

Evidencias de Validez de la Escala de Conducta Prosocial en una Muestra Colombiana

Evidence of Validity of the Prosocial Behavior Scale in a Colombian Sample

Javier Villalba¹, Juan Hernández¹, Moisés Betancort¹ y Bertha Avendaño²

¹ Escuela de Doctorado y Estudios de Posgrado de la Universidad de La Laguna

² Doctorado en Psicología de la Universidad Católica de Colombia

El objetivo de este estudio fue recolectar evidencias de validez basadas en el contenido de los ítems y en la estructura interna de la Escala de Conducta Prosocial (Auné & Attorresi, 2017), para obtener una adaptación preliminar para la población colombiana. Respecto al contenido de los ítems, los expertos aprobaron nueve de los quince ítems, se realizaron ajustes a cinco ítems y se decidió eliminar uno. El acuerdo global entre los tres expertos más exigentes fue del 73,33% con un Kappa de Fleiss $k=0,585$ y un $p=0,000$. La dimensión Ayuda tuvo fiabilidades $\alpha=0,85$ y $\omega=0,83$ y en la dimensión Confortar al otro se obtuvo un $\alpha=0,77$ y $\omega=0,58$. Se confirmó la estructura factorial del modelo bidimensional integrado con las dimensiones Ayuda y Confortar al otro, con índices de bondad de ajuste $NFI=0,958$, $NNFI=0,979$, $CFI=0,982$, $TLI=0,979$ y $RMSEA=0,046$. Esta estructura fue invariante con restricción configural, métrica, escalar y residual entre hombres y mujeres. La calibración de los ítems con el modelo de Rasch permitió identificar que las personas que invierten su tiempo libre en ayudar, sumado a la disposición de brindar apoyo físico a otros, son las que presentan los niveles más altos de conducta prosocial. Estas evidencias avalan el uso del cuestionario en Colombia para fines investigativos.

Palabras clave: Validez, Estructura Interna y Conducta Prosocial

The aim of this study was to collect validity evidence based on the content of the items and the internal structure of the Prosocial Behavior Scale (Auné & Attorresi, 2017), in order to obtain a preliminary adaptation for the Colombian population. Regarding the content of the items, the experts approved nine of the fifteen items, adjustments were made to five items and it was decided to eliminate one item. The overall agreement among the three most demanding experts was 73.33% with a Fleiss Kappa $k=0.585$ and $p=0.000$. The Helping dimension had reliabilities $\alpha=0.85$ and $\omega=0.83$ and in the Comforting the other dimension an $\alpha=0.77$ and $\omega=0.58$. The factorial structure of the two-dimensional model integrated with the dimensions Helping and Comforting the other was confirmed, with goodness-of-fit indices $NFI=0.958$, $NNFI=0.979$, $CFI=0.982$, $TLI=0.979$ and $RMSEA=0.046$. This structure was invariant with configural, metric, scalar and residual restriction between men and women. The calibration of the items with the Rasch model allowed to identify that people who invest their free time in helping, added to the willingness to provide physical support to others, are those who present the highest levels of prosocial behavior. These evidences support the use of the questionnaire in Colombia for research purposes.

Keywords: Validity, Internal Structure and Prosocial Behavior

En los últimos 15 años se ha incrementado el interés por estudiar la conducta prosocial, porque está relacionada con aspectos positivos del ser humano, específicamente el bienestar social y personal, lo que contribuye a la conformación de lazos solidarios y modera las conductas agresivas de las personas (Auné et al., 2014). También se ha encontrado que hay una asociación entre la conducta prosocial y la satisfacción con la vida, lo que permite predecir el bienestar subjetivo y especialmente en adolescentes (Gillham et al., 2011;

Javier Andrés Villalba Garzón  <https://orcid.org/0000-0002-5358-2353>

Juan A. Hernández-Cabrera  <https://orcid.org/0000-0001-6731-5670>

Moisés Betancort  <https://orcid.org/0000-0002-7790-8557>

Bertha Lucía Avendaño  <https://orcid.org/0000-0002-8136-5380>

El artículo es parte de la tesis para Optar al Grado de Doctor en Psicología de la Universidad de La Laguna. No existe ningún conflicto de intereses que revelar.

La correspondencia relativa a este artículo debe ser dirigida a Javier Andrés Villalba Garzón, Calle 25 #68B-47 Interior 3 apartamento 702, Bogotá D. C., Colombia. Email: alu0101298733@ull.edu.es

Ripoll-Núñez et al., 2020). Inicialmente, la conducta prosocial fue definida como un antónimo de la conducta antisocial (Wispé, 1972), pero con el tiempo este comportamiento ha sido redefinido a tal punto que, a la fecha, ha sido posible identificar al menos seis tipologías diferentes (Auné et al., 2014). En términos generales, la conducta prosocial se entiende como el intento de satisfacer la necesidad de apoyo físico y emocional de otra persona, diferente al altruismo que se define como el resultado de anteponer la necesidad de los demás a la propia (Auné et al., 2014). Se identifica un mayor número de investigaciones relacionadas con el estudio de la conducta prosocial en la infancia en comparación con la edad adulta, lo cual puede explicarse porque desde edades muy tempranas se empiezan a observar una gran variedad de comportamientos sociales positivos, que están orientados a interactuar de manera colaborativa con los demás (Ulber & Tomasello, 2020).

El desarrollo de instrumentos psicométricos ha sido fundamental para avanzar en el estudio de la conducta prosocial, porque permite obtener medidas altamente confiables y válidas, propiedades que se logran con las condiciones mínimas requeridas, que están establecidas en los estándares de calidad para pruebas psicológicas y educativas publicados por la American Educational Research Association, la American Psychological Association y el National Council on Measurement in Education (AERA, APA & NCME, 2014). Las cinco fuentes de evidencia de validez de estos estándares se basan en: a) la relación del contenido de los ítems con el marco de referencia del atributo medido; b) los procesos de respuesta involucrados; c) la estructura interna de la escala; d) la relación de los puntajes con otras variables y e) las consecuencias derivadas del uso de la prueba.

Se han identificado 12 pruebas de auto aplicación que permiten medir la conducta prosocial de forma global o como componente en una subescala. Estas pruebas son: a) la Prosocial Behavior Scale (PB) de Caprara y Pastorelli (1993); b) la Prosocialness Scale for Adults (PSA) de Caprara et al. (2005); c) la Prosocial Tendencies Measure (PTM) de Carlo y Randall (2002); d) el Teenage Inventory of Social Skills (TISS) de Inderbitzen y Foster (1992); e) la Escala de Habilidades Prosociales para Adolescentes (EHP-A) de Morales Rodríguez y Suárez Pérez (2011); f) el Cuestionario de Conducta Prosocial de Sánchez-Queija et al. (2006); g) la Escala para la Medición de la Conducta Prosocial-Antisocial en el Ámbito Vital y en el Tráfico de López de Cózar et al. (2008); h) la Batería de Socialización para Adolescentes (BAS-3) de Silva y Martorell (1987); i) la Prosocial Personality Battery (PSB) de Penner et al. (1995); j) el Empathy Questionnaire for Children and Adolescents (EmQue-CA) de Overgaauw et al. (2017); k) la Encuesta de Comportamientos Agresivos y Prosociales (COPRAG) de Agudelo et al. (2000) y l) la Escala de Conducta Prosocial (ECP) de Auné & Attorresi (2017).

Desarrollo de la Escala de Conducta Prosocial (ECP)

De las anteriores pruebas, la ECP de Auné y Attorresi (2017) es la única prueba diseñada para medir la conducta prosocial en adultos de un contexto latinoamericano y de la cual se dispone de una amplia variedad de evidencias, que permiten identificar su desarrollo y calidad psicométrica. Por otra parte, la ECP dispone de un mayor número de estudios realizados con métodos psicométricos propios de la Teoría de Respuestas al Ítem (TRI), que son más robustos en contraste con los métodos de la Teoría Clásica de los Tests (TCT).

Para elaborar este cuestionario se elaboraron 114 ítems, después de recolectar evidencias relacionadas con el contenido de los ítems con una revisión de un grupo de expertos, se seleccionaron 64 ítems que tuvieron una segunda revisión y se sometieron a una prueba piloto con 56 participantes que eran estudiantes de varias carreras profesionales de la Universidad de Buenos Aires (Auné & Abal, 2016; Auné, 2018). Con los resultados de la prueba piloto se obtuvo la primera versión de la ECP conformada por 29 ítems, con la que se recolectaron evidencias de validez de estructura interna con un Análisis Factorial Exploratorio (AFE), un Análisis Factorial Confirmatorio (AFC) y un análisis de detección de Funcionamiento Diferencial de los Ítems (DIF). Previo a la realización del AFE, se seleccionó aleatoriamente 761 participantes de una muestra de 1383 y se reportó una medida de adecuación muestral de Kaiser-Meyer-Olkin (*KMO*) de 0,90 y en la prueba de esfericidad de Bartlett un $\chi^2=7441$ con un $p=0,00001$ (Auné, 2018). En el AFE se identificaron dos factores que explicaron el 44,67% de la varianza, el primer factor se denominó “Ayuda” conformado por 9 ítems con una confiabilidad de consistencia interna $\alpha=0,85$ y el segundo factor “Confortar al otro” con 7 ítems con una confiabilidad de consistencia interna $\alpha=0,77$ (Auné, 2018).

Para realizar el AFC, se seleccionó aleatoriamente a 662 de los 1383 participantes y se evaluó la bondad de ajuste de cuatro modelos que tuvieron como base la estructura de dos factores con 16 ítems (Auné, 2018). Se utilizó el estimador mínimo cuadrados ponderados robustos (*WLSMV*), el índice de ajuste comparativo

(*CFI*), el índice de Tucker-Lewis (*TLI*) y el error medio cuadrático de aproximación (*RMSEA*) con un intervalo de confianza (*IC*) del 90%. Auné (2018) reportó índices *CFI*>0,610, *TLI*>0,550 y *RMSEA (IC 90%)*>0,0720; el modelo de dos factores relacionados tuvo la mejor bondad de ajuste con *CFI*=0,921, *TLI*=0,908 y *RMSEA (IC 90%)*=0,070 (0,063-0,077).

Auné (2018) reportó cuatro ítems con DIF entre mujeres y hombres que detectó con el estadístico χ^2 de la prueba Mantel-Haenszel y su *p* valor, el cociente de razones común (MH), la Diferencia de Proporciones Estandarizada (DPE) y el Delta-DIF de Mantel-Haenszel (MH D-DIF). Adicional a los análisis AFE, AFC y DIF, se calibraron los ítems de las dos dimensiones con el modelo logístico de tres parámetros para respuestas graduadas (MRG) y de crédito parcial, con el objetivo de identificar en qué niveles de los rasgos latentes se obtenía una mayor información en la estimación de los parámetros. Auné (2018) encontró que el factor Ayuda fue una dimensión más informativa en los niveles altos y con una precisión aceptable en un rango muy amplio del rasgo que permite ubicar a la mayoría de las personas. Respecto al factor Confortar al otro, se obtuvo una dimensión más informativa en niveles bajos del rasgo y con errores de medición que crecieron sustancialmente en los niveles altos del rasgo.

Con estos resultados se obtuvo la primera versión definitiva de la ECP (Auné, 2018; Auné & Attorresi, 2017) con 15 ítems que conforman dos dimensiones que permiten medir la conducta prosocial en población adulta de argentina, con una consistencia interna satisfactoria ($\alpha=0,84$) que se obtuvo en una muestra de 667 estudiantes de la Universidad de Buenos Aires. En este estudio se confirma la estructura de Ayuda y Confortar al Otro con un AFC con índices *GFI*=0,99 y *RMSR*=0,04. Adicionalmente, Auné y Attorresi (2017) encontraron con un AFE que estas dimensiones explican el 50,61% de la varianza total. Por otro lado, la dimensión Ayuda mostró mejores bondades psicométricas en comparación con la dimensión Confortar al otro y por esto Auné et al. (2020) proponen utilizar los siete primeros ítems de la subescala Ayuda como un cuestionario independiente denominado Escala de Ayuda (EA), a partir del análisis con un Modelo de Respuesta Graduada (MRG) de dos parámetros, con el cual detectaron que el ítem ocho presentó DIF no uniforme entre mujeres y hombres.

A parte de los estudios con población argentina, Canales Reyes (2020) reporta resultados de consistencia interna, evidencias de validez de estructura interna y de la relación con otras variables en población peruana. Con una muestra de 309 participantes entre estudiantes y trabajadores, se obtuvo un $\alpha=0,86$ en la dimensión Ayuda y $\alpha=0,79$ en Confortar al otro. Con el Omega de McDonald los resultados fueron $\omega=0,87$ y $\omega=0,86$ respectivamente. En el AFC se confirmó el modelo no jerárquico oblicuo bidimensional con 11 ítems (2, 3, 4, 5, 6, 7, 10, 11, 12, 13 y 14) y se obtuvo un $\chi^2/df=2,97$, *CFI*=0,94, *TLI*=0,92, *SRMR*=0,05 y *RMSEA*=0,07. Se encontraron correlaciones significativas entre los puntajes de la Escala de Conducta Empática y la dimensión Ayuda $r=0,85$ y la dimensión Confortar al otro $r=0,49$. Estos resultados son similares y consistentes con los obtenidos en los estudios de Auné (2018), Auné y Attorresi (2017) y Auné et al. (2020).

Estudio de la Conducta Prosocial en Colombia

En Colombia, Lengemann Méndez (2019) hizo una revisión de investigaciones entre el 2013 y 2018, en las bases de datos Redalyc, Dialnet, EBSCOhost, APA PsycNET, APA PsycINFO y APA PsycARTICLES, con el objetivo de describir los principales estudios relacionados con la conducta prosocial en Colombia. De este estudio se resalta que hay un mayor porcentaje de investigaciones cuantitativas con un 35%, que en su mayoría fueron realizadas con muestras de población adolescente y la prueba que más se ha utilizado es el TISS, que también está diseñado para adolescentes. La única prueba encontrada que fue diseñada y construida para población colombiana fue la (COPRAG) de Agudelo et al. (2000), que se aplica a menores de edad entre los tres y once años. Una de las conclusiones más interesantes del trabajo de Lengemann Méndez (2019), se relaciona con la afirmación que en Colombia el estudio de la conducta prosocial se encuentra en una etapa inicial y que se requiere madurar, tanto en el desarrollo de estudios a gran escala de naturaleza psicométrica como en la comparación con diferentes variables, que permitan reconocer de forma precisa las conductas prosociales en ese país.

Adaptación de la ECP en Colombia

Aunque en Colombia hay una clara tendencia de estudiar la conducta prosocial en menores de edad, también se considera importante estudiar este comportamiento en adultos por ser un buen predictor de bienestar individual y colectivo (Auné et al., 2014). Al no encontrar escalas adaptadas para adultos

colombianos, se considera como buena opción adaptar la ECP (Auné, 2018; Auné & Attorresi, 2017), por los resultados positivos reportados en los estudios psicométricos realizados con población argentina y peruana.

En la búsqueda realizada no se encontraron reportes empíricos del funcionamiento de la ECP en población colombiana, razón por la cual en este estudio se propuso iniciar con la evaluación de la confiabilidad de este cuestionario, con métodos de consistencia interna y adicionalmente, recolectar evidencias de validez relacionadas con el contenido de los ítems y la estructura interna de esta escala. Se plantea realizar este análisis para tener una adaptación preliminar y facilitar su uso en esta población.

Método

Participantes

Se conformó una muestra por conveniencia de 344 personas adultas alfabetizadas, de las cuales 249 (72,38%) fueron mujeres, 94 (27,33%) hombres y una persona (0,29%) se identificó con otro sexo. La edad de los participantes osciló entre 18 y 77 años ($M=37,59$; $DE=12,54$). Respecto al nivel educativo máximo logrado, el 0,58% reportó una formación de primaria, el 7,85% bachillerato, el 6,10% técnica, el 4,36% tecnológica, el 34,17% profesional, el 22,67% especialista, el 21,80% maestría y el 1,45% doctorado. Al momento de diligenciar la escala el 81,69% tenía un trabajo, el 8,43% era estudiante, el 3,78% estaba en condición de desempleo, el 2,62% se dedicaba a las labores del hogar, el 2,33% se había jubilado y el 1,16% estudiaba y trabajaba al mismo tiempo. El 48,55% de los participantes reportó ser solteros, el 30,23% casados, el 15,70% en unión libre, el 4,07% separados y el 1,45% viudos. Se obtuvieron aplicaciones de personas que nacieron en 12 lugares de Colombia, con la mayor concentración en la ciudad de Bogotá con una participación del 84,51%.

Instrumentos

Escala de Conducta Prosocial (ECP) de Auné & Attorresi (2017). Conformada por las dimensiones de Ayuda y Confortar al otro, con un total de 15 ítems de respuesta tipo Likert de seis opciones (*Nunca, Casi Nunca, A Veces, Con Frecuencia, Casi Siempre y Siempre*). La dimensión Ayuda incluye nueve comportamientos asociados con la solidaridad y que también se plantean como una escala independiente denominada “Escala de Ayuda (EA)” (Auné et al., 2020), mientras que en la dimensión Confortar a otro se describen seis comportamientos que se relacionan con el apoyo emocional que puede darse a otras personas.

Cuestionario de datos sociodemográficos. Se diseñó un instrumento con seis preguntas para recolectar la información sociodemográfica de los participantes. El sexo, estado civil y educación se abordaron con pregunta cerrada; la edad (años), el nombre del título obtenido y la ocupación fue con pregunta abierta.

Procedimiento

Conforme con la Ley 1090 (2006), por la cual se establece el código deontológico de la psicología en Colombia, se obtuvo el permiso de los autores originales para usar y adaptar la escala en población colombiana. Durante todas las fases de la investigación se citaron debidamente a los autores para garantizar la propiedad intelectual. Solo se aplicó el cuestionario a las personas que aceptaron participar voluntariamente y lo expresaron explícitamente en el consentimiento informado. Finalizada la investigación se enviaron los informes con los resultados individuales a los correos electrónicos de los participantes que los solicitaron. La participación fue anónima y se guardó total reserva de los correos electrónicos y datos sociodemográficos registrados, en cumplimiento con la Ley Estatutaria 1581 (2012) por la cual se dictan disposiciones generales para la protección de datos personales en Colombia.

La versión original tenía redactados los ítems en español, por lo que no fue necesario realizar traducción y re-traducción. Cuatro jueces expertos revisaron el cuestionario dos veces, en la primera se revisó la versión original y en la segunda validaron que los ajustes sugeridos en la primera valoración se realizaron correctamente. En ambas valoraciones se revisó cualitativamente la pertinencia del contenido y la formulación de los ítems. Al momento de realizar la evaluación los cuatro tenían más de diez años de experiencia en revisión, adaptación y construcción de pruebas psicométricas. Tres de los expertos tenían un nivel de formación de doctorado y uno de maestría.

Con base en las directrices internacionales sobre pruebas basadas en computadores e Internet de la International Test Commission (ITC, 2005), se subió la ECP a un formulario Google Forms online que podía ser diligenciado desde una computadora o dispositivo móvil con acceso a internet, con cualquier navegador y no hubo restricciones de sistema operativo, software o hardware. Al ser un cuestionario de autorreporte y ejecución típica, no requirió de supervisión directa y los participantes tuvieron la libertad de elegir el dispositivo y la cantidad de tiempo para diligenciar el cuestionario.

La aplicación se realizó durante el confinamiento generado por la pandemia del COVID-19 y se implementaron dos estrategias: a) el cuestionario se distribuyó por correo electrónico y WhatsApp, a partir de un listado de contactos de origen colombiano que disponían los autores y b) se publicó en Facebook y LinkedIn con la condición de limitar el acceso solo a personas colombianas. La recolección de la información tuvo una duración de cuatro meses y medio y se detuvo al obtener un tamaño de muestra superior a 200, para poder realizar los análisis con el modelo de Rasch (Smith et al., 2008) y el AFC (Vallejo, 2013). Se buscó tener una mayor participación de población colombiana no universitaria, para disminuir un probable sesgo de selección y se excluyeron las aplicaciones realizadas por personas que reportaron ser menores de edad, por no pertenecer a la población objetivo.

Plan de Análisis

Las evidencias de validez basadas en el contenido de los ítems (AERA, APA & NCME, 2014) se recolectaron con la valoración de los expertos con dos categorías “Aprobado” y “Requiere ajuste”. Se estimó la fiabilidad de las valoraciones de los cuatro expertos con acuerdo directo y con el coeficiente Kappa de Fleiss (Davies & Fleiss, 1982). También se revisó el acuerdo entre parejas de expertos de forma directa y con el coeficiente Kappa de Cohen (1968).

Se realizaron análisis psicométricos con la TCT y se complementaron con estimaciones del modelo de Rasch de la TRI para respuestas politómicas (Samejima, 1969). Con la TCT se obtuvo la discriminación de los ítems, con la correlación ítem-total y también se estimó la fiabilidad de consistencia interna con el coeficiente Alfa de Cronbach (1951) y el Omega de McDonald (1999). Para obtener el ajuste del modelo de Rasch, se estimaron los residuales cuadráticos medios (*MNSQ*) en los índices de ajuste interno (*Infit*) y externo (*Outfit*), junto con la prueba de hipótesis χ^2 para comparar las distribuciones empíricas con la del modelo.

Para recolectar las evidencias de validez basadas en la estructura interna de la escala (AERA, APA & NCME, 2014), se realizaron AFC a tres modelos (unidimensional de conducta prosocial; bidimensional integrado de las subescalas de Ayuda y Confortar al otro y bidimensional con las dos subescalas de forma independiente). El modelo con mejor bondad de ajuste se sometió a una evaluación progresiva de invarianza factorial (Elosua, 2005), con restricciones configural, métrica, escalar y residual respecto al sexo. Adicionalmente, en ese modelo se calibraron los ítems con el modelo de Rasch para identificar las conductas que caracterizan a las personas con mayores niveles de conducta prosocial.

El procesamiento de la información se realizó con el lenguaje de programación R (R Core Team, 2017), con la interfaz R Studio y el complemento ULLRToolbox (Hernández, 2019). Se utilizaron las librerías psych (Revelle, 2022), irr (Gamer et al., 2012), nortest (Gross & Ligges, 2012), eRm (Mair & Hatzinger, 2008), lavaan (Rosseel, 2012) y seemTools (Jorgensen et al., 2019).

Resultados

Revisión de contenido

La valoración de los expertos tuvo un acuerdo global del 60% con un Kappa de Fleiss $k=0,317$ y un $p=0,003$. El mayor acuerdo se obtuvo entre los expertos 2 y 4 con un 93,33% de coincidencias y un Kappa de Cohen $k=0,860$, seguido por los acuerdos entre los expertos 1 y 3 y 3 y 4 con un 80% de coincidencias y Kappas de Cohen $k=0,000$ y $0,550$ respectivamente. Los expertos 2 y 3 tuvieron un 73,33% de coincidencias con un Kappa de Cohen $k=0,330$, seguido por los expertos 1 y 2 con un acuerdo del 66,67% con un Kappa de Cohen $k=0,000$ y el menor acuerdo logrado fue entre los expertos 1 y 4 con un 60,00% de coincidencias y un Kappa de Cohen $k=0,000$. Al revisar las valoraciones individuales se identificó que el experto 1 fue el más laxo, porque aprobó todos los ítems sin realizar observaciones. Se volvió a estimar el acuerdo global excluyendo al experto 1 y se obtuvo un acuerdo directo del 73,33% con un Kappa de Fleiss $k=0,585$ y un $p=0,000$.

Los cuatro expertos aprobaron nueve ítems (1, 3, 4, 8, 10, 11, 12, 13 y 14) y se realizaron sugerencias a los restantes. Se contrastaron las sugerencias con los resultados psicométricos reportados por Auné (2018) y Auné y Attorresi (2017), específicamente con las curvas de probabilidad de las opciones de respuesta y las cargas factoriales. Se decidió modificar los ítems 2, 5, 6, 7 y 9 con base en las sugerencias de los expertos y eliminar el ítem 15 “Aconsejo a conocidos sobre trabajo” con fundamento en los reportes empíricos del estudio original y la relación del contenido de este con la respectiva dimensión. Los expertos revisaron nuevamente la ECP, aprobaron los ajustes y la decisión del eliminar el ítem 15 para la adaptación del cuestionario a población colombiana.

Fiabilidad y Análisis de Ítems

Respecto a la fiabilidad, los puntajes globales tuvieron un $\alpha=0,87$, en la dimensión Ayuda fue de $\alpha=0,85$ y en la dimensión Confortar al otro fue de $\alpha=0,77$. Estos resultados evidencian una alta intercorrelación de los ítems, lo que permite inferir que las variaciones de las puntuaciones en la escala global y las subescalas presentaron un bajo error no sistemático. Según los lineamientos mencionados por Peterson (1994) estos valores corresponden a un nivel moderado-alto de fiabilidad, que se encuentra por encima de la media y la mediana obtenidas generalmente con escalas Likert asociadas a reportes conductuales, pero que están por debajo de los valores sugeridos para investigación aplicada.

En la estimación de la fiabilidad con el coeficiente Omega se obtuvieron Omegas jerárquicos $\omega=0,59$ (global), $\omega=0,74$ (Ayuda) y $\omega=0,58$ (Confortar al otro); Omegas asintóticos $\omega=0,66$ (global), $\omega=0,83$ (Ayuda) y $\omega=0,67$ (Confortar al otro); y Omegas totales $\omega=0,90$ (global), $\omega=0,89$ (Ayuda) y $\omega=0,86$ (Confortar al otro). En contraste con los valores obtenidos con el Alfa de Cronbach, los resultados con el Omega de McDonald presentan grandes diferencias que evidencian el incumplimiento del principio de tau-equivalencia requerido para emplear el Alfa de Cronbach. En este sentido, el coeficiente Omega es el más adecuado para analizar la consistencia interna del ECP 14 y sus dos dimensiones.

Con el Omega Jerárquico se pudo identificar que la dimensión Ayuda presentó la mayor fiabilidad asociada a un factor general y esta puede aumentar hasta un $\omega=0,83$ (Omega asintótico), cuando los ítems tienden a un incremento infinito (De Reizábal, 2017). La consistencia interna de los 14 ítems de la ECP se afectó por la existencia de las dos dimensiones y por la moderada intercorrelación de los ítems en la dimensión Confortar al otro. A pesar de estos hallazgos, con el Omega total se identificó una varianza total de 0,90 asociada al conjunto de ítems de la escala.

Por otro lado, el análisis de ítems se hizo por subescala con la TCT y se complementó con estimaciones del modelo de Rasch, con el cual no se pudieron realizar los cálculos para el ítem 14 porque ningún participante respondió “*Nunca*” y el modelo no pudo converger exitosamente con esta respuesta *outlier* (0%). En la Tabla 1 se describen los resultados obtenidos en el análisis de ítems y se puede identificar que todos los ítems presentaron correlaciones altas y no aportaron errores significativos a la fiabilidad de las respectivas subescalas, porque la discriminación estimada con la Correlación Ítem-Total tuvo valores mayores de 0,20 (Likert, 1932) y el α no aumenta sustancialmente si alguno de los ítems es eliminado.

Para los análisis con el modelo de Rasch, se aplicaron los criterios propuestos por Hodge y Morgan (2017) y se identificó que solo tres ítems (6, 10 y 11) tuvieron índices Outfit e Infit por fuera de los valores aceptados para la bondad de ajuste (entre 0,70 y 1,30). Esto significa que el modelo de Rasch tuvo una baja predicción de las respuestas dadas por los participantes a estos tres ítems. El ítem 12 obtuvo un valor levemente superior en el índice *Infit*, en este caso particular significa que el modelo de Rasch perdió poder predictivo en las respuestas de los participantes con parámetros θ cercanos al parámetro b de este ítem. Respecto a la prueba χ^2 y su p-valor, se encontró que las distribuciones de las probabilidades de los ítems 1 y 12 presentaron diferencias estadísticamente significativas en contraste con la distribución establecida para el modelo de Rasch.

En contraste de ambos análisis, con la TCT no hubo hallazgos para realizar la revisión cualitativa de los ítems, sin embargo, con el modelo de Rasch se tuvo que revisar el contenido de cuatro ítems (6,10, 11 y 12) porque no hubo ajuste del modelo en uno o dos de los índices y esto indica una baja predicción de las respuestas en estos ítems.

Tabla 1
Análisis de ítems

Ítem	Dimensión	TCT		Modelo de Rasch			
		Corr. ítem subescala	α si se elimina el ítem	X^2	p	Infit MSQ	Outfit MSQ
1. Relego mi beneficio personal para ayudar a otros.	Ayuda	0,39	0,86	402,070	0,015*	1,095	1,169
2. En mi tiempo libre realizo actividades como voluntario.	Ayuda	0,66	0,83	343,159	0,487	0,977	0,998
3. Cuando siento que alguien está mal le demuestro que lo entiendo.	Confortar al otro	0,61	0,73	277,180	0,996	0,771	0,808
4. Actúo como bastón de los demás.	Confortar al otro	0,63	0,73	259,082	1,000	0,758	0,755
5. Siento el dolor ajeno como propio.	Confortar al otro	0,59	0,74	285,771	0,988	0,843	0,833
6. Acostumbro a comprometerme con causas benéficas.	Ayuda	0,81	0,82	218,196	1,000	0,637	0,634
7. Dedico un tiempo importante de mi vida para realizar acciones que mejoren el mundo en el que vivimos.	Ayuda	0,63	0,84	326,882	0,726	0,929	0,950
8. Me ocupo del bienestar de cualquier individuo, grupo o comunidad.	Ayuda	0,72	0,83	247,919	1,000	0,729	0,721
9. Vivo con lo necesario y reparto todo lo demás.	Ayuda	0,58	0,84	335,466	0,604	0,997	0,975
10. Me pongo en el lugar del otro.	Confortar al otro	0,64	0,72	223,691	1,000	0,646	0,652
11. Participo en actividades solidarias.	Ayuda	0,83	0,81	183,726	1,000	0,529	0,534
12. Hago donaciones a organizaciones benéficas.	Ayuda	0,55	0,85	431,791	0,001**	1,322	1,255
13. Si una persona me cuenta un conflicto, intento que comprenda el punto de vista de la otra parte.	Confortar al otro	0,53	0,76	374,350	0,110	1,184	1,033
14. Intento "levantar" la autoestima a mis amigos.	Ayuda	0,59	0,74			No converge	

Nota. * $p < 0,05$; ** $p < 0,01$.

Evidencias basadas en la estructura interna

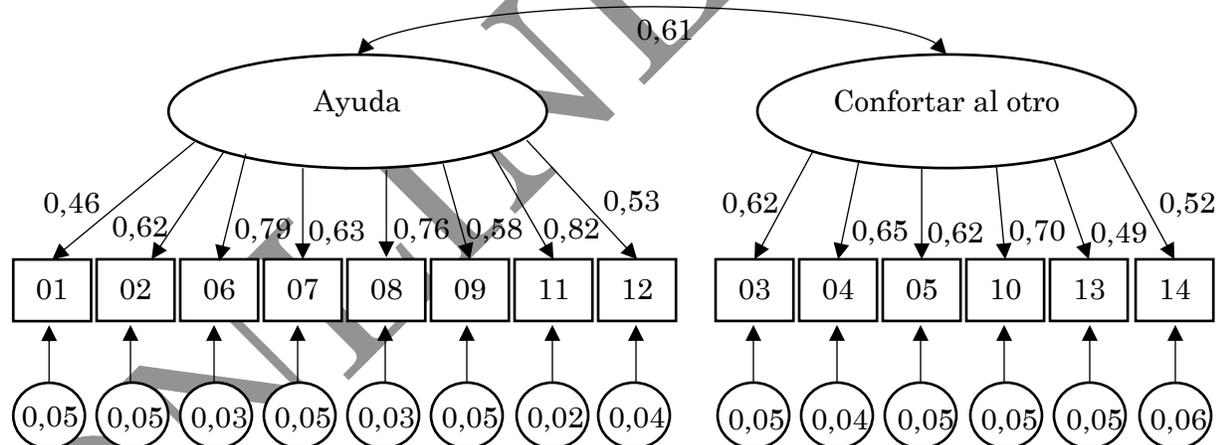
Estructura Factorial. Se sometieron a prueba tres modelos AFC: a) un modelo unidimensional que contempló la conducta prosocial en general; b) uno bidimensional que se conformó por dos dimensiones relacionadas Ayuda y Confortar al otro y c) uno bidimensional conformado por dos dimensiones separadas Ayuda y Confortar al otro. En la Tabla 2 se relacionan los índices de bondad de ajuste de los tres modelos analizados, de los cuales el modelo con mejor bondad de ajuste fue el que contempló dos dimensiones relacionadas, con $p < 0,05$ e índices de ajuste normalizado (*NFI*), no normalizado (*NNFI*), comparativo (*CFI*) y Tucker-Lewis (*TLI*) superiores a 0,95 y el error medio cuadrático de aproximación (RMSEA) menor a 0,05, conforme con los criterios de Browne y Cudeck (1993), Bentler y Bonnet (1980) y Martínez et al. (2012).

Tabla 2
Modelos AFC

Bondad de ajuste	Modelo 1	Modelo 2	Modelo 3	
	Unidimensional	Bidimensional integrado	Bidimensional segmentado	
			Ayuda	Confortar al otro
X^2	269,976	130,176	21,034	35,831
gl	77	76	20	9
p	0,000	0,000	0,395	0,000
NFI	0,913	0,958	0,987	0,941
$NNFI$	0,925	0,979	0,999	0,925
CFI	0,936	0,982	0,999	0,955
TLI	0,925	0,979	0,999	0,925
$RMSEA$	0,085	0,046	0,012	0,093
$[IC\ 90\%]$	[0,075; 0,097]	[0,032; 0,059]	[0,000; 0,04]	[0,063; 0,126]

En la Figura 1 se ilustra el modelo bidimensional integrado y se puede observar que los ítems tuvieron pesos de regresión estandarizados superiores a 0,40 en las respectivas dimensiones y errores menores o iguales a 0,06.

Figura 1
Estructura Factorial Modelo 2



Se evaluó que el modelo 2 fuera invariante respecto a la variable “sexo”, para verificar que la estructura factorial no variara entre mujeres y hombres y descartar un posible sesgo. En la Tabla 3 se muestra el análisis progresivo de invarianza realizado al modelo bidimensional y se puede observar que este modelo fue invariante entre mujeres y hombres, en la configuración básica del modelo (configural), en la equivalencia de los pesos factoriales (métrica) e interceptos (escalar) y en las varianzas y covarianzas residuales (residual). Esta interpretación se fundamenta en que los contrastes de X^2 y gl , se obtuvieron valores $p > 0,50$ con $\Delta TLI < 0,050$ (Lippke et al., 2007), sumado a valores $\Delta CFI > -0,010$ y $\Delta RMSEA < 0,010$ (Cheung & Rensvold, 2002).

Tabla 3
Invarianza Factorial Modelo 2 con la variable Sexo

Invarianza	gl	X^2	ΔX^2	Δgl	P	Contraste	ΔTLI	ΔCFI	$\Delta RMSEA$
Configural	152	164,11							
Métrica	164	184,66	10,176	12	0,601	Métrica- Configural	0,039	0,027	-0,012
Escalar	176	190,84	10,684	12	0,556	Escalar- Métrica	0,007	0,002	-0,002
Residual	190	197,27	10,473	14	0,727	Residual- Escalar	0,012	0,007	-0,004

Calibración de los ítems. Para ubicar y organizar las conductas prosociales descritas en los ítems en una escala invariante, se procedió a estimar los parámetros b y θ con el modelo de Rasch. En la Tabla 4 se relacionan los parámetros b calculados de forma independiente en las dos dimensiones Ayuda y Confortar al otro, para poder cumplir con el supuesto de unidimensionalidad que se requiere en este modelo. Los resultados en la dimensión Ayuda tuvo un rango de 17,328 logits ubicado entre -0,787 y 16,541, mientras que en la dimensión Confortar al otro el rango fue de 12,393 que se ubicó entre -0,681 y 11,712 logits. El orden ascendente establecido con el modelo de Rasch para las conductas de los ítems de la dimensión Ayuda fue 1, 7, 8, 11, 6, 9, 12 y 2, lo que indica que el ítem 1 describe la conducta que presentó el menor nivel de ayuda y el ítem 2 la conducta con el nivel más alto. En la dimensión Confortar al otro el orden de los ítems fue 3, 10, 13, 5 y 4.

En la dimensión Ayuda los parámetros con menor error estándar se obtuvieron en las estimaciones realizadas a la categoría 1 (*Nunca*) y 2 (*Casi Nunca*) con valores menores a 0,33, criterio que se emplea para considerar que un ítem está calibrado (Romero Morales et al., 2006). En la dimensión Confortar al otro, esto se logró solo en la categoría 1 y hubo errores estándar superiores a 1,00, que indican que las estimaciones de los ítems de esta dimensión fueron menos precisas en comparación con la dimensión Ayuda.

Tabla 4
Parámetros b

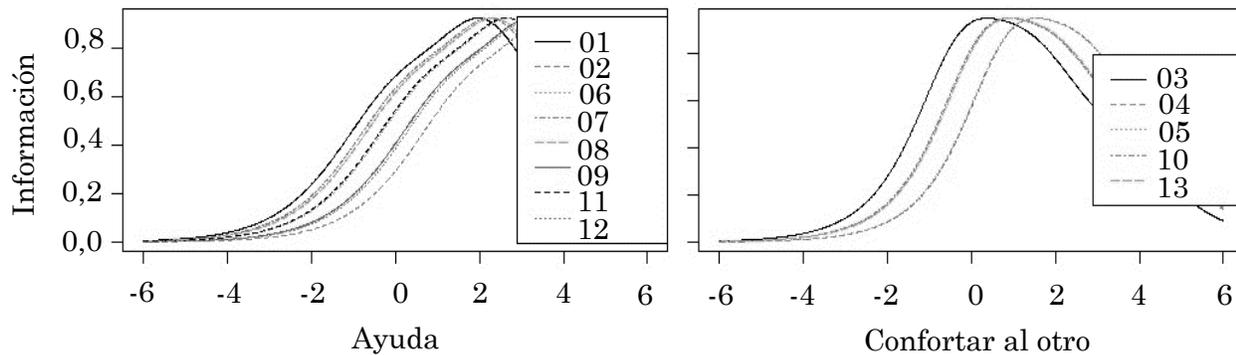
Ítem	Dimensión	Parámetro b [Error estándar]				
		$b1$	$b2$	$b3$	$b4$	$b5$
1.	Ayuda	-0,787 [0,060]	-0,358 [0,161]	1,472 [0,260]	3,939 [0,368]*	7,948 [0,493]*
2.	Ayuda	0,931 [0,065]	3,080 [0,191]	6,628 [0,329]	10,813 [0,486]*	16,541 [0,667]*
3.	Confortar al otro	-0,681 [0,065]	-0,893 [0,349]*	0,218 [0,607]*	2,137 [0,878]*	5,866 [1,157]**
4.	Confortar al otro	0,488 [0,057]	1,446 [0,353]*	3,726 [0,625]*	6,814 [0,919]*	11,712 [1,229]**
5.	Confortar al otro	0,467 [0,057]	1,403 [0,352]*	3,661 [0,624]*	6,728 [0,917]*	11,605 [1,226]**
6.	Ayuda	-0,063 [0,056]	1,090 [0,162]	3,644 [0,272]	6,835 [0,399]*	11,568 [0,550]*
7.	Ayuda	-0,521 [0,057]	0,175 [0,159]	2,271 [0,261]	5,005 [0,375]*	9,281 [0,511]*
8.	Ayuda	-0,433 [0,057]	0,350 [0,159]	2,534 [0,263]	5,354 [0,379]*	9,718 [0,517]*
9.	Ayuda	0,448 [0,059]	2,113 [0,173]	5,178 [0,297]	8,881 [0,439]*	14,126 [0,605]*
10.	Confortar al otro	-0,155 [0,057]	0,158 [0,345]*	1,794 [0,607]*	4,239 [0,884]*	8,493 [1,175]**
11.	Ayuda	-0,106 [0,056]	1,005 [0,161]	3,516 [0,271]	6,664 [0,396]*	11,355 [0,546]*
12.	Ayuda	0,532 [0,060]	2,280 [0,176]	5,429 [0,302]	9,215 [0,446]*	14,543 [0,615]*
13.	Confortar al otro	-0,119 [0,057]	0,231 [0,345]*	1,904 [0,607]*	4,385 [0,885]*	8,676 [1,177]**
14.	Confortar al otro			No converge		

Nota. *Error estándar >0,33; **Error estándar >1,00.

La Función de Información del Test permite ilustrar la función inversa de los errores de medida y se obtiene de la suma de las Funciones de Información de los Ítems (Prieto & Delgado, 2003). En la Figura 2 se grafica de forma segmentada esta función en ambas dimensiones, lo que permitió identificar que la mayor precisión en la dimensión Ayuda se obtuvo entre 2,00 y 4,00 logits, porque el menor error estándar en la estimación de los parámetros b se ubicó en esa posición. En la dimensión Confortar al otro la mayor precisión se ubicó entre 0,00 y 2,00 logits. Estos resultados muestran que la dimensión Ayuda abarcó el mayor rango de medición y con mayor precisión, en comparación con la dimensión Confortar al otro.

En la estimación de los parámetros θ de la dimensión Ayuda se distribuyó en un rango de 6,68 logits (entre -1,40 y 5,29), con una $M=2,03$ y una $DE=1,06$. Los errores estándar de las estimaciones en esta dimensión tuvieron una $M=0,40$ con una $DE=0,04$, un valor mínimo de 0,38 y un máximo de 0,75. Respecto a los parámetros θ de la dimensión Confortar al otro, el rango fue de 5,67 logits (entre 0,28 y 5,95), una $M=2,81$ y una $DE=1,09$. Los errores estándar en esta dimensión tuvieron una $M=0,55$ con una $DE=0,11$, un valor mínimo de 0,46 y un máximo de 1,08. A una persona no se le pudo estimar el parámetro θ , porque tuvo un comportamiento *outlier* (100%) que se caracterizó por puntuar la categoría más alta en todos los ítems de la dimensión.

Figura 2
Funciones de Información del Test



Discusión

El principal objetivo de este estudio fue recolectar evidencias de validez del contenido de los ítems y de la estructura interna de la escala, para realizar una adaptación preliminar de la ECP de Auné & Attorresi (2017) y proyectar su uso con población colombiana. En este proceso se siguieron las directrices de traducción y adaptación de pruebas de la ITC (2017), aunque no fue necesario traducir los ítems porque la escala original se construyó en español y fue planteada dentro de un contexto latinoamericano, que contempla aspectos contextuales similares entre la población con la que se diseñó y validó la escala (Argentina) y la población a adaptar (Colombia). Esta condición permitió minimizar la influencia de diferencias culturales y lingüísticas irrelevantes para la interpretación de los puntajes obtenidos en la población a adaptar.

Las evidencias de validez del contenido de los ítems se obtuvieron con una valoración de expertos, la cual conllevó a ajustar la redacción del enunciado de cinco ítems (2, 5, 6, 7 y 9) para aclarar y delimitar las conductas descritas, porque los expertos consideraron que no cumplía con las indicaciones sugeridas por Moreno et al. (2004) y Elejabarrieta e Iñiguez (2008), las cuales indican que: a) la redacción original le faltaba especificidad en la descripción de la conducta descrita; b) incluía más de una conducta en un mismo ítem o c) permitía un amplio margen de interpretación, lo cual no es deseado para ítems de respuesta cerrada. En comparación con la adaptación realizada por Canales Reyes (2020) con población peruana, se ajustaron menos ítems y solo coincidió que hubo modificación del ítem 9. En la versión original la redacción del ítem es “Me quedo con lo justo y necesario para vivir y reparto todo lo demás”, en la adaptación peruana se ajustó a “Comparto lo que tengo con los demás” y en este estudio el ítem quedó redactado como “Vivo con lo necesario y reparto todo lo demás”. En el ajuste realizado en este estudio al ítem 9, se buscó mantener la literalidad

del estudio original y solo se modificó la extensión de la afirmación con un planteamiento más corto y conciso. Los resultados psicométricos evidencian que este ítem no presentó problemas de interpretación, porque no se encontraron hallazgos con los métodos TCT y el modelo de Rasch.

En las opciones de respuesta de todos los ítems se incluyó información cuantitativa, para facilitar la comprensión y diferenciación de las frecuencias descritas en la lectura de las opciones (*Nunca-el 0% de las veces; Casi Nunca-entre el 1% y el 19% de las veces; A veces-entre el 20% y el 59% de las veces; Con frecuencia-entre el 60% y el 79% de las veces; Casi siempre-entre el 80% y el 99% de las veces y Siempre-el 100% de las veces*), para disminuir la subjetividad al momento de leer las afirmaciones y relacionarlas con una frecuencia.

La eliminación del ítem 15 “Aconsejo a conocidos sobre trabajo” se fundamentó en los reportes psicométricos de los estudios originales y el contenido de este respecto a la dimensión que mide. En los resultados psicométricos reportados por Auné (2018) y Auné y Attorresi (2017), se identifica que las curvas de probabilidad mostraron diferencias solo en las categorías 3 y 6, lo que evidencia que no funciona como una respuesta graduada de seis opciones. Adicionalmente, en el AFE realizado por Auné (2018) y Auné y Attorresi (2017), este ítem presentó las menores cargas factoriales (0,41 en la matriz de configuración y 0,44 en la matriz de estructura). Sumado a esto y relacionado con el contenido del ítem, el verbo “aconsejar” no es sinónimo de “confortar” y se relaciona más con conductas verbales de orientación u opinión que se brinda a otra persona.

Para los análisis psicométricos, a pesar de realizar un muestreo no probabilístico, se obtuvo una muestra suficientemente variada en edad, nivel educativo máximo alcanzado y ocupación y solo el 9,59% eran estudiantes universitarios. Aunque se logró una amplia variabilidad de diferentes grupos poblacionales de Colombia, no se recomienda generalizar los resultados a regiones diferentes a la ciudad capital porque la muestra se concentró significativamente en personas que nacieron en esta locación. Conforme con las recomendaciones de Vallejo (2013) se obtuvo un tamaño de muestra que permitió aplicar los métodos estadísticos planteados para los análisis de datos, especialmente el AFC que requiere una relación igual o superior de 10 datos por cada ítem. Sin embargo, Meade y Bauer (2007) recomiendan tener al menos una muestra de 200 por grupo comparado, en los análisis de invarianza factorial y en este estudio el tamaño de muestra de los hombres fue insuficiente porque solo participaron 94, lo que afectó puntualmente el cálculo del coeficiente Gamma.

Los resultados del análisis con los métodos de la TCT evidencian que todos los ítems discriminaron y se encontraron fiabilidades sustancialmente iguales, entre la escala original aplicada a una muestra de población argentina (Auné & Attorresi, 2017) y la versión de 14 ítems que se aplicó a la muestra colombiana. En ambas dimensiones se obtuvo la misma fiabilidad, un $\alpha=0,85$ en la dimensión de Ayuda y un $\alpha=0,77$ en Confortar al otro. El panorama fue similar en contraste con la adaptación con población peruana, porque las diferencias fueron ínfimas con un $\alpha=0,86$ en la dimensión de Ayuda y un $\alpha=0,79$ en Confortar al otro y $\omega=0,87$ y $\omega=0,83$ respectivamente.

A pesar de esta similitud en los resultados de fiabilidad con el Alfa de Cronbach, se sugiere plantear como base la fiabilidad obtenida con el coeficiente Omega de McDonald debido a que los ítems no presentaron medidas tau-equivalentes. Estos análisis permiten identificar que la dimensión Ayuda presentó mayor fiabilidad y discriminación en los ítems en comparación con la dimensión Confortar al otro.

En los análisis con el modelo de Rasch, se logró cumplir el supuesto de unidimensionalidad en la estructura confirmada con el AFC y esto conllevó a realizar las estimaciones de los parámetros de ambas dimensiones por separado. En la dimensión Ayuda en cuatro de los nueve ítems (6, 11, 12 y 14) el modelo no ajustó. Aunque en estos ítems hubo hallazgos, los contenidos planteados fueron coherentes con la conceptualización de la dimensión Ayuda y no se identifican aspectos que explicaran estos resultados psicométricos. Al no encontrar argumentos para sugerir la eliminación o modificación de estos ítems, se mantienen en la versión que se utilizará para medir la conducta prosocial en población colombiana.

En la dimensión Confortar al otro el modelo no ajustó en el ítem 10 y la conducta descrita en este ítem “Me pongo en el lugar del otro” se relaciona directamente con el constructo “empatía”. En sus versiones preliminares, se contemplaron cuatro dimensiones para la ECP (comportamientos empáticos, altruismo, ayuda y compartir) y este ítem pertenecía a la primera dimensión (Auné & Abal, 2016). En el siguiente desarrollo del cuestionario, los autores lo redujeron a dos dimensiones, porque no se logró la unidimensionalidad esperada (Auné & Attorresi, 2017). En este cambio se produjo una reinterpretación del factor “comportamientos empáticos” y se redefinió a “confortar al otro”, como consecuencia de la incorporación de ítems que pertenecían a otras dimensiones. Con este contexto se puede concluir que, al

momento de redactar el ítem, se basó desde una propuesta de comportamiento empático y por esto el gran énfasis del contenido hacia la empatía. A pesar del desajuste del modelo de Rasch, se recomienda mantener este ítem con base en el contexto del desarrollo del cuestionario y adicionalmente, se requiere confirmar el desajuste con una muestra más grande para descartar errores tipo I y II. Además, en los estudios de Garaigordobil y García de Galdeano (2006) se encontraron correlaciones positivas significativas entre la conducta prosocial y la empatía en una muestra de infantes. En coherencia con este hallazgo, Silke, et al. (2018) realizaron una revisión sistemática y encontraron 168 artículos científicos que documentan la asociación entre estas dos variables.

Con relación a la recolección de evidencias de la estructura interna de la escala, el AFC permite descartar el modelo 1 (unidimensional), porque la razón X^2/df fue mayor a 3,00, los índices *NFI*, *NNFI*, *CFI* y *TLI* fueron inferiores a 0,95 y el *RMSEA* fue mayor a 0,05. Se confirmaron los modelos 2 y 3 que tienen estructuras de dos dimensiones similares a los establecidos por Auné y Attorresi (2017). El modelo 2 (bifactorial con las dimensiones relacionadas) tuvo una razón X^2/df menores a 3,00, *NFI*, *NNFI*, *CFI* y *TLI* superiores a 0,95 y *RMSEA* menor a 0,05. El modelo 3 (bifactorial con dimensiones independientes) tuvo mejores índices *NFI*, *NNFI*, *CFI* y *TLI* en la dimensión de Ayuda, pero con un $p > 0,05$ y en la dimensión Confortar al otro, la razón X^2/df fue mayor a 3,00, los índices *NFI*, *NNFI* y *TLI* fueron menores a 0,95 y el *RMSEA* superior a 0,05. Los resultados del modelo 3 son coherentes con la propuesta de Auné et al. (2020) de contemplar la posibilidad de solo medir la dimensión Ayuda, sin embargo, por obtener un $p > 0,05$ en el AFC, se recomienda mantener el modelo bidimensional con dimensiones relacionadas.

El modelo de dos dimensiones relacionadas no mostró diferencias significativas entre hombres y mujeres a nivel configural, métrico, escalar y residual. Los resultados evidencian que la estructura factorial analizada es invariante entre ambos sexos. En la muestra hubo una persona que se identificó con otro sexo, por lo que se sugiere realizar otros estudios en los que se incluyan más participantes de este sexo para evaluar si la invarianza factorial se mantiene.

Con la calibración de los ítems con el modelo de Rasch se pudo identificar que la dimensión Ayuda tuvo un rango más amplio y preciso en comparación de la dimensión Confortar al otro, esto indica una mejor medición y es coherente con los resultados de los análisis TCT y AFC. El ítem 2 “En mi tiempo libre realizo actividades como voluntario” obtuvo el mayor parámetro *b* en la dimensión Ayuda, lo que muestra que esta conducta es la que evidencia mayores niveles de conducta prosocial y es coherente con los hallazgos de Picazo et al. (2020), quienes encontraron que las personas que invierten su tiempo libre en ayudar a otros suelen hacerlo de forma más prolongada, porque es parte de su realización personal y lo hacen para buscar la felicidad. En la dimensión Confortar al otro el ítem con el mayor parámetro *b* fue el 4 “Actúo como bastón de los demás”, que describe una conducta que expone de forma explícita la disposición física de apoyo a otras personas, en comparación a las conductas de los otros ítems que se refieren a disposiciones y actitudes hacia la acción de confortar. En la adaptación con población peruana se decidió cambiar la palabra “bastón” por “apoyo” que es una expresión más general, sin embargo, en esta adaptación a población colombiana se optó por modificar lo menos posible el cuestionario. A pesar de que la expresión “actúo como bastón” puede estar asociada a una jerga específica, los resultados psicométricos evidenciaron que no tuvo problemas de interpretación, porque los índices de la TCT y el ajuste del modelo de Rasch no detectaron anomalías en las distribuciones de las respuestas. Estos resultados demuestran que esta expresión se comparte en las ciudades capitales de Argentina y Colombia, porque las muestras de ambos estudios fueron obtenidas en estas ubicaciones geográficas. Para adaptaciones en poblaciones de otros países, se recomienda tener en cuenta la expresión utilizada en este ítem por la carga cultural que evidentemente tiene y puede contemplarse la modificación realizada en la adaptación con población peruana.

Para realizar un adecuado uso de la ECP, se recomienda calificarla en las dos dimensiones y transformar los puntajes con base en las especificaciones para escalas Thurstone, Guttman y Likert propuestas por García Sánchez et al. (2011), para asignar las frecuencias descritas en los ítems a los puntajes obtenidos en cada dimensión. Este resultado se obtiene de sumar los valores de las respuestas (entre 1 y 6) y luego el total se divide por la cantidad de ítems que conforman cada dimensión.

Los resultados de este estudio permiten conocer las calidades psicométricas de la ECP de Auné y Attorresi (2017) y se amplían para la EA de Auné et al. (2020) con los resultados obtenidos en la subescala de Ayuda. Se recomienda la ECP para medir la autopercepción de la conducta prosocial desde una perspectiva bidimensional y la EA para medir la autopercepción de conductas relacionadas con acciones positivas y voluntarias para beneficiar a otros (Auné et al., 2020). Tanto la ECP (Auné & Attorresi, 2017)

como la EA (Auné et al., 2020) pueden utilizarse en población colombiana, pero se recomienda hacer uso en investigaciones que apliquen los cuestionarios en una muestra que permita obtener resultados psicométricos y compararlos con los resultados de este estudio.

Los hallazgos no solo dan cuenta de la calidad psicométrica de la ECP para su uso en población colombiana, también aportan evidencias sobre las dimensiones que conforman la conducta prosocial. Los resultados del AFC muestran que los comportamientos asociados con ayudar a otras personas tuvieron mayor peso en el modelo, en comparación con las conductas dirigidas a confortar a las otras personas. Adicionalmente, la confiabilidad estimada con la TCT también fue más alta en esta dimensión, lo que indica una mejor correlación entre los ítems que describen las conductas de ayuda. En coherencia con estos resultados, el modelo de Rasch permitió identificar que los ítems de la dimensión Ayuda abarcaron un rango más amplio y con mejor precisión de medida, lo que significa que hay una mayor representación de conductas de ayuda dentro de un continuo expresado en escala logit y que se relaciona con la conducta prosocial.

Dentro de las tipologías que definen y clasifican la conducta prosocial, Warneken y Tomasello (2009) mencionan tres dimensiones a partir de la revisión de varios planteamientos teóricos y resultados empíricos. Estas dimensiones son: a) confortar; b) compartir información y c) ayuda instrumental, esta última más compleja que las otras dos y paradójicamente menos estudiada. La ayuda instrumental abarca aquellos comportamientos inclinados a actuar en nombre de los demás y se puede identificar un origen ontogénico y filogénico (Warneken & Tomasello, 2009). Al revisar el contenido de los ítems de la dimensión Ayuda de la ECP, se logra reconocer que en cada uno de estos se declaran conductas en las que sus objetivos están encaminados a actuar hacia el bienestar de otras personas y en este sentido, la dimensión de Ayuda encaja dentro del planteamiento de ayuda instrumental que es mencionado por Warneken y Tomasello (2009). Al tener componentes cognitivos y motivacionales de origen ontogénico, la ayuda instrumental puede expresarse en una mayor cantidad de conductas dentro de un amplio margen de probabilidad de ocurrencia, dada la alta variabilidad que puede generarse por las diferencias individuales en estos procesos psicológicos. Esta característica puede estar relacionada con el hecho de haber obtenido una mayor amplitud en la dimensión Ayuda, al transformarla en una escala logit con el modelo de Rasch. Con este tipo de análisis, la dimensión de Ayuda de la ECP o EA de Auné et al. (2020) se puede emplear para continuar con el estudio del planteamiento conceptual de la ayuda instrumental, específicamente en población adulta.

En términos generales, aunque los resultados de este estudio fueron similares a los que obtuvieron Auné y Attorresi (2017) con la escala original y Canales Reyes (2020) en la adaptación con población peruana, este estudio tuvo limitaciones en el diseño muestral y tamaño de la muestra, específicamente para realizar las estimaciones de la invarianza factorial, que conllevan a tomar con precaución la posibilidad de generalizar de las conclusiones sobre las propiedades psicométricas de la ECP en toda la población colombiana. Esta misma limitación impidió el uso de métodos para detectar DIF, porque para analizar cuestionarios de 20 o menos ítems se requiere como mínimo una muestra de 200 en cada grupo (Scott et al., 2009). Otra limitación fue la falta de identificación de la procedencia o residencia de los participantes, debido a las diferencias culturales que se han identificado entre las regiones que conforman a Colombia.

Finalmente, para las futuras investigaciones que analicen las propiedades psicométricas de la escala, se sugiere seguir las recomendaciones de Pedrosa et al. (2013), que implica aumentar el número de expertos en la revisión del contenido de los ítems e incorporar más métodos que permitan contrastar la valoración del acuerdo de los expertos. Para ampliar la recolección de evidencias de la estructura interna de la escala, se recomienda realizar estudios de invarianza y DIF entre regiones colombianas y latinoamericanas para conocer sus propiedades psicométricas a nivel intercultural. También sería relevante estimar fiabilidad test-retest para identificar la estabilidad de las puntuaciones de la escala a través del tiempo. Por último, se requiere continuar con este tipo de estudios para recolectar más evidencias de validez con otras fuentes (AERA, APA & NCME, 2014), por ejemplo, las consecuencias derivadas en el uso de la escala y la relación entre los puntajes de la ECP y variables asociadas con la conducta prosocial.

Referencias

- Agudelo, L.M, Giraldo, C. A., Gaviria, M., B., Rodríguez, M., A., Gómez, J. F., Gallón, A. & Pérez, A. (2000). Encuesta de Comportamientos Agresivos y Prosociales (COPRAG). COLCIENCIAS – Instituto de Ciencias de la Salud CES – Universidad de Antioquia.
- American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.

- 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]. https://ri.conicet.gov.ar/bitstream/handle/11336/83189/CONICET_Digital_Nro.2da682ec-dff9-4198-814c-5cc6d3738198_A.pdf
- Auné, S. E. & Abal, F. J. P. (2016). Diseño y construcción de una escala de conducta prosocial para adultos. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 2(42), 15-25. http://doi.org/10.21865/RIDEP42_15
- Auné, S. E., Abal, F. J. P. & Attorresi, H. F. (2020). Análisis psicométrico de una escala de ayuda con el modelo de respuesta graduada. *Psykhé*, 29(2), 1-14. <http://doi.org/10.7764/psykhe.29.2.1472>
- Auné, S. & Attorresi, H. (2017). Dimensionalidad de un test de conducta prosocial. *Revista Evaluar*, 17(1), 29-37. <https://doi.org/10.35670/1667-4545.v17.n1.17072>
- Auné, S., Blum, G., Abal, F., Lozzia, G. & Attorresi, H. (2014). La conducta prosocial: Estado actual de la investigación. *Perspectivas en Psicología*, 11(2), 21-33. <https://www.redalyc.org/articulo.oa?id=483547666003>
- Bentler, P. M. & Bonnet, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606. <https://doi.org/10.1037/0037-2909.88.3.588>
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*, (pp. 136-162). Newbury Park, CA: Sage.
- Canales Reyes, L. E. (2020). Adaptación de la escala de conducta prosocial (ECP-P) en adultos de Lima [Trabajo de Grado, Pregrado, Facultad de Ciencias de la Salud, Universidad César Vallejo, Perú]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/55516>
- Caprara, G. V. & Pastorelli, C. (1993). Early emotional instability, prosocial behavior, and aggression: some methodological aspects. *European Journal of Personality*, 7(1), 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(2), 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(1), 31-44. <https://doi.org/10.1023/A:1014033032440>
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4), 213-220. <https://doi.org/10.1037/h0026256>
- Cronbach, L. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16, 1-16. <https://link.springer.com/article/10.1007/bf02310555>
- Davies, M. & Fleiss, J. L. (1982). Measuring agreement for multinomial data. *Biometrics*, 1047-1051. <https://doi.org/10.2307/2529886>
- De Reizábal, M. L. (2017). El aprendizaje gestual en dirección de orquesta mediante la observación (auto y hetero) [Tesis de doctorado, Universidad del País Vasco-Euskal Herriko Unibertsitatea]. <https://addi.ehu.es/handle/10810/27792>
- Elejabarrieta, F. & Iñiguez, L. (2008). Construcción de escalas de actitud, tipo Thurstone y Likert. *La Sociología en sus escenarios*, 17. <https://revistas.udea.edu.co/index.php/ceo/article/view/6820>
- Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema*, 17(2), 356-362. <https://www.psicothema.com/pi?pi=3112>
- Gamer, M., Lemon, J. & Fellows, I. (2012). Package irr: Various coefficients of interrater reliability and agreement. (R package version 0.84) [Computer software]. <http://CRAN.R-project.org/package=irr>
- Garaigordobil, M. & García de Galdeano, P. (2006). Empatía en niños de 10 a 12 años. *Psicothema*, 18, 180-186. <https://www.redalyc.org/pdf/727/72718203.pdf>
- García Sánchez, J., Aguilera Terrats, J. R. & Castillo Rosas, A. (2011). Guía técnica para la construcción de escalas de actitud. *Odiseo, Revista Electrónica de Pedagogía*, 8(16). <http://www.odiseo.com.mx/2011/8-16/garcia-aguilera-castillo-guia-construccion-escalas-actitud.html>
- Gillham, J., Adams-Deutsch, Z., Werner, J., Reivich, K., Coulter-Heindl, V., Linkins, M., Winder, B., Peterson, C., Park, N., Abenavoli, R., Contero, A. & Seligman, M. E. P. (2011). Character strengths predict subjective well-being during adolescence. *The Journal of Positive Psychology*, 6, 31-44. <https://doi.org/10.1080/17439760.2010.536773>
- Gross, J. & Ligges U. (2012). nortest: Tests for Normality (R package, version 1.0-2) [Computer software]. <http://CRAN.R-project.org/package=nortest>
- Hernández, J. A. (2019). ULLRToolbox [Computer Software]. <https://sites.google.com/site/ullrtoolbox/>
- Hodge, K. J. & Morgan, G. B. (2017). Stability of INFIT and OUTFIT compared to simulated estimates in applied setting. *Journal of Applied Measurement*, 18(4), 383-392. <https://euopepmc.org/article/med/29252207>
- Inderbitzen, H. M. & Foster, S. L. (1992). The Teenage Inventory of Social Skills: Development, reliability, and validity. *Psychological Assessment*, 4(4), 451-459. <https://psycnet.apa.org/buy/1993-11963-001>
- International Test Commission (2005). International guidelines on computer-based and internet delivered testing. <http://www.intest.com.org>
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (Second edition). www.InTestCom.org
- Jorgensen, T. D., Pomprasertmanit, S., Schoemann, A. M. & Rosseel Y. (2019) semTools: Useful tools for structural equation modeling (R package version 0.5-5) [Computer software].
- Lengemann Méndez, C. G. (2019) Estado del arte conductas prosociales en Colombia del año 2013 al 2018 investigaciones más relevantes [Trabajo de Grado, Pregrado, Escuela de Ciencias Sociales, Artes y Humanidades, Universidad Nacional Abierta y a Distancia, Colombia]. <https://repositorio.unad.edu.co/handle/10596/27976>
- Ley 1090: Código Deontológico y Bioético de Psicología. Congreso de la República de Colombia (2006). http://www.secretariassenado.gov.co/senado/basedoc/ley_1090_2006.html
- Ley Estatutaria 1581 Por la cual se dictan disposiciones generales para la protección de datos personales (2012). <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Likert, R. (1932) A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55. <https://psycnet.apa.org/record/1933-01885-001>
- Lippke, S., Nigg, C. R. & Maddock, J. E. (2007). The theory of planned behavior within the stages of the transtheoretical model: Latent structural modeling of stage-specific prediction patterns in physical activity. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 649-670. <https://doi.org/10.1080/10705510701575586>

- López de Cózar, E., Alonso Plá, F., Esteban Martínez, C., Calatayud Miñana, C. & Alamar Rocati, B. (2008). Diseño y validación de una escala para la medición de la conducta prosocial - antisocial en el ámbito vital y en el tráfico. Instituto Universitario de Tráfico y Seguridad Vial. Universidad de Valencia. https://www.uv.es/metras/docs/2008_encuestas_lopez_de_cozar_2_acta.pdf
- Mair, P. & Hatzinger, R. (2008). eRm: Extended Rasch Modeling (R package version 0.10-1) [Computer software].
- Martínez, E. R., García-Alandete, J., Sellés Nohales, P., Bernabé Valero, G. & Soucase Lozano, B. (2012). Análisis factorial confirmatorio de los principales modelos propuestos para el purpose-in-life test en una muestra de universitarios españoles. *Acta Colombiana de Psicología*, 15(1), 67-76. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-91552012000100007
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Meade, A. W. & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 14(4), 611-635. <https://doi.org/10.1080/10705510701575461>
- Morales Rodríguez, M. & Suárez Pérez, C. D. (2011). Construcción y validación de una escala para evaluar habilidades prosociales en adolescentes. Ponencia. XI Congreso Nacional de Investigación Educativa. Consejo Mexicano de Investigación Educativa, Delegación Coyoacán, México, Distrito Federal.
- Moreno, R., Martínez, R. J. & Muñoz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490-497. <http://www.redalyc.org/articulo.oa?id=72716324>
- Overgaauw, S., Rieffe, C., Broekhof, E., Crone, E. A. & Gürgölu, B. (2017) Assessing empathy across childhood and adolescence: Validation of the Empathy Questionnaire for children and adolescents (EmQue-CA). *Frontiers in Psychology*, 8, 870. <https://doi.org/10.3389/fpsyg.2017.00870>
- Pedrosa, I., Suárez-Álvarez, J. & García-Cueto, E. (2013). Evidencias sobre la validez de contenido: Avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3-18. <http://doi.org/10.5944/ap.10.2.11820>
- 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). Lawrence Erlbaum Associates, Inc.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21, 381-391. <https://www.jstor.org/stable/2489828>
- Picazo, C., López, M. & Gamboa, J. P. (2020). Análisis de la felicidad durante el tiempo libre: El papel de la conducta prosocial y material. *Diversitas*, 16(2), 233 - 244. <https://doi.org/10.15332/22563067.6291>
- Prieto, G. & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100. <https://psycnet.apa.org/record/2003-04505-016>
- R Core Team (2017). *R: A language and environment for statistical computing*. R Found. Stat. Comput. Vienna, Austria.
- Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. (R package version 2.2.5) [Computer software]. <https://CRAN.R-project.org/package=psych>.
- Ripoll-Núñez, K., Carrillo, S., Gómez, Y. & Villada, J. (2020). Predicting well-being and life satisfaction in Colombian adolescents: The role of emotion regulation, proactive coping, and prosocial behavior. *Psykhē*, 29(2), 1-16. <https://doi.org/10.7764/psykhe.29.1.1420>
- Romero Morales, C., Ventura Soto, S., Hervás Martínez, C., Rider Jiménez, I. & Martín-Palomo García, S. (2006). Herramienta Autor para la gestión de tests informatizados dentro del sistema AHAI. *IE comunicaciones: Revista Iberoamericana de Informática Educativa*. <https://redined.educacion.gob.es/xmlui/handle/11162/3670>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v48.i02/>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4), 1-100. <https://psycnet.apa.org/record/1972-04809-001>
- Sánchez-Queija, I., Oliva, A. & Parra, A. (2006). Empatía y conducta prosocial durante la adolescencia. *Revista de Psicología Social*, 21(3), 259-271. <https://doi.org/10.1174/021347406778538230>
- Scott, N. W., Fayes, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M. & Quality of Life Cross-Cultural Meta-Analysis Group. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62(3), 288-295. <https://doi.org/10.1016/j.jclinepi.2008.06.003>
- Silke, C., Brady, B., Boylan, C. & Dolan, P. (2018). Factors influencing the development of empathy and pro-social behaviour among adolescents: A systematic review. *Children and Youth Services Review*, 94, 421-436. <https://doi.org/10.1016/j.childyouth.2018.07.027>
- Silva, F. & Martorell, M. C. (1987). *BAS-3. Batería de Socialización*. Madrid: TEA.
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G. & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8, 1-11. <https://doi.org/10.1186/1471-2288-8-33>
- Ulber, J. & Tomasello, M. (2020). Young children's prosocial responses toward peers and adults in two social contexts. *Journal of Experimental Child Psychology*, 198, 104888. <https://doi.org/10.1016/j.jecp.2020.104888>
- Vallejo, P. M. (2013). *Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios*. Madrid: Universidad Pontificia Comillas. [http://www.rubenjoserodriguez.com.ar/wp-content/uploads/2015/04/An%23U00e1lisis Factorial Test - y Escalas Pedro Morales Vallejo.pdf](http://www.rubenjoserodriguez.com.ar/wp-content/uploads/2015/04/An%23U00e1lisis%20Factorial%20Test%20y%20Escalas%20Pedro%20Morales%20Vallejo.pdf)
- Warneken, F. & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, 100(3), 455-471. <https://doi.org/10.1348/000712608X379061>
- Wispe, L. G. (1972). Positive forms of social behaviour: An overview. *Journal of Social Issues*, 28(3), 1-19. <https://doi.org/10.1111/j.1540-4560.1972.tb00029.x>

Fecha de recepción: Agosto de 2022.

Fecha de aceptación: Octubre de 2023.