Application of the Graded Response Model to a Scale of Empathic Behavior

Aplicación del modelo de respuesta graduada a una escala de comportamiento empático

Sofía E. Auné[1, 2, *], Facundo J. P. Abal[1, 2], and Horacio F.Attorresi[1]

Abstract

The results obtained from the application of the Graded Response Model (GRM) to the items of the Scale of Empathic Behavior whose authors are Auné, Abal and Attorresi (2017) are presented. The sample was obtained by accessibility and consisted of Argentine university students (80% Women). None of the items presented uniform or non-uniform gender differential item functioning. The GRM assumptions of local independence and unidimensionality were successfully corroborated. Threshold parameters tended to be located at low levels of the trait scale, whereas discrimination parameters were high. The analysis of the information function showed acceptable precision in low and middle empathic behavior level.

Resumen

Se presentan los resultados obtenidos a partir de la aplicación del Modelo de Respuesta Graduada a los ítems de la Escala de Comportamientos Empáticos de Auné, Abal y Attorresi (2017), un instrumento compuesto por ocho ítems que se responden en una escala Likert de seis categorías de respuesta. La muestra fue obtenida por accesibilidad y estuvo compuesta por estudiantes universitarios argentinos (80% Mujeres). Ningún ítem presentó funcionamiento diferencial del ítem uniforme o no uniforme. Se corroboraron los supuestos del MRG de independencia local y unidimensionalidad, considerándose satisfechos. Los parámetros de umbral tendieron a ubicarse en niveles bajos de la escala del rasgo, mientras que los parámetros de discriminación tuvieron valores altos. El análisis con la Función de Información del Test mostró una precisión alta para niveles de rasgo bajos y medios.

Keywords:
Empathic Behavior, Empathy, Graded Response Model, Item Response Theory.

Palabras Claves:
Comportamientos Empáticos, Empatía, Modelo de Respuesta Graduada, Teoría de la Respuesta al Ítem.

1. Faculty of Psychology, University of Buenos Aires, Argentina.
2. National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina.
* Corresponding author: 577 Virrey Liniers Street (15th floor, apartment B), Autonomous City of Buenos Aires, Argentina. Postal Code 1220.
Email: sofiaaune@psi.uba.ar sofiaaune177@hotmail.com. Telephone number: +541149313374

Introduction

Empathy is generally described as the capacity to feel and share the feelings and thoughts of others (Decety, 2017). Actually, it is considered to be a multifaceted construct, with diverse components that interact: a) through emotional contact, the capacity to share the emotional states of others; b) under the empathetic inclination, a tendency to contribute to the wellbeing of others; and c) through taking perspective, the deliberate cognitive capacity to understand what another person thinks and feels (Decety, 2015; Decety & Svetlova, 2012; Derntl et al., 2010; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). Each of these components emanates from specific neurobiological processes (Decety, 2017).

Exercising empathy reinforces prosocial behavior (Decety & Cowell, 2014). This capacity can be shown to be present or absent in specific actions that may be taken in reaction to the suffering or pain experienced by others. In this sense, one who empathizes must be able to relieve such suffering (Batson, Van Lange, Ahmad, & Lishner, 2003). On the other hand, according to many studies, the capacity to show empathy runs counter to aggressive and violent behavior (e.g. Winter, Spengler, Bermpohl, Singer, & Kanske, 2017). Also, anomalies have been detected in schizophrenic patients in their cognitive capacity to exercise empathy (Vistoli, Lavoie, Sutliff, Jackson, & Achim, 2017). People with

schizophrenia showed an abnormal modulation of activation in the temporoparietal junction (TPJ) bilaterally (Vistoli et al., 2017). TPJ is the specific brain system associated with cognitive capacity to exercise empathy.

The relationship between empathy and conduct is reflected in the concept of empathic behavior. This term has been used in different studies, often loosely, without a clear understanding behind it. Auné, Blum, Abal, Lozzia and Atorresi (2014) defined empathic behavior as those actions that show comprehension, effort and emotional support. These authors have developed the scale of empathic behavior (SEB) to measure the tendency to behave with empathy.

The Scale of Empathic Behavior (SEB)

The SEB (Auné, Abal, & Attorresi, 2017) is composed of eight items that are responded to using a 6-point frequency scale *(Never, Almost Never, Sometimes, Often, Almost Always, Always)*. By means of exploratory factor analysis (EFA), the developers of the SEB provided evidence that a single dimension runs through the SEB items and that the SEB items are internally consistent (α = 0.82; glb = 0.89). In addition, the authors demonstrated convergent validity between the measurements of the SEB and the Argentinian version of the Assumed Perspective of the Scale of Prosocial Abilities (SPA, Morales Rodriguez & Suarez Perez, 2011), as well as with the study participants' self-perception of their own prosocial behavior. The study participants were 1168 Argentine university students (81% Women).

Modeling of the SEB with the Item Response Theory (IRT)

Even though validity evidence for the SEB scores was established using EFA and classical test theory (CTT) methods, it is also useful to generate psychometric evidence by modeling the SEB with item response theory (IRT, De Ayala, 2009). It is important to mention that the general use of IRT models in Argentina and Latin America is recent and of limited extent, and even more limited in applications to tests of typical behaviors by those who evaluate such modes of behavior. The application of IRT in non-educational environments poses challenges that have been surmounted primarily with narrow constructs. IRT offers specific models for examining and organizing items in instruments and a process of evaluation that cannot be conducted with CTT and EFA methods.

The GRM developed by Fumiko Samejima (1969) is an extension of the two-parameter logistic (2PL) model for items of dichotomous response to the polytomous case. It is appropriate when the responses to an item can be classified into more than two ordered categories, such as to represent different degrees of achievement in the solution to a problem or levels of agreement or frequency to a certain statement. The responses to a given item are classified in M + 1 ordered categories, where M is the total number of original categories of the item, so that those with the smallest numbers represent less of the feature measured by the item than higher categories. We will represent the generic category of the item as m. The scores for the item are successive integers. Despite this, it is assumed that the discrete response s of an individual to an item is part of a continuum of response S.

For each item, a slope (a) parameter and a set of threshold parameters of category b_m are estimated. The b_m are the points of the latent trait where the probability of choosing category m or higher is 0.5. The discrimination parameter a is constant for all the item category characteristic curves (ICCCs) within an item and reports the degree to which their response categories distinguish between trait levels. This means the a parameter can vary across items within the same instrument. The parameter a in the GRM is denoted in different ways. Some authors call it discrimination as is done for the 2PL model (e.g., Escurra Mayaute, & Salas Blas, 2014) and others call it the slope parameter (e.g., Hernández Baeza, Muñiz Fernández, & García Cueto, 2000).

Samejima (1969) developed a two-step process to obtain the probability that an individual chooses a certain category in a given item. In the first step, the probability that an individual with a certain trait level chooses a category m or greater in a given item is expressed by the following formula (Equation 1):

$$P_m^*(\theta) = \frac{1}{1 + e^{-1.7a(\theta - b_m)}}$$

where:

1.7 is the specialized D parameter, since this value gives the best approximation within one decimal point of the logistic function to the accumulated normal.

a is the parameter of slope or discrimination of the item.

b_m is the point of the latent trait where the probability of choosing category m or higher is 0.5.

e is the base of the natural logarithm: 2.718

θ is the level in the latent trait, in this case the level of empathic behavior.

$P_m^*(\theta) = 1$ if $m = 1$ since the cumulative probability of responding in the lowest category or in all the major ones is a certain event.

$P_m^*(\theta) = 0$ if $m = M + 1$ since the probability of responding in a category following the largest is null.

In the second step, the probability that an individual responds in a given category is defined as (Equation 2):

$$P_m(\theta) = P_m^*(\theta) - P_{(m+1)}^*(\theta)$$

The use of this formula, which subtracts the cumulative probability to the right of a given category and that of the next one, allows us to obtain the ICCCs. Each ICCC describes the probability of giving a certain response option to the item for each level of the trait it measures.

Purpose

The general purpose of this study was to evaluate the psychometric properties of the SEB in a sample of Argentine university students using the GRM.

The specific aims of the study were as follows:

a. To present findings related to differential item functioning (DIF) in the SEB. Analyses of gender were performed.
b. To investigate the psychometric properties of an 8-item SEB scale administered to Argentine university students using IRT.

Method

Participants

The sample consisted of 1136 Argentine university students (80.7% Women). The mean age was 21.5 years with a standard deviation of 5.8. The majors with the greatest weight in the sample were Medicine (13.4%), Law (9.6%), Public Accounting (9%), Psychology (5.3%), Business Administration (5.3%) and Audiovisual Arts. (4.4%). The rest of the participants were distributed, with very low frequency, in more than 40 different university careers.

Procedure

The sampling method was convenience sampling. Informed consent was included in the administered protocol. In addition, it was clarified that the use was exclusively for research purposes and participation was entirely voluntary and could be discontinued at any time.

Instruments

Sociodemographic Questionnaire. This included variables such as age, gender, academic major and current academic level of the questionnaire participant.

Scale of Empathic Behavior (SEB). Details about the SEB item type were previously mentioned. The specific items that comprise the SEB are as follows: 1) *When I feel that somebody is distressed, I show them that I understand them;* 2) *I behave as a support to others;* 3) *I put myself in the other person's shoes;* 4) *I tell others how much they are valued;* 5) *I share the bad experiences of my life with friends, so that they won't make the same mistakes I did;* 6) *If a person tells me about a conflict, I try to explain to him the point of view of the other person;* 7) *I try to "raise" the self-esteem of my friends; and* 8) *I am there for my associates in times of sadness.*

Data Analysis

The assessment of the unidimensionality assumption assumed by the GRM was carried out by means of an EFA of the polychoric correlation matrix on the eight items that make up the SEB using FACTOR version 10.8.04 (Lorenzo-Seva & Ferrando, 2013). The factor extraction method was Unweighted Least Squares (ULS) with promin rotation. The procedure for determining the number of dimensions was optimal implementation of Parallel Analysis (Timmerman & Lorenzo-Seva, 2011).

The assumption of local independence is also required by the GRM. Through the χ^2_{LD} index proposed by Chen and Thissen (1997), each pair of items was tested. If the value of the χ^2_{LD} is greater than 10, it indicates possible local dependence (LD) and the non-fulfillment of the assumption. The index is composed of two parts: the χ^2 statistic and a standardized odds ratio difference. The latter can cause the result to be negative; this would indicate that the observed frequencies of the pair of items have less LD than that predicted by the model (Chen & Thissen, 1997).

Once the assumptions of the GRM were corroborated, the GRM was implemented through the Item Response Theory Patient Reported Outcomes (IRTPRO) 4.2 program (Cai, Thissen, & du Toit, 2011). The adjustment of each item to the model was evaluated with the S-χ^2 index (Orlando & Thissen, 2000, 2003) and considered to fit the GRM if the associated $p > 0.01$ (Toland, 2014). Global adjustment of the model was calculated with the M_2 statistic (Maydeu Olivares & Joe, 2005, 2006) and the associated RMSEA index; the adjustment was accepted if the RMSEA was less than or equal to 0.05 (Hu & Bentler, 1999; Yu, 2002). The GRM parameters were estimated by marginal maximum likelihood (Bock & Aitkin, 1981).

The DIF by gender was analyzed following the procedure designed by Woods (2009) for its detection. DIF was first tested by applying the modified Wald test (Cai, 2012; Cai et al., 2011; Langer, 2008) to each SEB item, considering remaining items as anchors. Thus, the possibility that some items could exhibit DIF was explored. Then, a second analysis was performed, specific for one item with potential DIF. For this, both groups -women and men- were anchored with one item which was presumed free of DIF. The goal of this second step was to identify a group invariant designated anchor item that more accurately produced Wald´s statistic test results than those using all other items as anchors (Woods, 2009).

Since a unidimensional IRT model was implemented, only the latent trait θ was considered to be empathic behavior. The model calculated, for each item, a slope parameter a and five threshold parameters b_m, given that there were six response options. Higher parameters a reflected a greater strength of the relationship between each item on the SEB scale an the latent trait empathic behavior.

Once the GRM was found tenable and DIF was not present, the item information function (IIF) was computed for each item to indicate how much reliable information each item was able to provide to the measurement of the latent trait of empathic behavior. Also, the reliability of the SEB was assessed in the framework of the IRT by the test information function (TIF).

Results

Assessment of GRM Assumptions

The EFA on the eight items that make up the SEB indicated that the advised number of dimensions was one. This satisfied the requirement for the presence of a single preponderant factor. Bartlett's statistic = 2399.5 (df = 28; p = 0.000010) and Kaiser-Meyer-Olkin (KMO) test = 0.86926 showed that the correlation matrix could be factorized. The cumulative proportion of variance of the first eigenvalue was 0.4925.

With respect to local independence, the Chen and Thissen´s χ^2_{LD} statistic took values for each pair of items between -0.8 and 6.8, lower than the limit value of 10. Therefore, the pairs of items were considered sufficiently locally independent to implement the GRM.

Assessment of Differential Item Functioning for Gender

Table 1 shows DIF analysis of the SEB. There is no DIF by gender.

Table 1
DIF Statistics for Gender

Item	Total χ^2 (df = 6)	p	χ^2_a (df = 5)	p	χ^2_b (df = 1)	p
1	3.0	0.8124	0.3	0.5691	2.7	0.7542
2	4.7	0.5805	0.0	0.9329	4.7	0.4524
3	12.3	0.0553	4.4	0.0363	7.9	0.1602
4	4.4	0.6225	0.7	0.4094	3.7	0.5904
5	3.2	0.7792	0.2	0.6577	3.0	0.6946
7	4.3	0.6414	0.4	0.5135	3.8	0.5738
8	0.8	0.9918	0.2	0.6238	0.6	0.9894

Note. The Total χ^2 refers to the omnibus test for DIF, χ^2_a refers to the test for non-uniform DIF, and χ^2_b refers to the test for uniform DIF

Graded Response Model Item Parameters Estimation

The global assessment (M_2 = 829.24, df = 692, p = 0.0001, RMSEA = 0.02) was adequate. The individual assessment of each item to the GRM indicated that the item 6 didn´t fit the GRM ($p < 0.01$). To ensure the appropriate implementation of the GRM to the data, item 6 was excluded from the analysis (Toland, 2014). Table 2 summarizes the item level fit results.

Table 2
S-χ^2 index values

Item	χ^2	df.	p
1	99.14	82	0.0956
2	122.94	106	0.1245
3	98.15	88	0.2152
4	112.85	100	0.1787
5	129.81	103	0.0381
6	125.90	89	0.0061
7	88.83	73	0.1000
8	111.95	94	0.0998

df: degrees of freedom

The results of the item parameters obtained with the GRM are shown in Table 3. Overall, the items traverse a wide portion of the latent trait, from -4.05 (b_1 item 1) to 2.01 (b_5 item 2).

With respect to a parameters, they had values from 1.19 to 1.99, which represents a high capacity of the response categories to distinguish between participants with different trait levels according to the criteria of Reise and Waller (1990).

By way of illustration, the ICCCs of items 2 and 7 are shown in Figures 1 and 2, respectively. With respect to item 2, the distance between the b_m was wide, allowing the discrimination between different levels of empathic behavior and each of the response categories. However, since the a parameter was comparatively low and its b_m was far between, the general discriminative capacity of the item was relatively low. A middle level is needed in the feature to select *Often* or a higher category. With respect to item 7 it is possible to indicate that, although a parameter is very high, this discriminatory capacity occurs at the medium-low and low levels of the trait. A medium level of empathic behavior is enough to make it much more likely to choose the category than the other response options.

Item Information Function and Test Information Function

Table 4 shows values derived from the IIF and FIT for some levels of empathic behavior trait. As we can observe, the test, and most of the results, are more informative for middle and low levels of the trait, where the standard error (SE) in the measurement is less than that of the higher levels. Items 3 and 4 provide more information at a positive level of the trait. The SE increases with higher levels of the trait, where both the reliability and the level of information that the test offers diminish.

Figure 3 displays the TIF. For the GRM, the TIF reached its maximum value of 6.45, in which θ = -0.40 with a minimum s.e. value of at this point of 0.39. The

Table 3

Graded response model parameter estimates for the SEB

Item	a(s.e.)	b₁(s.e.)	b₂(s.e.)	b₃(s.e.)	b₄(s.e.)	b₅(s.e.)
1	1.82 (0.12)	-4.05 (0.34)	-2.74 (0.15)	-1.42 (0.08)	-0.60 (0.05)	0.25 (0.05)
2	1.25 (0.08)	-3.00 (0.19)	-1.95 (0.12)	-0.31 (0.06)	0.86 (0.08)	2.01 (0.13)
3	1.45 (0.09)	-4.24 (0.33)	-2.91 (0.18)	-1.00 (0.07)	0.05 (0.05)	1.14 (0.08)
4	1.77 (0.11)	-2.52 (0.14)	-1.43 (0.08)	-0.33 (0.05)	0.45 (0.05)	1.17 (0.07)
5	1.19 (0.08)	-3.50 (0.24)	-2.57 (0.17)	-0.96 (0.08)	0.04 (0.06)	0.97 (0.08)
7	1.99 (0.14)	-3.54 (0.25)	-3.07 (0.19)	-1.61 (0.08)	-0.76 (0.05)	-0.04 (0.05)
8	1.52 (0.10)	-3.60 (0.25)	-2.62 (0.15)	-1.28 (0.08)	-0.50 (0.06)	0.34 (0.06)

*Note. **a** refers to the slope parameter; **b₁**, … **b₅** refer to the five threshold parameters; **s.e.** refers to the standard error*
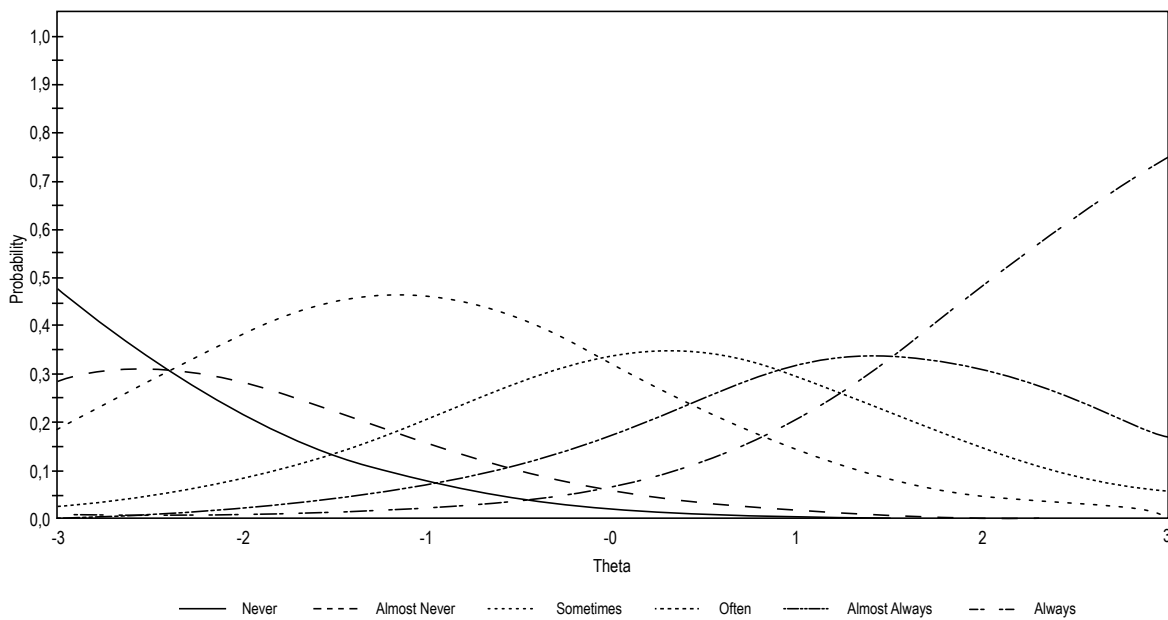

Figure 1. Graded Response Model Item Category Characteristic Curves for Item 2. Theta: level of empathic behavior on the latent trait continuum.
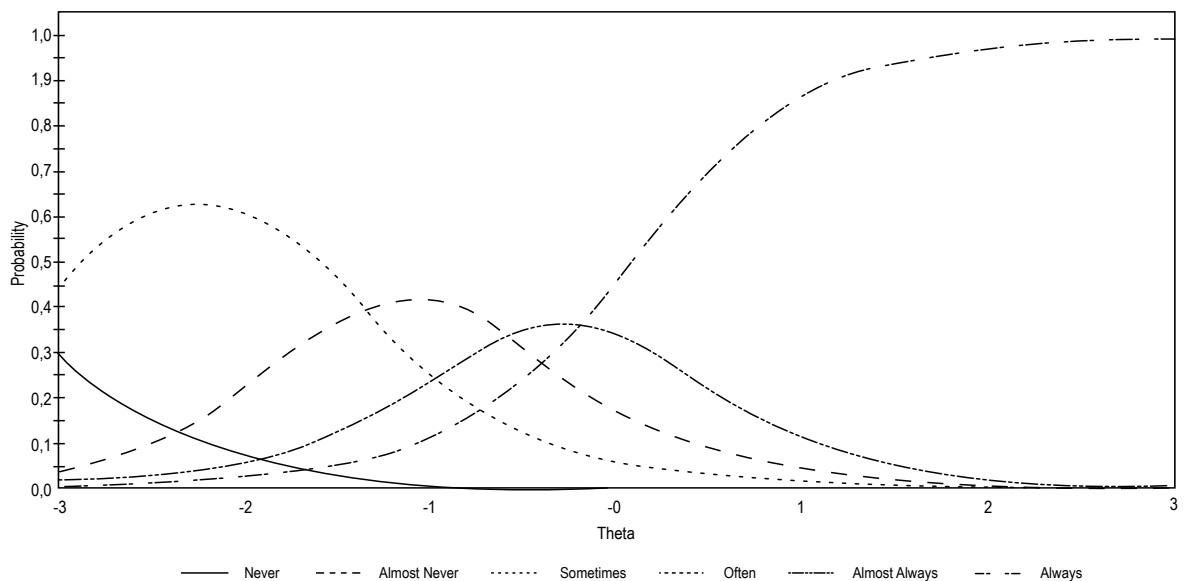

Figure 2. Graded Response Model Item Category Characteristic Curves for Item 7 Theta: level of empathic behavior on the latent trait continuum.

Table 4
Graded response model Item Information Function and Test Information Function

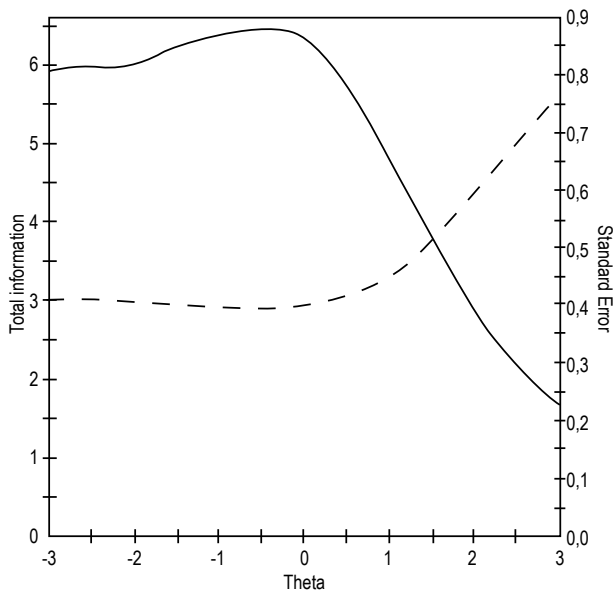| Item \ θ	-2.4	1.6	0.8	0	0.8	1.6	2.4
1	0.93	0.96	1.02	0.98	0.66	0.24	0.06
2	0.47	0.46	0.46	0.48	0.48	0.47	0.39
3	0.56	0.56	0.63	0.65	0.62	0.49	0.25
4	0.88	0.93	0.93	0.97	0.96	0.71	0.29
5	0.43	0.43	0.44	0.45	0.42	0.32	0.19
7	1.00	1.15	1.21	1.10	0.53	0.14	0.03
8	0.69	0.69	0.73	0.71	0.54	0.26	0.09
Test	5.97	6.19	6.42	6.32	5.21	3.62	2.30
s.e.	0.41	0.40	0.39	0.40	0.44	0.53	0.66

Figure 3. Test Information Function of the Scale of Empathic Behavior (Graded Response Model).
Theta = level of empathic behavior on the latent trait continuum.

level of information held relatively constant through the low and mid-levels of the trait, decreasing considerably in the high levels.

Discussion

Empathic behavior develops around the second year of life, based on children's capacities to distinguish between themselves and others (Decety & Meyer, 2008) and regulate their own negative emotions (Paulus, Kühn-Popp, Licata, Sodian, & Meinhardt, 2013). The comfort dimension of prosocial conduct defined by Dunfield (2014) and Warneken and Tomasello (2009) can be related to empathic behavior. Hay and Cook (2007) describe this as feelings for others, and

Caprara, Steca, Zelli and Capanna (2005) call it feeling empathic with others, showing that this behavior has been described many times.

The empathic behaviors included in the criteria of the SEB, in which one comforts, assists, and emotionally supports another, are, above all, compassionate. The actions listed are part of the daily life of most of the people surveyed, and generally lead to a circle of mutual social support. Psychometric properties typical of the narrow constructs are demonstrated therein with high correlations of the items with the test (Auné et al., 2017), elevated parameters of inclination and a one-dimensional core (Reise, & Revicki, 2015).

Analysis of the SEB items with the GRM provides evidence that the SEB offers a greater level of information in the middle and low levels of the trait. The standard error grows substantially with higher levels of empathic behavior. As the level of information increases, this investigative capacity yields appropriate values for the totality of the criteria that comprise the SEB. Therefore, if new items are added to SEB to produce a favorable finding, it could be suggested that such items will uncover a high level of empathic behavior.

With respect to a detailed analysis of the items, items 1, 7 and 8 were found to be negative and offered more information in very low, low and intermediate levels of empathic behavior. Items 3, 4 and 5 yielded information in a broader range. As for item 2, it yielded relatively little information along the entirety of the scale.

The analysis presented shows the importance of performing modeling with IRT, because it provides in depth information about the psychometric attributes of each item, to detect if each has DIF, and to evaluate the relationship of each one of the response options of the Likert scale with the latent trait that it attempts to measure. DIF was not evidenced with the SEB regarding gender, indicating that comparisons of gender groups on the SEB don't contain construct irrelevant variance in this instance.

The most important limitation of the present study is the degree of specificity of the sampling. Future investigations will replicate the analysis in the general population and in other specific contexts. Also, the scale needs to include more items to better measure the top end of the scale. Additionally, some of the items could be revised as they are not always clear. Future research should include cognitive interviews and content expert reviews to build test content evidence validity for the SEB and validity evidence of response processes.

References

Auné, S., Abal, F., & Attorresi, H. (2017). Propiedades Psicométricas de una Prueba de Conducta Empática. *Revista Iberoamericana de Diagnóstico y Evaluación-e Avaliação Psicológica, 3* (45), 47-56. doi:10.21865/RIDEP45.3.04

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.

Batson, C., Van Lange, P., Ahmad, N., & Lishner, D. (2003). Altruism and helping behavior. In M. A. Hogg & J. Cooper (Eds.), *Sage Handbook of Social Psychology* (pp. 279–295). London, UK: Sage Publication.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443-459.

Cai, L. (2012). *flexMIRT: Flexible multilevel item factor analysis and test scoring* [Computer software]. Seattle, WA: Vector Psychometric Group, LLC.

Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO user's guide*. Lincolnwood, IL: Scientific Software International.

Caprara, G., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European Journal of Psychological Assessment, 21*(2), 77-89. doi:10.1027/1015-5759.21.2.77

Chen, W., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265-289. doi:10.3102/10769986022003265

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Decety, J. (2015). The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences, 3*, 1-6. doi:10.1016/j.cobeha.2014.12.001

Decety, J. (2017). Moral Progress Requires a Coupling Between Empathy and Reason? A Social Neuroscience Perspective. In *Cultural Patterns and Neurocognitive Circuits II* (Vol. Volume 5, pp. 187–209). WORLD SCIENTIFIC. doi:10.1142/9789813230484_0009

Decety, J., & Cowell, J. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences, 18*(7), 337-339. doi:10.1016/j.tics.2014.04.008

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and psychopathology, 20*(4), 1053-1080. doi:10.1017/S0954579408000503

Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience, 2*(1), 1-24. doi:10.1016/j.dcn.2011.05.003

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology, 35*(1), 67-82. doi:10.1016/j.psyneuen.2009.10.006

Dunfield, K. A. (2014). A construct divided: prosocial behavior as helping, sharing, and comforting subtypes. *Frontiers in Psychology, 5*, 958. doi:10.3389/fpsyg.2014.00958

Escurra Mayaute, M., & Salas Blas, E. (2014). Construcción y validación del cuestionario de adicción a redes sociales (ARS). *Liberabit, 20*(1), 73-91.

Hay, D., & Cook, K. (2007). The transformation of prosocial behavior from infancy to childhood. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 100-131). New York, NY: The Guilford Press.

Hernández Baeza, A., Muñiz Fernández, J., & García Cueto, E. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema, 12*(2), 288-291.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.

Lorenzo-Seva, U., & Ferrando, P.J. (2013). Factor 9.2: A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement, 37*(6), 497-498. doi:10.1177/0146621613487794

Maydeu Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*(471), 1009-1020. doi:10.1198/016214504000002069

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713-732. doi:10.1007/s11336-005-1295-9

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.

Reise, S. & Revicki, D. (2015). Age-old problems and modern solutions. In Reise, S. & Revicki, D. (eds.) *Handbook of ítem Response Theory Modeling. Applications to Typical Performance Assessment* (pp. 3-12). New York, NY: Taylor & Francis.

Reise, S., & Waller, N. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*(1), 45-58. doi:10.1177/014662169001400105

Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*(1), 50-64. doi:10.1177/01466216000241003

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-χ2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27*, 289-298. doi:10.1177/0146621603027004004

Paulus, M., Kühn-Popp, N., Licata, M., Sodian, B., & Meinhardt, J. (2013). Neural correlates of prosocial behavior in infancy: different neurophysiological mechanisms support the emergence of helping and comforting. *Neuroimage, 66*, 522-530. doi:10.1016/j.neuroimage.2012.10.041

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement, 17*(4), 2. doi:10.1002/j.2333-8504.1968.tb00153.x

Shamay-Tsoory, S., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617-627. doi:10.1093/brain/awn279

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods, 16*(2), 209-220. doi:10.1037/a0023353

Toland, M. (2014). Practical guide to conducting an item response theory analysis. The *Journal of Early Adolescence, 34*(1), 120-151. doi:10.1177/0272431613511332

Vistoli, D., Lavoie, M., Sutliff, S., Jackson, P. L., & Achim, A. M. (2017). Functional MRI examination of empathy for pain in people with schizophrenia reveals abnormal activation related to cognitive perspective-taking but typical activation linked to affective sharing. *Journal of Psychiatry & Neuroscience: JPN, 42*(4), 262. doi:10.1503/jpn.160136

Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology, 100*(3), 455-471. doi:10.1348/000712608X379061

Winter, K., Spengler, S., Bermpohl, F., Singer, T., & Kanske, P. (2017). Social cognition in aggressive offenders: Impaired empathy, but intact theory of mind. *Scientific reports, 7*(1), 670. doi:10.1038/s41598-017-00745-0

Woods, C. (2009). Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement, 33*(1), 42-57. doi:10.1177/0146621607314044

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation). Los Angeles, CA. Retrieved from http://www.statmodel.com/download/Yudissertation.pdf