

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

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

Javier Villalba<sup>1</sup>, Juan Hernández<sup>1</sup>, Moisés Betancort<sup>1</sup> and Bertha Avendaño<sup>2</sup>

<sup>1</sup> School of Doctoral and Postgraduate Studies of the University of La Laguna

<sup>2</sup> Doctorate in Psychology from the Catholic University of Colombia

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 behaviour. These evidences support the use of the questionnaire in Colombia for research purposes.

*Keywords:* Validity, Internal Structure and Prosocial Behaviour

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

In the last 15 years, there has been an increased interest in studying prosocial behaviour, because it is related to positive aspects of human beings, specifically social and personal well-being, which contributes to the formation of solidarity bonds and moderates aggressive behaviours (Auné et al., 2014). An association between prosocial behaviour and life satisfaction has also been found to predict subjective well-being, especially in adolescents (Gillham et al., 2011; Ripoll-Núñez et al., 2020). Initially, prosocial behaviour was defined as an antonym of antisocial behaviour (Wispé, 1972), but over time this behaviour has been redefined

---

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>

The article is part of the thesis for the Doctoral Degree in Psychology at the University of La Laguna. There is no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Javier Andrés Villalba Garzón, Calle 25 #68B-47 Interior 3 apartamento 702, Bogotá D. C., Colombia. Email: [alu0101298733@ull.edu.es](mailto:alu0101298733@ull.edu.es)

to such an extent that, to date, it has been possible to identify at least six different typologies (Auné et al., 2014). In general terms prosocial behaviour is understood as the attempt to satisfy the physical and emotional support of the other persons, as opposed to altruism, which is defined as the result of putting others' needs before one's own (Auné et al., 2014). A greater number of research studies related to the study of prosocial behaviour are identified in childhood compared to adulthood, which can be explained by the fact that from a very early age a wide variety of positive social behaviours begin to be observed, which are oriented towards interacting collaboratively with others (Ulber & Tomasello, 2020).

The development of psychometric instruments has been fundamental to progress in the study of prosocial behaviour, because it allows us to obtain highly reliable and valid measures, properties that are achieved with the minimum required conditions, which are established in the quality standards for psychological and educational tests published by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education (AERA, APA & NCME, 2014). The five sources of validity evidence in these standards are based on: a) the relations of item content to the frame of reference of the attribute measured or test content; b) the response processes involved; c) the internal structure of the scale; d) the relations of scores to other variables; and e) the consequences of testing.

Twelve self-administered tests have been identified that measure prosocial behaviour globally or as a component in a subscale. These tests are: a) the Prosocial Behaviour Scale (PB) by Caprara and Pastorelli (1993); b) the Prosocialness Scale for Adults (PSA) by Caprara et al. (2005); c) the Prosocial Tendencies Measure (PTM) by Carlo and Randall (2002); d) the Teenage Inventory of Social Skills (TISS) by Inderbitzen and Foster (1992); e) the Prosocial Skills Scale for Adolescents (EHP-A) by Morales Rodríguez and Suárez Pérez (2011); f) the Prosocial Behaviour Questionnaire by Sánchez-Queija et al. (2006); g) the scale for the Measurement of Prosocial-Antisocial Behavior in the vital sphere and Traffic by López de Cózar et al. (2008); h) the Battery of Socialization for Teenagers (BAS-3) by Silva and Martorell (1987); i) the Prosocial Personality Battery (PSB) by Penner et al. (1995); j) the Empathy Questionnaire for Children and Adolescents (EmQue-CA) by Overgaauw et al. (2017); k) the Aggressive and Pro-social Behaviour Survey (COPRAG) by Agudelo et al. (2000) and l) the Prosocial Behaviour Scale (PBS) by Auné & Attorresi (2017).

### **Development of the Prosocial Behaviour Scale (PBS)**

Of the above tests, the PBS by Auné and Attorresi (2017) is the only test designed to measure prosocial behaviour in adults in a Latin American context and for which a wide variety of evidence is available, allowing us to identify its development and psychometric quality. On the other hand, the PBS has a greater number of studies carried out with psychometric methods from Item Response Theory (IRT), which are more robust in contrast to methods from Classical Test Theory (CTT).

To develop this questionnaire 114 items were developed, after collecting evidence related to the content of the items with a review by a group of experts, 64 items were selected for a second review and were subjected to a pilot test with 56 participants who were students from various professional careers at the University of Buenos Aires (Auné & Abal, 2016; Auné, 2018). With the results of the pilot test, the first version of the PBS was obtained, consisting of 29 items, with which evidence of internal structure validity was collected with an Exploratory Factor Analysis (EFA), a Confirmatory Factor Analysis (CFA) and a Differential Item Functioning (DIF) analysis. Prior to conducting the EFA, 761 participants were randomly selected from a sample of 1383 and reported a Kaiser-Meyer-Olkin (*KMO*) sample adequacy measure of 0.90 and in Bartlett's test of sphericity an  $\chi^2=7441$  with a  $p=0.00001$  (Auné, 2018). In the EFA two factors were identified that explained 44.67% of the variance, the first factor was called "Helping" made up of 9 items with an internal consistency reliability  $\alpha=0.85$  and the second factor "Comforting the other" with 7 items with an internal consistency reliability  $\alpha=0.77$  (Auné, 2018).

To conduct the CFA, 662 of the 1383 participants were randomly selected and the goodness-of-fit of four models based on the two-factor structure with 16 items was assessed (Auné, 2018). The weighted least squares robust least squares estimator (*WLSMV*), the comparative fit index (*CFI*), the Tucker-Lewis index (*TLI*) and the root mean squared error of approximation (*RMSEA*) with a confidence interval (*CI*) of 90% were used. Auné (2018) reported  $CFI>0.610$ ,  $TLI>0.550$  and  $RMSEA (90\% CI)>0.0720$ ; the related two-factor model had the best goodness-of-fit with  $CFI=0.921$ ,  $TLI=0.908$  and  $RMSEA (90\% CI)=0.070 (0.063-0.077)$ .

Auné (2018) reported four items with DIF between women and men that he detected with the  $\chi^2$  of the Mantel-Haenszel test with p-value, the ratio of common ratios (MH), the Standardised Difference of Proportions (SDP) and the Mantel-Haenszel Delta-DIF (MH D-DIF). In addition to the EFA, CFA and DIF

analyses, the items of the two dimensions were calibrated with the three-parameter logistic model for graded responses (MRG) and partial credit, in order to identify at which levels of the latent traits more information was obtained in the parameter estimation. Auné (2018) found that the Helping factor was a more informative dimension at high levels and with acceptable accuracy over a very wide range of the trait allowing most people to be located. Regarding the Comforting the Other factor, a more informative dimension was obtained at low levels of the trait and with measurement errors that grew substantially at high levels of the trait.

With these results, the first definitive version of the PBS was obtained (Auné, 2018; Auné & Attorresi, 2017) with 15 items that make up two dimensions that allow measuring prosocial behaviour in the adult population of Argentina, with a satisfactory internal consistency ( $\alpha=0.84$ ) obtained in a sample of 667 students from the University of Buenos Aires. This study confirms the structure of Helping and Comforting the Other with an AFC with  $GFI=0.99$  and  $RMSR=0.04$ . Additionally, Auné and Attorresi (2017) found with an AFA that these dimensions explain 50.61% of the total variance. On the other hand, the Helping dimension showed better psychometric goodness compared to the Comforting the other dimension and for this reason Auné et al. (2020) propose to use the first seven items of the Helping subscale as an independent questionnaire called the Helping Scale (HS), based on the analysis with a two-parameter Graded Response Model (GRM), with which they detected that item eight presented non-uniform DIF between women and men.

Apart from the studies with the Argentinean population, Canales Reyes (2020) reports results of internal consistency, evidence of internal structure validity and the relationship with other variables in the Peruvian population. With a sample of 309 participants, including students and workers, an  $\alpha=0.86$  was obtained for the Helping dimension and  $\alpha=0.79$  for Comforting the other. With McDonald's Omega the results were  $\omega=0.87$  and  $\omega=0.86$  respectively. The two-dimensional oblique non-hierarchical oblique model with 11 items (2, 3, 4, 5, 6, 7, 10, 11, 12, 13 and 14) was confirmed in the CFA with  $\chi^2/df=2.97$ ,  $CFI=0.94$ ,  $TLI=0.92$ ,  $SRMR=0.05$  and  $RMSEA=0.07$ . Significant correlations were found between the Empathic Behaviour Scale scores and the Helping dimension  $r=0.85$  and the Comforting Other dimension  $r=0.49$ . These results are similar and consistent with those obtained in the studies by Auné (2018), Auné and Attorresi (2017) and Auné et al. (2020).

### **Study of Prosocial Behaviour in Colombia**

In Colombia, Lengemann Méndez (2019) conducted a review of research between 2013 and 2018, in the databases Redalyc, Dialnet, EBSCOhost, APA PsycNET, APA PsycINFO and APA PsycARTICLES, with the aim of describing the main studies related to prosocial behaviour in Colombia. From this study it is highlighted that there is a higher percentage of quantitative research with 35%, most of which was conducted with samples of adolescent population and the test that has been used the most is the TISS, which is also designed for adolescents. The only test found that was designed and constructed for the Colombian population was the (COPRAG) by Agudelo et al. (2000), which is applied to minors between the ages of three and eleven. One of the most interesting conclusions of the work of Lengemann Méndez (2019) is related to the statement that in Colombia the study of prosocial behaviour is at an initial stage and that it needs to mature, both in the development of large-scale studies of a psychometric nature and in the comparison with different variables, which allow us to accurately recognize prosocial behaviours in that country.

### **Adaptation of the PBS in Colombia**

Although in Colombia there is a clear tendency to study prosocial behaviour in minors, it is also considered important to study this behaviour in adults because it is a good predictor of individual and collective well-being (Auné et al., 2014). As there are no scales adapted for Colombian adults, it is considered a good option to adapt the PBS (Auné, 2018; Auné & Attorresi, 2017), due to the positive results reported in psychometric studies conducted with Argentinean and Peruvian populations.

In the search carried out, no empirical reports were found on the functioning of the PBS in the Colombian population, which is why this study proposed to begin with the evaluation of the reliability of this questionnaire, using internal consistency methods and, additionally, to collect evidence of validity related to the content of the items and the internal structure of this scale. It is proposed to carry out this analysis in order to have a preliminary adaptation and facilitate its use in this population.

## Method

### Participants

A convenience sample of 344 literate adults was drawn, of whom 249 (72.38%) were female, 94 (27.33%) were male and one person (0.29%) identified with another gender. The age of the participants ranged from 18 to 77 years ( $M=37.59$ ,  $SD=12.54$ ). Regarding the highest level of education attained, 0.58% reported primary education, 7.85% high school, 6.10% technical, 4.36% technological, 34.17% professional, 22.67% specialist, 21.80% master's degree and 1.45% doctorate. At the time of filling in the scale, 81.69% had a job, 8.43% were students, 3.78% were unemployed, 2.62% were engaged in housework, 2.33% had retired and 1.16% were studying and working at the same time. 48.55% of the participants reported being single, 30.23% were married, 15.70% were in union, 4.07% were separated and 1.45% were widowed. Applications were obtained from people born in 12 locations in Colombia, with the highest concentration in the city of Bogotá with a participation of 84.51%.

### Instruments

**Prosocial Behaviour Scale (PBS) by Auné & Attorresi (2017).** It consists of the dimensions of Helping and Comforting others, with a total of 15 Likert-type response items with six options (*Never, Almost Never, Sometimes, Frequently, Almost Always and Always*). The Helping dimension includes nine behaviours associated with solidarity and which are also presented as a separate scale called the "Helping Scale (HS)" (Auné et al., 2020), while the Comforting Another dimension describes six behaviours that relate to the emotional support that can be given to others.

**Sociodemographic data questionnaire.** An instrument with six questions was designed to collect socio-demographic information from the participants. Sex, marital status and education were addressed with a closed question; age (years), name of degree obtained and occupation were addressed with an open question.

### Procedure

In accordance with Law 1090 (2006), which establishes the code of ethics for psychology in Colombia, permission was obtained from the original authors to use and adapt the scale in the Colombian population. During all phases of the research, the authors were duly cited to guarantee intellectual property rights. The questionnaire was only administered to those who agreed to participate voluntarily and explicitly expressed this in the informed consent form. At the end of the research, reports with individual results were sent to the e-mail addresses of the participants who requested them. Participation was anonymous and total confidentiality was maintained about the e-mail addresses and socio-demographic data recorded, in compliance with Statutory Law 1581 (2012), which establishes general provisions for the protection of personal data in Colombia.

The original version of the questionnaire was written in Spanish, so translation and re-translation was not necessary. Four expert judges reviewed the questionnaire twice, the first time the original version was reviewed and the second time they validated that the adjustments suggested in the first assessment were carried out correctly. In both assessments, the relevance of the content and the wording of the items were reviewed qualitatively. At the time of the assessment, all four had more than ten years of experience in reviewing, adapting and constructing psychometric tests. Three of the experts had a doctoral degree and one had a master's degree.

Based on the International Test Commission's (ITC, 2005) international guidelines on computer- and internet-based testing, the PBS was uploaded to an online Google Forms that could be completed from a computer or mobile device with internet access, with any browser, and there were no operating system, software or hardware restrictions. As a typical self-report questionnaire, it did not require direct supervision and participants were free to choose the device and amount of time to complete the questionnaire.

The application was carried out during the confinement generated by the COVID-19 pandemic and two strategies were implemented: a) the questionnaire was distributed by email and WhatsApp, based on a list of contacts of Colombian origin available to the authors, and b) it was published on Facebook and LinkedIn with the condition of limiting access only to Colombian people. Data collection lasted four and a half months and was stopped when the sample size exceeded 200, to carry out the analyses using the Rasch model (Smith et al., 2008) and the CFA (Vallejo, 2013). A greater participation of the non-university Colombian population was sought to reduce a probable selection bias, and applications made by people who reported being minors were excluded, as they did not belong to the target population.

## Analysis Plan

Evidence of content-based validity of the items (AERA, APA & NCME, 2014) was collected with expert ratings with two categories "Approved" and "Requires adjustment". The reliability of the four experts' ratings was estimated with direct agreement and Fleiss' Kappa coefficient (Davies & Fleiss, 1982). The agreement between pairs of experts was also checked directly and with Cohen's Kappa coefficient (1968).

Psychometric analyses were carried out with the CTT and supplemented with estimates of the Rasch model of IRT for polytomous responses (Samejima, 1969). Item discrimination was obtained with the CTT, with item-total correlation and internal consistency reliability was also estimated with Cronbach's Alpha (1951) and McDonald's Omega (1999) coefficients. To obtain the fit of the Rasch model, mean squared residuals (*MNSQ*) were estimated for the internal (*Infit*) and external (*Outfit*) fit indices, together with the hypothesis test  $\chi^2$  to compare the empirical distributions with that of the model.

To collect validity evidence based on the internal structure of the scale (AERA, APA & NCME, 2014), CFAs were conducted on three models (unidimensional of prosocial behaviour; two-dimensional integrated of the Helping and Comforting the other subscales; and two-dimensional with the two subscales independently). The model with the best goodness-of-fit was subjected to a progressive factorial invariance assessment (Elosua, 2005), with configural, metric, scalar and residual restrictions with respect to gender. Additionally, the items in this model were calibrated with the Rasch model to identify the behaviours that characterise people with higher levels of prosocial behaviour.

Data processing was performed using the R programming language (R Core Team, 2017), with the R Studio interface and the ULLRToolbox add-on (Hernández, 2019). The libraries psych (Revelle, 2022), irr (Gamer et al., 2012), nortest (Gross & Ligges, 2012), eRm (Mair & Hatzinger, 2008), lavaan (Rosseel, 2012) and seemTools (Jorgensen et al., 2019) were used.

## Results

### Content review

The experts' assessment had an overall agreement of 60% with a Fleiss Kappa  $k=0.317$  and a  $p=0.003$ . The highest agreement was obtained between experts 2 and 4 with 93.33% agreement and a Cohen's Kappa  $k=0.860$ , followed by the agreements between experts 1 and 3 and 3 and 4 with 80% agreement and Cohen's Kappas  $k=0.000$  and  $0.550$  respectively. Experts 2 and 3 had 73.33% agreement with a Cohen's Kappa  $k=0.330$ , followed by experts 1 and 2 with 66.67% agreement with a Cohen's Kappa  $k=0.000$  and the lowest agreement achieved was between experts 1 and 4 with 60.00% agreement and a Cohen's Kappa  $k=0.000$ . When reviewing the individual ratings, it was identified that expert 1 was the laxest, because he passed all items without making any observations. The global agreement was re-estimated excluding expert 1 and a direct agreement of 73.33% was obtained with a Fleiss Kappa  $k=0.585$  and a  $p=0.000$ .

Nine items (1, 3, 4, 8, 10, 11, 12, 13 and 14) were approved by the four experts and suggestions were made to the remaining items. The suggestions were contrasted with the psychometric results reported by Auné (2018) and Auné and Attorresi (2017), specifically with the probability curves of the response options and the factor loadings. It was decided to modify items 2, 5, 6, 7 and 9 based on the experts' suggestions and to remove item 15 "I advise acquaintances about work" based on the empirical reports of the original study and the relationship of its content to the respective dimension. The experts reviewed the PBS again, approved the adjustments and the decision to remove item 15 for the adaptation of the questionnaire to the Colombian population.

## Reliability and Item Analysis

Regarding reliability, the overall scores had an  $\alpha=0.87$ , in the Helping dimension it was  $\alpha=0.85$  and in the Comforting the other dimension it was  $\alpha=0.77$ . These results show a high intercorrelation of the items, which allows us to infer that the variations of the scores in the global scale and the subscales presented a low non-systematic error. According to the guidelines mentioned by Peterson (1994), these values correspond to a moderate-high level of reliability, which is above the mean and median generally obtained with Likert scales associated with behavioural reports, but which are below the values suggested for applied research.

In the estimation of reliability with the Omega coefficient, hierarchical Omegas  $\omega=0.59$  (overall),  $\omega=0.74$  (Help) and  $\omega=0.58$  (Comfort the other); asymptotic Omegas  $\omega=0.66$  (overall),  $\omega=0.83$  (Help) and  $\omega=0.67$  (Comfort the other); and total Omegas  $\omega=0.90$  (overall),  $\omega=0.89$  (Help) and  $\omega=0.86$  (Comfort the other) were obtained. In contrast to the values obtained with Cronbach's Alpha, the results with McDonald's Omega present large differences that show the non-compliance with the tau-equivalence principle required to use Cronbach's Alpha. In this sense, the Omega coefficient is the most appropriate for analyzing the internal consistency of the PBS 14 and its two dimensions.

With the Hierarchical Omega it could be identified that the Helping dimension presented the highest reliability associated with a general factor and this can increase up to a  $\omega=0.83$  (asymptotic Omega), when the items tend to an infinite increase (De Reizábal, 2017). The internal consistency of the 14 PBS items was affected by the existence of the two dimensions and by the moderate intercorrelation of the items in the Comforting Other dimension. Despite these findings, a total variance of 0.90 associated with the set of scale items was identified with the Omega total.

On the other hand, the item analysis was done by subscale with the CTT and was complemented with estimates of the Rasch model, with which calculations could not be made for item 14 because no participant answered "Never" and the model could not successfully converge with this *outlier* response (0%). Table 1 describes the results obtained in the item analysis and it can be identified that all items presented high correlations and did not contribute significant errors to the reliability of the respective subscales, because the discrimination estimated with the Item-Total Correlation had values greater than 0.20 (Likert, 1932) and the  $\alpha$  does not increase substantially if any of the items is eliminated.

For the analyses with the Rasch model, the criteria proposed by Hodge and Morgan (2017) were applied and it was identified that only three items (6, 10 and 11) had Outfit and Infit indices outside the accepted values for goodness of fit (between 0.70 and 1.30). This means that the Rasch model had a low prediction of the responses given by the participants to these three items. Item 12 obtained a slightly higher value in the *Infit* index, in this case it means that the Rasch model lost predictive power in the responses of the participants with parameters  $\theta$  close to the parameter  $b$  of this item. Regarding the  $\chi^2$  and its p-value, it was found that the distributions of the probabilities of items 1 and 12 presented statistically significant differences in contrast to the distribution established for the Rasch model.

In contrast to both analyses, with the CTT there were no findings for the qualitative review of the items, however, with the Rasch model the content of four items (6, 10, 11 and 12) had to be reviewed because there was no model fit in one or two of the indices and this indicates a low prediction of the responses on these items.

**Table 1**  
*Item analysis*

| Item   | Dimension             | CTT                 |                                 | Rasch model       |         |                  |                   |
|--|-----------------------|---------------------|---------------------------------|-------------------|---------|------------------|-------------------|
|  |                       | Corr. item subscale | $\alpha$ if the item is deleted | $\chi^2$          | $p$     | <i>Infit MSQ</i> | <i>MSQ Outfit</i> |
| 1. I relegate my personal benefit to help others.  | Help                  | 0.39                | 0.86                            | 402.070           | 0.015*  | 1.095            | 1.169             |
| 2. In my spare time I do volunteer activities.   | Help                  | 0.66                | 0.83                            | 343.159           | 0.487   | 0.977            | 0.998             |
| 3. When I feel that someone is wrong, I show them that I understand.                                     | Comforting each other | 0.61                | 0.73                            | 277.180           | 0.996   | 0.771            | 0.808             |
| 4. I act as a baton for others.  | Comforting each other | 0.63                | 0.73                            | 259.082           | 1.000   | 0.758            | 0.755             |
| 5. I feel other people's pain as my own.   | Comforting each other | 0.59                | 0.74                            | 285.771           | 0.988   | 0.843            | 0.833             |
| 6. I am used to committing myself to charitable causes.  | Help                  | 0.81                | 0.82                            | 218.196           | 1.000   | <b>0.637</b>     | <b>0.634</b>      |
| 7. I dedicate an important part of my life to actions that improve the world we live in.                 | Help                  | 0.63                | 0.84                            | 326.882           | 0.726   | 0.929            | 0.950             |
| 8. I care for the well-being of any individual, group or community.                                      | Help                  | 0.72                | 0.83                            | 247.919           | 1.000   | 0.729            | 0.721             |
| 9. I live with the bare necessities and distribute everything else.                                      | Help                  | 0.58                | 0.84                            | 335.466           | 0.604   | 0.997            | 0.975             |
| 10. I put myself in the place of the other.  | Comforting each other | 0.64                | 0.72                            | 223.691           | 1.000   | <b>0.646</b>     | <b>0.652</b>      |
| 11. I participate in solidarity activities.  | Help                  | 0.83                | 0.81                            | 183.726           | 1.000   | <b>0.529</b>     | <b>0.534</b>      |
| 12. I donate to charities.   | Help                  | 0.55                | 0.85                            | 431.791           | 0.001** | <b>1.322</b>     | 1.255             |
| 13. If a person tells me about a conflict, I try to make them understand the other side's point of view. | Comforting each other | 0.53                | 0.76                            | 374.350           | 0.110   | 1.184            | 1.033             |
| 14. I try to "raise" the self-esteem of my friends.  | Help                  | 0.59                | 0.74                            | Does not converge |         |                  |                   |

Note. \* $p < 0.05$ , \*\* $p < 0.01$ .

### Evidence based on internal structure

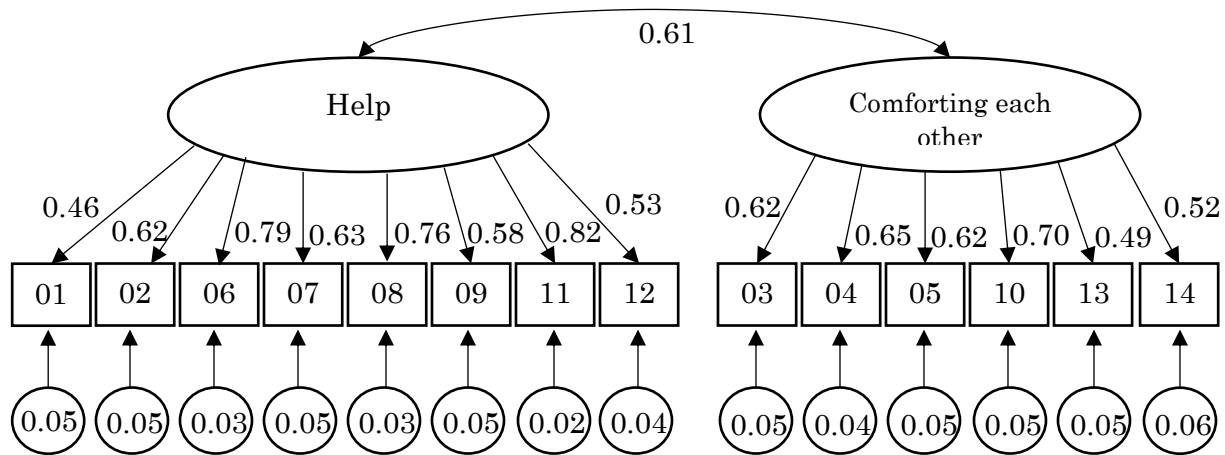
**Factor structure.** Three CFA models were tested: a) a unidimensional model that considered prosocial behaviour in general; b) a two-dimensional one that consisted of two related dimensions Helping and Comforting the other and c) a two-dimensional one consisting of two separate dimensions Helping and Comforting the other. Table 2 lists the goodness-of-fit indices of the three models analyzed, of which the model with the best goodness-of-fit was the one with two related dimensions, with  $p < 0.05$  and normal goodness-of-fit indices. 0.05 and normalized (*NFI*), non-normalized (*NNFI*), comparative (*CFI*) and Tucker-Lewis (*TLI*) fit indices greater than 0.95 and the root mean square error of approximation (RMSEA) less than 0.05, in accordance with the criteria of Browne and Cudeck (1993), Bentler and Bonnet (1980) and Martínez et al. (2012).

**Table 2**  
*AFC models*

| Goodness of fit | Model 1         | Model 2                    | Model 3       |                       |
|-----------------|-----------------|----------------------------|---------------|-----------------------|
|                 | One-dimensional | Two-dimensional integrated | Help          | Comforting each other |
| $\chi^2$        | 269.976         | 130.176                    | 21.034        | 35.831                |
| $df$            | 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                 |
| $IFC$           | 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]        |

Figure 1 illustrates the integrated two-dimensional model and the items had standardized regression weights above 0.40 in the respective dimensions and errors less than or equal to 0.06.

**Figure 1**  
*Factor Structure Model 2*



Model 2 was assessed to be invariant with respect to the "sex" variable, to verify that the factor structure did not vary between women and men and to rule out possible bias. Table 3 shows the progressive invariance analysis performed on the two-dimensional model and this model was invariant between women and men, in the basic configuration of the model (configural), in the equivalence of the factor weights (metric) and intercepts (scalar) and in the residual variances and covariances (residual). This interpretation is because in the  $\chi^2$  and  $df$  tests,  $p$ -values  $>0.50$  with  $\Delta TLI < 0.050$  were obtained (Lippke et al., 2007), together with  $\Delta CFI \geq -0.010$  and  $\Delta RMSEA < 0.010$  (Cheung & Rensvold, 2002).



**Table 3**  
*Factorial Invariance Model 2 with the variable Sex*

| Invariance | <i>df</i> | $\chi^2$ | $\Delta\chi^2$ | $\Delta df$ | <i>P</i> | Contrast              | $\Delta TLI$ | $\Delta CFI$ | $\Delta RMSEA$ |
|------------|-----------|----------|----------------|-------------|----------|-----------------------|--------------|--------------|----------------|
| Configural | 152       | 164,11   |                |             |          |                       |              |              |                |
| Metric     | 164       | 184,66   | 10,176         | 12          | 0,601    | Metric-<br>Configural | 0,039        | 0,027        | -0,012         |
| Climb      | 176       | 190,84   | 10,684         | 12          | 0,556    | Scalar-<br>Metric     | 0,007        | 0,002        | -0,002         |
| Residual   | 190       | 197,27   | 10,473         | 14          | 0,727    | Residual-<br>Escalar  | 0,012        | 0,007        | -0,004         |

**Item calibration.** In order to locate and organize the prosocial behaviours described in the items in an invariant scale, we proceeded to estimate the *b* and  $\theta$  parameters with the Rasch model. Table 4 lists the independently calculated *b-parameters* for the two dimensions Helping and Comforting the other, to check the assumption of unidimensionality required in this model. The results for the Helping dimension had a range of 17.328 logits located between -0.787 and 16.541, while for the Comforting the other dimension the range was 12.393 which was located between -0.681 and 11.712 logits. The ascending order established with the Rasch model for the behaviours of the Helping dimension items was 1, 7, 8, 11, 6, 9, 12 and 2, indicating that item 1 describes the behaviour with the lowest level of helping and item 2 the behaviour with the highest level. In the Comforting dimension the order of the items was 3, 10, 13, 5 and 4.

In the Helping dimension, the parameters with the lowest standard errors were obtained in the estimates made for category 1 (*Never*) and 2 (*Almost Never*) with values less than 0.33, a criterion used to consider an item to be calibrated (Romero Morales et al., 2006). In the Comforting the other dimension, this was achieved only in category 1 and there were standard errors greater than 1.00, indicating that the estimates of the items in this dimension were less accurate compared to the Helping dimension.

**Table 4**  
*Parameters b*

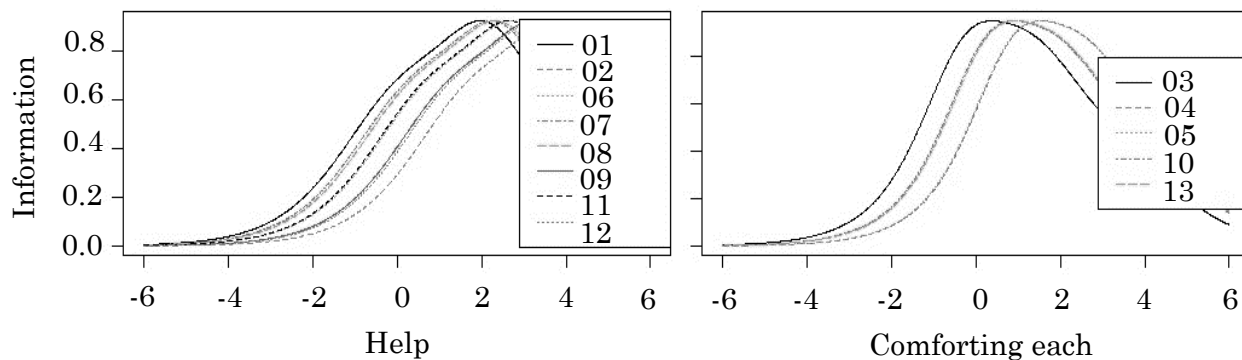
| Item | Dimension             | Parameter <i>b</i> [Standard error]. |                 |                |                 |                  |
|------|-----------------------|--------------------------------------|-----------------|----------------|-----------------|------------------|
|      |                       | <i>b1</i>                            | <i>b2</i>       | <i>b3</i>      | <i>b4</i>       | <i>b5</i>        |
| 1.   | Help                  | -0.787 [0.060]                       | -0.358 [0.161]  | 1.472 [0.260]  | 3.939 [0.368]*  | 7.948 [0.493]*   |
| 2.   | Help                  | 0.931 [0.065]                        | 3.080 [0.191]   | 6.628 [0.329]  | 10.813 [0.486]* | 16.541 [0.667]*  |
| 3.   | Comforting each other | -0.681 [0.065]                       | -0.893 [0.349]* | 0.218 [0.607]* | 2.137 [0.878]*  | 5.866 [1.157]**  |
| 4.   | Comforting each other | 0.488 [0.057]                        | 1.446 [0.353]*  | 3.726 [0.625]* | 6.814 [0.919]*  | 11.712 [1.229]** |
| 5.   | Comforting each other | 0.467 [0.057]                        | 1.403 [0.352]*  | 3.661 [0.624]* | 6.728 [0.917]*  | 11.605 [1.226]** |
| 6.   | Help                  | -0.063 [0.056]                       | 1.090 [0.162]   | 3.644 [0.272]  | 6.835 [0.399]*  | 11.568 [0.550]*  |
| 7.   | Help                  | -0.521 [0.057]                       | 0.175 [0.159]   | 2.271 [0.261]  | 5.005 [0.375]*  | 9.281 [0.511]*   |
| 8.   | Help                  | -0.433 [0.057]                       | 0.350 [0.159]   | 2.534 [0.263]  | 5.354 [0.379]*  | 9.718 [0.517]*   |
| 9.   | Help                  | 0.448 [0.059]                        | 2.113 [0.173]   | 5.178 [0.297]  | 8.881 [0.439]*  | 14.126 [0.605]*  |
| 10.  | Comforting each other | -0.155 [0.057]                       | 0.158 [0.345]*  | 1.794 [0.607]* | 4.239 [0.884]*  | 8.493 [1.175]**  |
| 11.  | Help                  | -0.106 [0.056]                       | 1.005 [0.161]   | 3.516 [0.271]  | 6.664 [0.396]*  | 11.355 [0.546]*  |
| 12.  | Help                  | 0.532 [0.060]                        | 2.280 [0.176]   | 5.429 [0.302]  | 9.215 [0.446]*  | 14.543 [0.615]*  |
| 13.  | Comforting each other | -0.119 [0.057]                       | 0.231 [0.345]*  | 1.904 [0.607]* | 4.385 [0.885]*  | 8.676 [1.177]**  |
| 14.  | Comforting each other | Does not converge                    |                 |                |                 |                  |

Note. \*Standard error >0.33; \*\*Standard error >1.00.

The Test Information Function illustrates the inverse function of the measurement errors and is obtained from the sum of the Item Information Functions (Prieto & Delgado, 2003). Figure 2 shows a segmented plot of this function in both dimensions, which allowed us to identify that the highest accuracy in the Helping dimension was obtained between 2.00 and 4.00 logits, because the lowest standard error in the estimation of the  $b$  parameters was located in that position. In the Comforting dimension, the highest precision was between 0.00 and 2.00 logits. These results show that the Helping dimension covered the largest measurement range and with the highest precision, compared to the Comforting the other dimension.

In the estimation of the parameters  $\theta$  of the Helping dimension, it was distributed in a range of 6.68 logits (between -1.40 and 5.29), with an  $M=2.03$  and a  $SD=1.06$ . The standard errors of the estimates in this dimension had an  $M=0.40$  with a  $SD=0.04$ , a minimum value of 0.38 and a maximum of 0.75. Regarding the parameters  $\theta$  of the Comforting the other dimension, the range was 5.67 logits (between 0.28 and 5.95), an  $M=2.81$  and a  $SD=1.09$ . The standard errors in this dimension had an  $M=0.55$  with a  $SD=0.11$ , a minimum value of 0.46 and a maximum of 1.08. For one person the parameter  $\theta$  could not be estimated, because he had an *outlier* behaviour (100%) characterized by scoring the highest category in all items of the dimension.

**Figure 2**  
*Test Information Functions*



## Discussion

The main objective of this study was to collect evidence of the content validity of the items and the internal structure of the scale, in order to make a preliminary adaptation of the PBS of Auné & Attorresi (2017) and project its use with the Colombian population. In this process, the guidelines for translation and adaptation of tests of the ITC (2017) were followed, although it was not necessary to translate the items because the original scale was constructed in Spanish and was proposed within a Latin American context, which contemplates similar contextual aspects between the population with which the scale was designed and validated (Argentina) and the population to be adapted (Colombia). This condition made it possible to minimise the influence of cultural and linguistic differences irrelevant to the interpretation of the scores obtained in the population to be adapted.

Evidence of the content validity of the items was obtained with an expert assessment, which led to adjusting the wording of five items (2, 5, 6, 7 and 9) to clarify and delimit the behaviours described, because the experts considered that it did not comply with the indications suggested by Moreno et al. (2004) and Elejabarrieta and Iñiguez (2008), which indicate that: a) the original wording lacked specificity in the description of the described behaviour; b) it included more than one behaviour in the same item; or c) it allowed a wide margin of interpretation, which is not desirable for closed-response items.

In comparison with the adaptation carried out by Canales Reyes (2020) with a Peruvian population, fewer items were adjusted and only item 9 was modified.

In the original version the wording of the item is "I keep just what is necessary to live and share everything else", in the Peruvian adaptation it was adjusted to "I share what I have with others" and in this study the item was reworded as "I live with what is necessary and share everything else". In the adjustment made in this study to item 9, we sought to maintain the literality of the original study and only modified the length of the statement with a shorter and more concise approach. The psychometric results show that this item did not present problems of interpretation, because no findings were found with the CTT methods and the Rasch model.

Quantitative information was included in the response options for all items to facilitate understanding and differentiation of the frequencies described in the reading of the options (*Never - 0% of the time; Almost never - between 1% and 19% of the time; Sometimes - between 20% and 59% of the time; Often - between 60% and 79% of the time; Almost always - between 80% and 99% of the time; and Always - 100% of the time*); *Sometimes-between 20% and 59% of the time; Frequently-between 60% and 79% of the time; Almost always-between 80% and 99% of the time; and Always-100% of the time*), in order to reduce subjectivity when reading the statements and relating them to a frequency.

The elimination of item 15 "I advise acquaintances about work" was based on the psychometric reports of the original studies and its content with respect to the dimension it measures. In the psychometric results reported by Auné (2018) and Auné and Attorresi (2017), it is identified that the probability curves showed differences only in categories 3 and 6, which shows that it does not function as a graded response of six options. Additionally, in the EFA conducted by Auné (2018) and Auné and Attorresi (2017), this item presented the lowest factor loadings (0.41 in the configuration matrix and 0.44 in the structure matrix). In addition to this, and related to the content of the item, the verb "to advise" is not synonymous with "to comfort" and is more related to verbal behaviours of guidance or opinion given to another person.

For the psychometric analyses, despite non-probability sampling, the sample was sufficiently varied in terms of age, highest level of education attained and occupation, and only 9.59% were university students. Although a wide variability of different population groups in Colombia was achieved, it is not recommended to generalize the results to regions other than the capital city because the sample was significantly concentrated in people who were born in this location. In accordance with the recommendations of Vallejo (2013), a sample size was obtained that allowed the application of the statistical methods proposed for data analysis, especially the CFA which requires a ratio equal to or greater than 10 data for each item. However, Meade and Bauer (2007) recommend having at least a sample of 200 per comparison group in factorial invariance analyses, and in this study the sample size for men was insufficient because only 94 participated, which affected the calculation of the Gamma coefficient.

The results of the analysis with the CTT methods show that all items discriminated and substantially equal reliabilities were found between the original scale applied to a sample of the Argentine population (Auné & Attorresi, 2017) and the 14-item version applied to the Colombian sample. In both dimensions the same reliability was obtained, an  $\alpha=0.85$  in the Helping dimension and an  $\alpha=0.77$  in Comforting the other. The picture was similar in contrast to the adaptation with the Peruvian population, because the differences were minimal with an  $\alpha=0.86$  in the Helping dimension and an  $\alpha=0.79$  in Comforting the other and  $\omega=0.87$  and  $\omega=0.83$  respectively.

Despite this similarity in the reliability results with Cronbach's alpha, it is suggested that the reliability obtained with McDonald's omega coefficient be used as a basis because the items did not present tau-equivalent measures. These analyses allow us to identify that the Helping dimension presented greater reliability and discrimination in the items compared to the Comforting the other dimension.

In the analyses with the Rasch model, the assumption of unidimensionality in the structure confirmed with the AFC was met and this led to separate parameter estimates for both dimensions. In the Help dimension, in four of the nine items (6, 11, 12 and 14) the model did not fit. Although there were findings in these items, the contents were consistent with the conceptualisation of the Helping dimension and no aspects were identified that could explain these psychometric results. As no arguments were found to suggest the elimination or modification of these items, they are maintained in the version that will be used to measure prosocial behaviour in the Colombian population.

In the Comforting the other dimension, the model did not fit item 10 and the behaviour described in this item "I put myself in the other's place" is directly related to the construct "empathy". In its preliminary versions, four dimensions were considered for PBS (empathic behaviours, altruism, helping and sharing) and this item belonged to the first dimension (Auné & Abal, 2016). In the next development of the questionnaire, the authors reduced it to two dimensions, because the expected unidimensionality was not achieved (Auné & Attorresi, 2017). In this change, the factor "empathic behaviours" was reinterpreted and redefined to "comforting the other", as a consequence of the incorporation of items that belonged to other dimensions. Against this background, it can be concluded that, at the time of writing the item, it was based on an empathic behaviour approach, hence the strong emphasis on empathy in the content. Despite the mismatch of the Rasch model, it is recommended to keep this item based on the context of the questionnaire development and additionally, it is required to confirm the mismatch with a larger sample to rule out type I and II errors. Furthermore, in the studies by Garaigordobil and García de Galdeano (2006), significant positive correlations were found between prosocial behaviour and empathy in a sample of infants. Consistent with this finding, Silke, et al. (2018) conducted a systematic review and found 168 scientific articles documenting the association between these two variables.

In relation to the collection of evidence of the internal structure of the scale, the CFA allows us to discard model 1 (one-dimensional), because the  $\chi^2/df$  was greater than 3.00, the *NFI*, *NNFI*, *CFI* and *TLI* indices were less than 0.95 and the *RMSEA* was greater than 0.05. Models 2 and 3 were confirmed as having two-dimensional structures similar to those established by Auné and Attorresi (2017). Model 2 (bifactor with related dimensions) had an  $\chi^2/df$  less than 3.00, *NFI*, *NNFI*, *CFI* and *TLI* greater than 0.95 and *RMSEA* less than 0.05. Model 3 (bifactorial with independent dimensions) had better *NFI*, *NNFI*, *CFI* and *TLI* indices in the Helping dimension, but with a  $p > 0.05$  and in the Comforting dimension, the  $\chi^2/df$  was greater than 3.00, the *NFI*, *NNFI* and *TLI* indices were less than 0.95 and the *RMSEA* greater than 0.05. The results of model 3 are consistent with the proposal of Auné et al. (2020) to contemplate the possibility of measuring only the Helping dimension, however, because of obtaining a  $p > 0.05$  in the CFA, it is recommended to keep the two-dimensional model with related dimensions.

The model of two related dimensions showed no significant differences between men and women at the configural, metric, scalar and residual levels. The results show that the factor structure analyzed is invariant between both sexes. There was one person in the sample who identified with another sex, so it is suggested that further studies should be carried out in which more participants of this sex are included to assess whether the factorial invariance is maintained.

With the calibration of the items with the Rasch model it was possible to identify that the Helping dimension had a wider and more precise range compared to the Comforting dimension, which indicates a better measurement and is consistent with the results of the CTT and CFA analyses. Item 2 "In my free time I do volunteer activities" obtained the highest parameter  $b$  in the Helping dimension, which shows that this behaviour is the one that evidences higher levels of prosocial behaviour and is consistent with the findings of Picazo et al. (2020), who found that people who invest their free time in helping others tend to do so for a longer period of time, because it is part of their personal fulfilment and they do it to seek happiness. In the Comforting others dimension, the item with the highest parameter  $b$  was 4 "I act as a cane for others", which describes a behaviour that explicitly states the physical willingness to support other people, compared to the behaviours of the other items that refer to dispositions and attitudes towards the action of comforting. In the adaptation with the Peruvian population, it was decided to change the word "cane" for "support", which is a more general expression; however, in this adaptation to the Colombian population, it was decided to modify the questionnaire as little as possible. Even though the expression "I act like a cane" may be associated with a specific jargon, the psychometric results showed that it had no problems of interpretation, because the CTT indices and the Rasch model adjustment did not detect anomalies in the distributions of the responses. These results show that this expression is shared in the capital cities of Argentina and Colombia, because the samples of both studies were obtained in these geographical locations. For adaptations in populations from other countries, it is recommended to consider the expression used in this item due to the cultural load that it evidently has, and the modification made in the adaptation with the Peruvian population can be considered.

To make proper use of the PBS, it is recommended to score it in the two dimensions and transform the scores based on the specifications for Thurstone, Guttman and Likert scales proposed by García Sánchez et al. (2011), in order to assign the frequencies described in the items to the scores obtained in each dimension. This result is obtained by adding the values of the responses (between 1 and 6) and then the total is divided by the number of items that make up each dimension.

The results of this study provide insight into the psychometric qualities of the PBS of Auné and Attorresi (2017) and are extended for the HS of Auné et al. (2020) with the results obtained for the Helping subscale. The PBS is recommended to measure self-perceived prosocial behaviour from a two-dimensional perspective and the EA to measure self-perceived behaviours related to positive and voluntary actions to benefit others (Auné et al., 2020). Both the PBS (Auné & Attorresi, 2017) and the HS (Auné et al., 2020) can be used in the Colombian population, but it is recommended that they be used in research that applies the questionnaires in a sample that allows psychometric results to be obtained and compared with the results of this study.

The findings not only attest to the psychometric quality of the PBS for use in the Colombian population, but also provide evidence on the dimensions that make up prosocial behaviour. The results of the CFA show that behaviours associated with helping others had greater weight in the model, compared to behaviours aimed at comforting others. Additionally, the estimated reliability with the CTT was also higher on this dimension, indicating a better correlation between items describing helping behaviours. Consistent with these results, the Rasch model allowed us to identify that the items of the Helping dimension covered a wider range and with better measurement precision, which means that there is a greater representation of helping behaviours within a continuum expressed on a logit scale and related to prosocial behaviour.

Within the typologies that define and classify prosocial behaviour, Warneken and Tomasello (2009) mention three dimensions based on a review of various theoretical approaches and empirical results. These dimensions are: a) comforting; b) information sharing; and c) instrumental helping, the latter being more complex than the other two and paradoxically less studied. Instrumental helping encompasses those behaviours inclined to act on behalf of others and an ontogenetic and phylogenetic origin can be identified (Warneken & Tomasello, 2009). When reviewing the content of the items of the PBS Helping dimension, it is possible to recognise that in each of these items, behaviours are stated in which their objectives are aimed at acting towards the well-being of other people, and in this sense, the Helping dimension fits within the instrumental helping approach mentioned by Warneken and Tomasello (2009). Having cognitive and motivational components of ontogenetic origin, instrumental helping can be expressed in a larger number of behaviours within a wide range of probability of occurrence, given the high variability that can be generated by individual differences in these psychological processes. This characteristic may be related to the fact that we have obtained a greater amplitude in the Helping dimension by transforming it into a logit scale with the Rasch model. With this type of analysis, the Helping dimension of the PBS or HS of Auné et al. (2020) can be used to continue the study of the conceptual approach to instrumental helping, specifically in the adult population.

In general terms, although the results of this study were similar to those obtained by Auné and Attorresi (2017) with the original scale and Canales Reyes (2020) in the adaptation with the Peruvian population, this study had limitations in the sample design and sample size, specifically to make the estimates of factorial invariance, which lead us to take with caution the possibility of generalizing the conclusions about the psychometric properties of the PBS to the entire Colombian population. This same limitation prevented the use of methods to detect DIF, because to analyze questionnaires of 20 or fewer items requires at least a sample of 200 in each group (Scott et al., 2009). Another limitation was the lack of identification of the origin or residence of the participants, due to the cultural differences that have been identified between the regions that make up Colombia.

Finally, for future research analyzing the psychometric properties of the scale, we suggest following the recommendations of Pedrosa et al. (2013), which implies increasing the number of experts in the review of the content of the items and incorporating more methods that allow us to contrast the assessment of the experts' agreement. To expand the collection of evidence on the internal structure of the scale, it is recommended that invariance and DIF studies be carried out between Colombian and Latin American regions to determine its psychometric properties at a cross-cultural level. It would also be relevant to estimate test-retest reliability to identify the stability of the scale scores over time. Finally, it is necessary to continue with this type of study to collect more evidence of validity with other sources (AERA, APA & NCME, 2014), for example, the consequences derived from the use of the scale and the relationship between PBS scores and variables associated with prosocial behaviour.

## References

- 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](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](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. *Psykhē*, 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/10.1037/0033-2909.88.3.588>
- 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](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?pii=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://europepmc.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.intestcom.org>
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (Second edition). [www.InTestCom.org](http://www.InTestCom.org)
- Jorgensen, T. D., Pornprasertmanit, 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](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>
- Lipke, 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](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](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, E., 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 AHA!IE comunicaciones: *Revista Iberoamericana de Informática Educativa*. <https://redined.educacion.gob.es/xmlui/handle/11162/5670>
- 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., Fayers, 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. Bateria 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>
- Warneken, F. & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, 100(3), 455-471. <https://doi.org/10.1348/000712608X379061>
- Wispé, 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>

Date of receipt: August 2022.

Acceptance date: October 2023.