se examinó si cada una de las categorías de respuesta resultó eficaz para discriminar entre participantes con diferentes niveles de ayuda.

El ajuste global del modelo se calculó con el estadístico M_2 y su RMSEA asociado (Cai, Maydeu-Olivares, Coffman & Thissen, 2006; Maydeu-Olivares & Joe, 2005, 2006; Thissen, 2009). El estadístico M_2 puede sobredimensionar pequeños desajustes del modelo; es por ello que se acompaña de un RMSEA, considerándose ajuste adecuado si $RMSEA < 0,05$ (Toland, 2014). Para evaluar el ajuste de cada ítem al modelo, se utilizó el índice $S\text{-}\chi^2$ (Orlando & Thissen, 2000, 2003). En el caso de que la muestra sea grande en combinación con una escala corta —como en este caso—, es pertinente considerar ajuste para el índice $S\text{-}\chi^2$ si $p > 0,01$ (Toland, 2014).

La función de información del ítem (FII) fue utilizada para verificar la precisión de cada ítem en la medición de la ayuda. Para medir la confiabilidad global en el marco de la TRI, se obtuvo la función de información del test (FIT) y el índice de confiabilidad marginal (Green, Bock, Humphreys, Linn & Reckase, 1984).

Resultados

Cumplimiento de Supuestos de la TRI: Independencia Local y Unidimensionalidad

Para comprobar la independencia local se estimó el valor del estadístico χ^2_{LD} de Chen y Thissen (1997) para cada uno de los pares de ítems. Su valor osciló entre -1,9 y 5,2, es decir, no se excedió el valor límite de 10. Así, se dio por satisfecho dicho supuesto.

Los indicadores de ajuste comparativos muestran un resultado muy adecuado para la verificación del modelo unidimensional, $CFI = 0,972$, $TLI = 0,961$, y asimismo resultó satisfactorio el índice RMSEA, 0,067, IC 90% [0,056, 0,077], por lo cual se acepta el cumplimiento del supuesto de unidimensionalidad para la EA. El valor de chi cuadrado fue $\chi^2(gl = 20, n = 1383) = 118,06$, $p < 0,0001$, donde su significación estadística puede ser debida al tamaño muestral. Las cargas factoriales para cada uno de los ítems fueron: ítem 1 = 0,37, ítem 2 = 0,53, ítem 3 = 0,76, ítem 4 = 0,71, ítem 5 = 0,57, ítem 6 = 0,51, ítem 7 = 0,69 e ítem 8 = 0,52. Con respecto a los índices de modificación, ninguno de los multiplicadores de Lagrange fue estadísticamente significativo y los valores residuales de la matriz de las predicciones de la covarianza y

correlación tuvieron un promedio de 0,0226, exhibiendo un valor máximo de 0,0668. Estos resultados indican que no se requiere una reespecificación del modelo (Cupani, 2012).

Análisis de Funcionamiento Diferencial del Ítem

En primer lugar, se testeó en cada uno de los ítems de la EA la presencia de DIF, utilizando el resto como anclaje. Los resultados de este análisis se encuentran en la Tabla 1, mientras que los parámetros de los ítems por género estimados mediante MRG se pueden observar en la Tabla 2. Como se puede apreciar, el ítem 8 —*Me quedo con lo justo y necesario para vivir y reparto todo lo demás*— mostró evidencias de DIF no uniforme por género. Sus categorías resultan más exigentes y menos discriminantes para los participantes varones con respecto a las participantes mujeres. Esta mayor exigencia se acentúa para las categorías superiores de respuesta.

Tabla 1
Análisis de Funcionamiento Diferencial del Ítem entre Mujeres y Varones: χ^2 de Wald

Ítem	χ^2 total	gl	p	χ^2_a	gl	p	χ^2_b	gl	p
1	7,4	6	0,2849	0,3	1	0,5736	7,1	5	0,2123
2	10,8	6	0,0958	0,0	1	0,8313	10,7	5	0,0572
3	10,4	6	0,1067	0,0	1	0,9854	10,4	5	0,0633
4	5,9	6	0,4413	0,3	1	0,5831	5,5	5	0,3539
5	8,8	6	0,1860	0,1	1	0,7746	8,7	5	0,1216
6	4,5	6	0,6080	0,1	1	0,7430	4,4	5	0,4933
7	1,0	6	0,9858	0,0	1	0,8974	1,0	5	0,9642
8	14,8	6	0,0220	10,0	1	0,0015	4,8	5	0,4472

Nota. El χ^2 total se refiere al test global, el χ^2_a , al test para DIF no uniforme y el χ^2_b , al test para DIF uniforme.

Tabla 2
Parámetros de los Ítems Estimados por MRG para el Análisis de DIF por Género

Ítem	Género	a (EE)	b_1 (EE)	b_2 (EE)	b_3 (EE)	b_4 (EE)	b_5 (EE)
1	M	0,65 (0,07)	-3,38 (0,35)	-1,33 (0,16)	1,52 (0,17)	3,73 (0,38)	5,74 (0,61)
	V	0,78 (0,13)	-3,35 (0,50)	-1,03 (0,20)	1,26 (0,29)	3,16 (0,58)	6,32 (1,34)
2	M	0,98 (0,07)	-2,65 (0,20)	-1,12 (0,10)	0,31 (0,07)	0,96 (0,09)	1,58 (0,13)
	V	1,01 (0,15)	-2,18 (0,28)	-0,77 (0,15)	0,66 (0,18)	1,25 (0,23)	1,97 (0,32)
3	M	2,06 (0,13)	-2,53 (0,14)	-1,30 (0,07)	0,16 (0,04)	0,92 (0,06)	1,66 (0,08)
	V	2,05 (0,25)	-2,19 (0,19)	-1,25 (0,12)	0,01 (0,09)	0,79 (0,12)	1,86 (0,21)
4	M	1,79 (0,11)	-1,75 (0,09)	-0,27 (0,05)	0,94 (0,06)	1,67 (0,09)	2,43 (0,13)
	V	1,72 (0,22)	-1,98 (0,19)	-0,52 (0,10)	0,71 (0,12)	1,48 (0,18)	2,38 (0,29)
5	M	1,11 (0,08)	-2,88 (0,20)	-1,60 (0,12)	-0,36 (0,07)	0,48 (0,07)	1,35 (0,10)
	V	1,21 (0,17)	-2,36 (0,26)	-1,39 (0,17)	-0,50 (0,12)	0,36 (0,14)	1,02 (0,19)
6	M	1,01 (0,08)	-2,46 (0,18)	-0,51 (0,08)	1,36 (0,11)	2,35 (0,17)	3,62 (0,27)
	V	0,97 (0,15)	-2,44 (0,31)	-0,72 (0,15)	1,10 (0,23)	2,36 (0,39)	3,85 (0,65)
7	M	1,69 (0,10)	-1,98 (0,11)	-0,65 (0,06)	0,81 (0,06)	1,81 (0,10)	2,69 (0,15)
	V	1,69 (0,21)	-1,99 (0,19)	-0,69 (0,10)	0,82 (0,13)	1,88 (0,22)	2,60 (0,33)
8	M	1,16 (0,08)	-1,03 (0,09)	0,66 (0,07)	2,34 (0,16)	3,23 (0,22)	4,73 (0,40)
	V	0,68 (0,13)	-1,50 (0,28)	1,12 (0,31)	3,29 (0,38)	4,76 (0,99)	6,17 (1,35)

Nota. M: mujer, V: varón; a : parámetro de inclinación calculado con el MRG; b_1 , b_2 , b_3 , b_4 y b_5 : parámetros de umbral calculados con el MRG; EE: error estándar.

Siguiendo el procedimiento recomendado por Woods (2009), se estudió en un segundo paso al ítem 8 como candidato a tener DIF, seleccionando otro ítem que funcionara como anclaje entre los dos grupos. El reactivo elegido para este fin fue el ítem 7, por tener menor χ^2 total. Los resultados corroboraron la presencia de DIF no uniforme en el ítem 8, $\chi^2_a(1, N = 1383) = 5,9, p = 0,015$. Por lo tanto, fue eliminado para los análisis del presente trabajo y de la escala.

Los análisis de invarianza por género se exponen en la Tabla 3. Si bien globalmente se puede aceptar la invarianza, existe un leve desajuste en el límite superior del RMSEA para los varones (M0 Varones) y en el modelo 2 (M2), lo cual puede ser compatible con el hallazgo realizado previamente de un ítem con DIF.

Tabla 3
Análisis de Invarianza por Género

	χ^2	gl	p	CFI	RMSEA [90% IC]
M0 Mujeres	89,123	20	< 0,001	0,960	0,056 [0,044, 0,068]
M0 Varones	44,930	20	< 0,001	0,946	0,067 [0,040, 0,094]
M1	169,802	20	< 0,001	0,967	0,074 [0,064, 0,084]
M2	296,334	26	< 0,001	0,941	0,087 [0,078, 0,096]

Nota. N = 1383. M0 Mujeres y M0 Varones: modelo factorial confirmatorio unidimensional aplicado solo a mujeres y solo a varones, respectivamente; M1: modelo confirmatorio aplicado a la totalidad de la muestra con restricción de igualdad de cargas factoriales entre mujeres y varones; M2: modelo confirmatorio aplicado a la totalidad de la muestra con restricción de igualdad de cargas factoriales y errores de varianzas entre mujeres y varones.

Estimación de Parámetros del MRG

El ajuste de los datos de la EA al MRG resultó adecuado, $M_2 = 700,22, gl = 518, p < 0,001, RMSEA = 0,02$. Asimismo, para todos los ítems el índice S- χ^2 se asoció con un $p > 0,01$, señalando un adecuado ajuste de los reactivos al modelo.

Como ya se ha mencionado, para cada ítem se calculó un parámetro de inclinación a y cinco parámetros de umbral b_m . Estos resultados se exponen en la Tabla 4. El parámetro b_1 indica el valor de umbral que separa las categorías de respuesta *Nunca* de *Casi nunca*. Es el mínimo nivel de ayuda necesario para tener una probabilidad mayor a 0,50 de elegir la categoría *Casi nunca* o una superior. El parámetro b_2 indica el valor de umbral que separa las categorías de respuesta *Casi nunca* de *A veces*. Es el mínimo nivel de ayuda necesario para tener una probabilidad de 0,50 de elegir la categoría *A veces* o una superior. De idéntica manera se debe interpretar el resto de los b_m . Estos parámetros permiten calcular las curvas características operantes, paso intermedio para la obtención de las CCCRI. Para calcular cada CCCRI se sustraen dos probabilidades acumuladas a derecha: la de una categoría dada y de la siguiente a esta. Estas probabilidades acumuladas a derecha refieren a que la respuesta de un participante al ítem esté en o por encima de un umbral b_m en función del nivel del rasgo latente θ .

Los parámetros de inclinación a tuvieron un valor promedio estimado de 1,37 ($EE = 0,53$), que representa una capacidad alta de las categorías de respuesta para distinguir entre los diferentes niveles de rasgo según el criterio de Reise y Waller (1990). El ítem 3 (*Me comprometo con causas nobles*) tuvo el parámetro de inclinación a más elevado con un valor de 2,17 ($EE = 0,13$). Los ítems 4 (*Dedico un tiempo importante de mi vida a mejorar el mundo*) y 7 (*Me ocupo del bienestar de cualquier individuo, grupo o comunidad*) también alcanzaron valores considerables en este parámetro, de 1,79 ($EE = 0,10$) y 1,73 ($EE = 0,10$), respectivamente.

Tabla 4
Parámetros Estimados Mediante MRG para los Ítems de la EA

Ítem	a (EE)	b_1 (EE)	b_2 (EE)	b_3 (EE)	b_4 (EE)	b_5 (EE)	b_{prom}
1	0,69 (0,06)	-3,24 (0,28)	-1,17 (0,13)	1,51 (0,15)	3,62 (0,32)	5,70 (0,53)	1,28
2	1,02 (0,07)	-2,37 (0,16)	-0,94 (0,08)	0,42 (0,07)	1,03 (0,08)	1,64 (0,11)	-0,04
3	2,17 (0,13)	-2,25 (0,10)	-1,17 (0,06)	0,18 (0,04)	0,93 (0,05)	1,70 (0,08)	-0,12
4	1,79 (0,10)	-1,70 (0,08)	-0,25 (0,04)	0,94 (0,06)	1,67 (0,08)	2,45 (0,12)	0,62
5	1,22 (0,08)	-2,50 (0,15)	-1,39 (0,09)	-0,30 (0,06)	0,50 (0,06)	1,27 (0,08)	-0,48
6	0,98 (0,07)	-2,41 (0,16)	-0,49 (0,07)	1,38 (0,10)	2,44 (0,16)	3,76 (0,26)	0,94
7	1,73 (0,10)	-1,87 (0,09)	-0,58 (0,05)	0,86 (0,05)	1,85 (0,09)	2,69 (0,14)	0,59

Nota. a : parámetro de inclinación calculado con el MRG; b_1 , b_2 , b_3 , b_4 y b_5 : parámetros de umbral calculados con el MRG; b_{prom} : promedio de los parámetros de umbral b ; EE: error estándar.

A modo de ejemplo, se exhiben las CCCRI de los ítems 3 en la Figura 1 y 7 en la Figura 2.

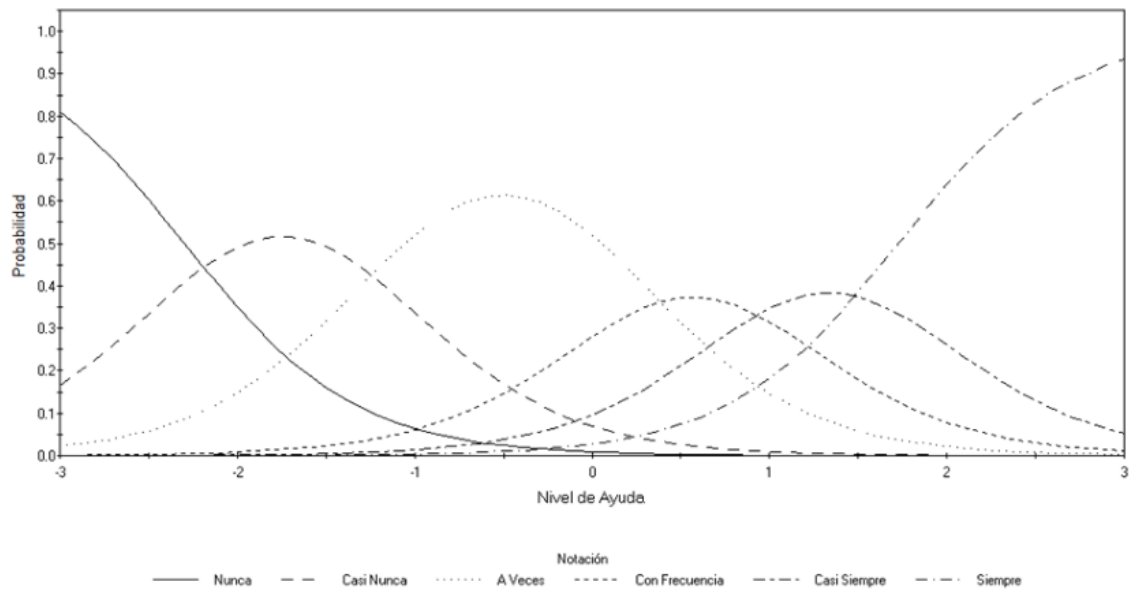


Figura 1. Curvas características de las categorías de respuesta del ítem 3.

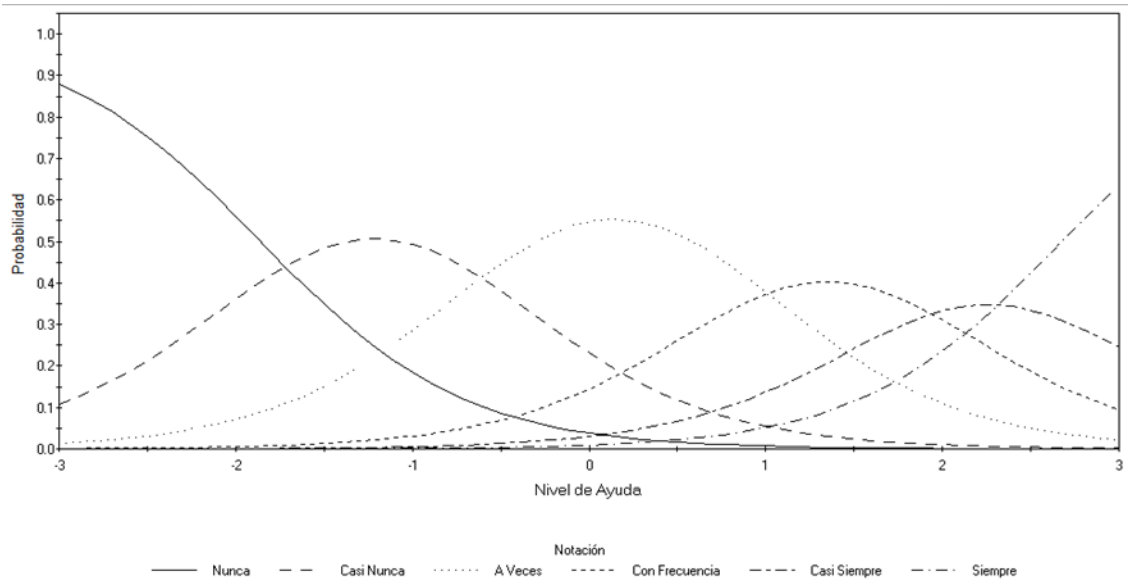


Figura 2. Curvas características de las categorías de respuesta del ítem 7.

Funciones de Información de los Ítems y del Test

En la Tabla 5 se exhiben los valores que alcanzan las FIIs y FIT para algunos niveles de ayuda. Como se puede observar, el test y la mayoría de los ítems son más informativos para niveles medio-bajos hasta altos del rasgo. El error estándar en la medición es mayor en niveles extremos de ayuda, donde a su vez disminuye el nivel de información que el test proporciona. Se puede observar que el ítem 3 es el que proporciona el mayor nivel de información, seguido por los ítems 4 y 7. Por el contrario, el ítem 1 contribuye pobremente y de forma achatada a lo largo de todo el espectro del rasgo latente a la información que brinda el test.

Tabla 5

Funciones de Información de los Ítems y del Test Obtenidas Mediante MRG

θ Ítem	$\theta = -2,4$	$\theta = -1,6$	$\theta = -0,8$	$\theta = 0$	$\theta = 0,8$	$\theta = 1,6$	$\theta = 2,4$
1	0,14	0,14	0,14	0,14	0,14	0,14	0,15
2	0,29	0,32	0,33	0,33	0,33	0,31	0,24
3	1,19	1,30	1,22	1,30	1,43	1,35	0,69
4	0,56	0,86	0,86	0,93	0,96	1,00	0,92
5	0,43	0,46	0,47	0,48	0,46	0,39	0,24
6	0,27	0,28	0,29	0,29	0,29	0,30	0,30
7	0,62	0,83	0,85	0,81	0,87	0,91	0,90
Test	4,50	5,21	5,17	5,28	5,50	5,40	4,43
EE	0,47	0,44	0,44	0,44	0,43	0,43	0,47

Nota. Se destacan en negrita los valores máximos de la función de información. MRG: modelo de respuesta graduada; θ : nivel de ayuda; Test: nivel de información de la EA para cada nivel de ayuda; EE: error estándar.

La Figura 3 muestra que el gráfico de la FIT tuvo forma de meseta. Es así que el test proporciona estimaciones de ayuda relativamente precisas en un rango amplio de rasgo de -2 a 2, mientras que el error en la medición crece hacia los niveles de ayuda menores a -2 y mayores a 2.

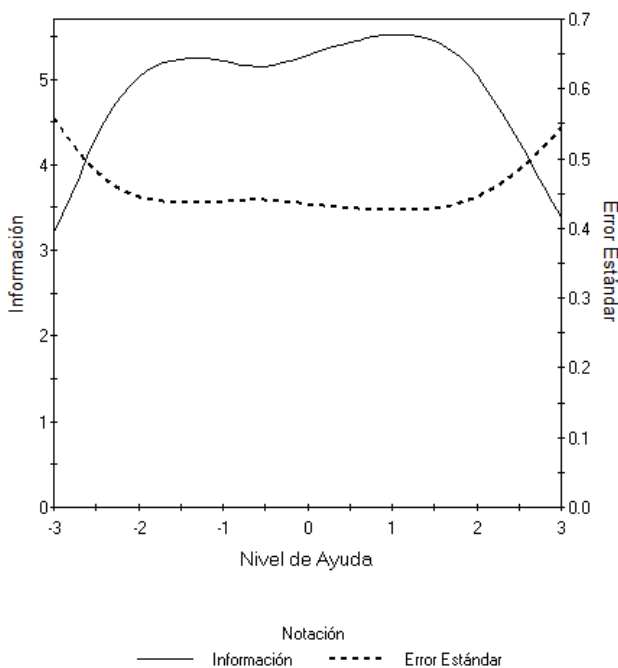


Figura 3. Función de información del test.

Confiabilidad

El índice de confiabilidad marginal, apto para medir la confiabilidad global en el marco de la TRI, resultó de 0,81, es decir, muy adecuado.

Discusión

La modelización de la EA con el MRG evidenció que el instrumento resulta más informativo para niveles medio-altos del rasgo, en comparación con niveles bajos. Es decir, proporciona menos información en participantes con bajos niveles de ayuda. Algunos de los ítems existentes suponen un nivel muy elevado de ayuda (eg., *Si tuviera un único plato de comida, lo compartiría con un desconocido*), compatible con la noción de altruismo. Conviene recordar que la existencia misma del altruismo es un tópico controvertido, con defensores y detractores (Batson & Powell, 2003). Dados los elevados índices de ajuste de la estructura a la unidimensionalidad, se puede plantear una hipótesis a corroborar en posteriores estudios: considerar que el altruismo es la forma más alta de la dimensión ayuda. Su existencia no es aislada, sino que formaría parte de un continuo. En este sentido, se ha hallado una asociación entre actitudes altruistas y comportamientos de ayuda (Kahana, Bhatta, Lovegreen, Kahana & Midlarsky, 2013). Además, este postulado sería coherente con la hipótesis social heurística, una teoría que postula que las estrategias beneficiosas en la vida cotidiana se manifiestan también en escenarios menos frecuentes y que requieren una respuesta rápida, como son las situaciones de emergencia (Rand & Epstein, 2014). La EA puede ser útil, por ejemplo, para la selección de personal en empleos que requieran niveles elevados de ayuda, la evaluación de programas de incremento de estas acciones y el estudio en el contexto local de la relación entre el rasgo y el bienestar personal, la calidad de vida y facetas de la personalidad, entre otros constructos.

El análisis con el MRG muestra una precisión aceptable en un rango muy amplio de niveles del rasgo, donde la mayoría de las personas podría encuadrarse. El error de medición en el MRG solo es muy elevado en personas con niveles extremos de ayuda. Si se incorporaran nuevos ítems a esta escala, sería aconsejable que los mismos requieran un nivel bajo de ayuda para ser respondidos en forma favorable.

Los ítems de la EA pueden discriminar —según el parámetro α — en forma alta según MRG. Esta capacidad discriminatoria, así como el nivel de información alcanzado fue mucho mayor para los ítems 3, 4

y 7, es decir, que comprometerse con causas nobles, dedicar tiempo a que el mundo sea mejor y ocuparse del bienestar de otros fueron contenidos efectivos en detectar pequeñas variaciones en el nivel de ayuda. En el caso del ítem 3, sus categorías permiten discriminar entre participantes con niveles desde muy bajos hasta altos de ayuda y los ítems 4 y 7, desde niveles bajos hasta niveles muy altos de ayuda, aunque para los tres el máximo de la información brindada se encuentra en los niveles altos del rasgo. Estos reactivos denotan acciones concretas, sostenidas y solidarias en relación con los otros: compromiso, responsabilidad y ocuparse y preocuparse por determinada causa. Estos reactivos parecen ser el núcleo central de la escala.

En cuanto a los reactivos 2, 5 y 6, si bien aportan un menor nivel de información que el anterior grupo, mostraron adecuadas propiedades psicométricas, siendo el de menor calidad el ítem 6. Este presenta un parámetro b_5 mayor a 3, lo cual podría deberse a que es un ítem que presenta un contenido bastante extremo y la opción *Siempre* fue poco frecuente (3,60%). Además, al ser muchas opciones de respuesta, parece lógico que en este punto se capte a las personas con niveles muy superiores del rasgo ayuda. Este conjunto de ítems, además, presenta una menor capacidad de discriminación entre los participantes con diferentes niveles de ayuda. Estos ítems tienen en común el referirse a beneficiarios de un exogrupo y dos de ellos implican dar comida o dinero, en contraposición a los ítems del anterior grupo que implicaban más un dar del propio tiempo personal. Será necesario investigar más profundamente estos aspectos.

Con respecto al ítem 1, si bien se ajustó al modelo, mostró pobres indicadores que hacen dudar de su calidad psicométrica: parámetros de umbral fuera de rango y un nivel de información bajo y plano. Si se analiza su contenido, es el único ítem que se refiere al ámbito estudiantil, es decir, para la respuesta de la muestra estudiada el contenido de este ítem, específico del ámbito estudiantil, no permitió distinguir entre participantes con distintos niveles de ayuda. En base a los resultados obtenidos en el presente estudio, este reactivo será reformulado y reevaluado en posteriores versiones de la escala.

Haber obtenido evidencia de DIF en un ítem extremo de ayuda podría ser resultado de la vigencia de los estereotipos acerca de cómo es y debe ser un varón y una mujer (Auné et al., 2015). Dar todo y quedarse sin nada puede responder a un ideal altruista más compatible con lo inculcado a la mujer y menos con el pragmatismo resultante de los estereotipos masculinos. Esto podría ser objeto de indagación en posteriores estudios. Es de destacar la importancia de analizar tanto el DIF uniforme como el no uniforme, lo cual se puede lograr fácilmente con los métodos basados en la TRI. Así, ningún ítem con DIF quedará sin detectar. Por otra parte, al estar la muestra circunscripta a estudiantes universitarios, los resultados obtenidos deberían ser corroborados en estudiantes de otros niveles educativos. Una muestra más heterogénea permitirá potenciar notablemente las posibilidades de generalización de los resultados; además, se podrán realizar comparaciones de interés.

Es necesario destacar que, dado que la EA se trata de un autorreporte, la misma presenta límites, dados principalmente por la distorsión en la percepción de la propia conducta de ayuda y en la deseabilidad social, así como sesgos de respuesta. Es una meta para futuros estudios correlacionar el puntaje en la EA con la realización efectiva de conductas de ayuda, así como con un test de deseabilidad social. En este sentido, es factible utilizar la versión argentina del Inventario Balanceado de Respuesta Deseable (Mikulic, Crespi & Caballero, 2016). En cuanto a las limitaciones de la presente investigación en forma global, el comportamiento del Modelo de Respuesta Graduada puede modificarse por la escala Likert que se propone y el número de opciones de respuesta. En futuros estudios se pueden comparar los indicadores psicométricos de los ítems según la cantidad de opciones de respuesta, para así determinar su número óptimo. Otra limitación del presente estudio es que la muestra estuvo compuesta exclusivamente por estudiantes universitarios. En el futuro se podría recolectar una muestra de estudiantes de otros niveles educativos y otra de población general. Así, se podrán comparar los grupos en sus resultados, así como generalizar las evidencias de calidad psicométrica de la EA.

Referencias

- Abad, F. J., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid, España: Síntesis.
- Aknin, L. B., Broesch, T., Hamlin, J. K. & Van de Vondervoort, J. W. (2015). Prosocial behavior leads to happiness in a small-scale rural society. *Journal of Experimental Psychology: General*, 144, 788-795. <https://doi.org/10.1037/xge0000082>
- Auné, S. E. (2018). *Construcción de un test de comportamiento prosocial y su modelización con la teoría de la respuesta al ítem* (Tesis de Doctorado, Facultad de Psicología, Universidad de Buenos Aires, Argentina). Extraído de https://ri.conicet.gov.ar/bitstream/handle/11336/83189/CONICET_Digital_Nro.2da682ec-dff9-4198-814c-5cc6d3738198_A.pdf?sequence=2&isAllowed=y

- Auné, S., Abal, F. & Attorresi, H. (2014). Versión argentina de la Escala de Habilidades Prosociales de Morales Rodríguez y Suárez Pérez (2011). *Praxis: Revista de Psicología*, 16(26), 31-48. Extraído de <http://www.praxis.udp.cl/index.php/praxis/article/view/24/14>
- Auné, S., Abal, F. & Attorresi, H. (2015). Conducta prosocial y estereotipos de género. *Praxis: Revista de Psicología*, 17(27), 7-19. Extraído de <http://www.praxis.udp.cl/index.php/praxis/article/view/15/9>
- Auné, S. E., Blum, D., Abal, F. J. P., Lozzia, G. S. & Attorresi, H. F. (2014). La conducta prosocial: estado actual de la investigación. *Perspectivas en Psicología*, 11(2), 21-33. Extraído de <http://www.seadpsi.com.ar/revistas/index.php/pep/article/view/153/pdf>
- Batson, C. D. & Powell, A. A. (2003). Altruism and prosocial behavior. En I. B. Weiner, T. Millon & M. J. Lerner (Eds.), *Handbook of psychology. Volume 5: Personality and social psychology* (pp. 463-484). New York, NY: John Wiley & Sons.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z. & Franic, S. (2009). The end of construct validity. En R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 135-170). Charlotte, NC: Information Age Publishing.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2ª ed.). New York, NY: Guilford Press.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.
- Cai, L. (2012). *flexMIRT: Flexible multilevel item factor analysis and test scoring* [Software computacional]. Seattle, WA: Vector Psychometric Group.
- Cai, L., Maydeu-Olivares, A., Coffman, D. L. & Thissen, D. (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2^p tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173-194. <https://doi.org/10.1348/000711005X66419>
- Cai, L., Thissen, D. & du Toit, S. H. C. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Caprara, G. V., Alessandri, G. & Eisenberg, N. (2012). Prosociality: The contribution of traits, values, and self-efficacy beliefs. *Journal of Personality and Social Psychology*, 102, 1289-1303. <https://doi.org/10.1037/a0025626>
- Caprara, G. V. & Pastorelli, C. (1993). Early emotional instability, prosocial behaviour, and aggression: Some methodological aspects. *European Journal of Personality*, 7, 19-36. <https://doi.org/10.1002/per.2410070103>
- Caprara, G. V., Steca, P., Zelli, A. & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European Journal of Psychological Assessment*, 21, 77-89. <https://doi.org/10.1027/1015-5759.21.2.77>
- Carlo, G. & Randall, B. A. (2002). The development of a measure of prosocial behaviors for late adolescents. *Journal of Youth and Adolescence*, 31, 31-44. <https://doi.org/10.1023/A:1014033032440>
- Chen, W. -H. & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289. <https://doi.org/10.3102/10769986022003265>
- Chou, S. Y. & Stauffer, J. M. (2015). A theoretical classification of helping behavior and helping motives. *Personnel Review*, 45, 871-888. <https://doi.org/10.1108/PR-03-2015-0076>
- Cupani, M. (2012). Análisis de ecuaciones estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*, 2, 186-199. Extraído de <https://revistas.unc.edu.ar/index.php/tesis/article/view/2884>
- Drummond, J. D. K., Hammond, S. I., Satlof-Bedrick, E., Waugh, W. E. & Brownell, C. A. (2017). Helping the one you hurt: Toddlers' rudimentary guilt, shame, and prosocial behavior after harming another. *Child Development*, 88, 1382-1397. <https://doi.org/10.1111/cdev.12653>
- Dunfield, K. A. (2014). A construct divided: Prosocial behavior as helping, sharing, and comforting subtypes. *Frontiers in Psychology*, 5, artículo 958. <https://doi.org/10.3389/fpsyg.2014.00958>
- Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist*, 64, 644-658. <https://doi.org/10.1037/0003-066X.64.8.644>
- Eisenberg, N., Hofer, C., Sulik, M. J. & Liew, J. (2014). The development of prosocial moral reasoning and a prosocial orientation in young adulthood: Concurrent and longitudinal correlates. *Developmental Psychology*, 50, 58-70. <https://doi.org/10.1037/a0032990>
- González Gutiérrez, J. L., Peñacoba Puente, C., Moreno Rodríguez, R., López López, A. & Velasco Furlong, L. (2006). Nursing Motives for Helping Scale (N-MHS): Reliability and validity. *The Spanish Journal of Psychology*, 9, 103-112. <https://doi.org/10.1017/S1138741600006028>
- González Portal, M. D. (2000). *Conducta prosocial: evaluación e intervención* (3ª ed.). Madrid, España: Morata.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360. <https://doi.org/10.1111/j.1745-3984.1984.tb01039.x>
- Harpole, J. K., Levinson, C. A., Woods, C. M., Rodebaugh, T. L., Weeks, J. W., Brown, P. J. ... Liebowitz, M. (2015). Assessing the straightforwardly-worded Brief Fear of Negative Evaluation Scale for differential item functioning across gender and ethnicity. *Journal of Psychopathology and Behavioral Assessment*, 37, 306-317. <https://doi.org/10.1007/s10862-014-9455-9>
- Inderbitzen, H. M. & Foster, S. L. (1992). The Teenage Inventory of Social Skills: Development, reliability, and validity. *Psychological Assessment*, 4, 451-459. <https://doi.org/10.1037/1040-3590.4.4.451>
- Kahana, E., Bhatta, T., Lovegreen, L. D., Kahana, B. & Midlarsky, E. (2013). Altruism, helping, and volunteering: Pathways to well-being in late life. *Journal of Aging and Health*, 25, 159-187. <https://doi.org/10.1177/0898264312469665>
- Klein, N. (2017). Prosocial behavior increases perceptions of meaning in life. *The Journal of Positive Psychology*, 12, 354-361. <https://doi.org/10.1080/17439760.2016.1209541>
- Langer, M. M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Disertación de Doctorado, University of North Carolina at Chapel Hill, NC, Estados Unidos). Extraído de: <https://pdfs.semanticscholar.org/204e/a4e24df2284f4c6833b73ec18a01964fe2ca.pdf>
- Lord, F. M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1, 95-100. <https://doi.org/10.1177/014662167700100115>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Maydeu-Olivares, A. & Joe, H. (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables: A Unified Framework. *Journal of the American Statistical Association*, 100, 1009-1020. <https://doi.org/10.1198/016214504000002069>
- Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71, artículo 713. <https://doi.org/10.1007/s11336-005-1295-9>
- McFarland, S., Brown, D. & Webb, M. (2013). Identification with all humanity as a moral concept and psychological construct. *Current Directions in Psychological Science*, 22, 194-198. <https://doi.org/10.1177/0963721412471346>

- Mikulic, I. M., Crespi, M. & Caballero, R. (2016). Estudio psicométrico del Inventario Balanceado de Respuesta Deseable. *Anuario de Psicología*, 46, 58-66. <https://doi.org/10.1016/j.anpsic.2016.07.002>
- Morales Rodríguez, M. & Suárez Pérez, C. D. (2011, Noviembre). *Construcción y validación de una escala para evaluar habilidades prosociales en adolescentes*. Ponencia presentada en el XI Congreso Nacional de Investigación Educativa, Consejo Mexicano de Investigación Educativa, México DF, México. Extraído de http://www.comie.org.mx/congreso/memoriaelectronica/v11/docs/area_01/0777.pdf
- Muñoz García, J. A. & Amón Uribe, I. (2013). Técnicas para detección de outliers multivariantes. *Revista en Telecomunicaciones e Informática*, 3(5), 11-25. Extraído de <https://revistas.upb.edu.co/index.php/telecomunicaciones/article/view/3308/2909>
- Muthén, L. K. & Muthén, B. O. (2010). *Mplus user's guide* (6ª ed.) Los Angeles, CA: Autores.
- Nickell, G. S. (1998, Agosto). *The Helping Attitude Scale*. Ponencia presentada en la 106th Annual Convention of the American Psychological Association, San Francisco, CA, Estados Unidos. Extraído de <http://web.mnstate.edu/nickell/APA%201998%20paper.pdf>
- Nielson, M. G., Padilla-Walker, L. & Holmes, E. K. (2017). How do men and women help? Validation of a multidimensional measure of prosocial behavior. *Journal of Adolescence*, 56, 91-106. <https://doi.org/10.1016/j.adolescence.2017.02.006>
- 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>
- Osterlind, S. J. & Everson, H. T. (2009). *Differential item functioning* (2ª ed.). Los Angeles, CA: SAGE. <https://doi.org/10.4135/9781412993913>
- Penner, L. A., Dovidio, J. F., Piliavin, J. A. & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, 56, 365-392. <https://doi.org/10.1146/annurev.psych.56.091103.070141>
- Penner, L. A., Fritzsche, B. A., Craiger, J. P. & Freifeld, T. S. (1995). Measuring the prosocial personality. En J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 147-163). Hillsdale, NJ: Lawrence Erlbaum.
- Rand, D. G. & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS ONE*, 9(10), artículo e109687. <https://doi.org/10.1371/journal.pone.0109687>
- Reise, S. P. & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, 14, 45-58. <https://doi.org/10.1177/014662169001400105>
- Romer, D., Gruder, C. L. & Lizzadro, T. (1986). A person-situation approach to altruistic behavior. *Journal of Personality and Social Psychology*, 51, 1001-1012. <https://doi.org/10.1037/0022-3514.51.5.1001>
- Ruci, L., van Allen, Z. M. & Zelenski, J. M. (2018). Pro-social personality traits, helping behavior, and ego-depletion: Is helping really easier for the dispositionally pro-social? *Personality and Individual Differences*, 120, 32-39. <https://doi.org/10.1016/j.paid.2017.08.013>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(1 Supplement), 1-97. <https://doi.org/10.1007/BF03372160>
- Samejima, F. (2016). Graded response models. En W. J. van der Linden (Ed.), *Handbook of item response theory, Volume One: Models* (pp. 95-108). Boca Raton, FL: CRC Press.
- Sánchez-Queija, I., Oliva Delgado, A. & Parra Jiménez, A. (2006). Empatía y conducta prosocial durante la adolescencia. *Revista de Psicología Social*, 21, 259-272. <https://doi.org/10.1174/021347406778538230>
- Spitzmuller, M. & Van Dyne, L. (2013). Proactive and reactive helping: Contrasting the positive consequences of different forms of helping. *Journal of Organizational Behavior*, 34, 560-580. <https://doi.org/10.1002/job.1848>
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics* (4ª ed.). Boston, MA: Allyn & Bacon.
- Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system—IRT software. En D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing* (pp. 1-32). Reston, VA: Graduate Management Admission Council. Extraído de <http://www.iacat.org/sites/default/files/biblio/cat09thissen.pdf>
- Thissen, D., Chen, W. -H. & Bock, R. (2003). *MULTILOG 7* [Software computacional]. Chicago, IL: Scientific Software International.
- Toland, M. D. (2014). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, 34, 120-151. <https://doi.org/10.1177/0272431613511332>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-482. <https://doi.org/10.1090/s0002-9947-1943-0012401-3>
- Warneken, F. & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, 100, 455-471. <https://doi.org/10.1348/000712608x379061>
- Woods, C. M. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*, 33, 42-57. <https://doi.org/10.1177/0146621607314044>
- Woods, C. M., Cai, L. & Wang, M. (2013). The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, 73, 532-547. <https://doi.org/10.1177/0013164412464875>
- Yavuzer, H., İşmen-Gazioğlu, E., Yildiz, A., Demir, İ., Meşeci, F., Kiliçaslan, A. & Serteln, Ç. (2006). The Teacher Altruism Scale: Development, validity and reliability. *Educational Sciences: Theory & Practice*, 6, 964-972. <https://doi.org/10.1037/t68695-000>

Fecha de recepción: Agosto de 2018.

Fecha de aceptación: Octubre de 2019.