_____

# Validity and factorial invariance of the Interpersonal Reactivity Index in Peruvian university students

# Validez e invarianza factorial del Índice de Reactividad Interpersonal en universitarios peruanos

# Validade e invariância fatorial do Índice de Reatividade Interpessoal em universitários peruanos

Mirian Grimaldo Muchotrigo[1], ORCID 0000-0003-4197-1906
Jossue Correa-Rojas[2], ORCID 0000-0002-4166-7210
Eduardo Manzanares Medina[3], ORCID 0000-0002-3377-311X
Karen Macavilca Milera[4], ORCID 0000-0001-6198-6318

[1] *Universidad Peruana de Ciencias Aplicadas, Peru*
[2] *Universidad Peruana de Ciencias Aplicadas, Peru*
[3] *Universidad Peruana de Ciencias Aplicadas, Peru*
[4] *Universidad Peruana de Ciencias Aplicadas, Peru*

## Abstract

The Interpersonal Reactivity Index (IRI) is one of the most used self-report measures in young people and adults to measure empathy, it considers cognitive and affective aspects in its composition. This research presents the structural validity, invariance and reliability of the measure in Peruvian university students. 859 psychology students, between 17 and 39 years old, from two universities in Metropolitan Lima were intentionally selected. The confirmatory factor analysis performed on three models: Model 1, replicates the original structure, Model 2, analyzes the original structure, but negatively worded items and model 3 were excluded, inverse items and those with irrelevant variance were excluded. construct (item 9). This last model is the one with the best fit (CFI = .924; RMSEA = .091; SRMR = 0.058). On this model, the configurational, metric, scalar and residual invariance according to age was established. In addition, a second order model was tested that demonstrated the presence of the latent factors Cognitive Empathy and Emotional Reactions, this model presents an adequate fit (CFI = .957; RMSEA = .078; SRMR = 0.068). It is concluded that the reduced version of the IRI is a valid and reliable measure in Peruvian university students.

*Keywords*: factorial analysis; psychometric properties; validity and reliability; empathy; Peruvian university students

## Resumen

El Índice de Reactividad Interpersonal (IRI) constituye una de las medidas de autoinforme más utilizadas en jóvenes y adultos para medir la empatía, ya que considera en su composición aspectos cognitivos y afectivos. En esta investigación se presenta la validez estructural, la invarianza y fiabilidad de la medida en universitarios peruanos. Se seleccionaron intencionalmente 859 estudiantes de Psicología, entre 17 y 39 años, de dos universidades de Lima Metropolitana. El análisis factorial confirmatorio realizado sobre tres modelos: modelo 1, replica la estructura original; modelo 2, analiza la estructura original sin los ítems redactados en forma negativa; modelo 3, se excluyeron los ítems inversos y los que presentaban varianza irrelevante de constructo (ítem 9). Este último modelo es el que presenta un mejor ajuste (CFI = .924; RMSEA = .091; SRMR = 0.058). Sobre este modelo se estableció la invarianza configuracional, métrica, escalar y residual

Mirian Grimaldo Muchotrigo, Jossue Correa-Rojas,
Eduardo Manzanares Medina & Karen Macavilca Milera

según edad. Adicionalmente, se probó un modelo de segundo orden que demostró la presencia de los factores latentes empatía cognitiva y reacciones emocionales, este modelo presenta un ajuste adecuado (CFI = .957; RMSEA = .078; SRMR = 0.068). Se concluye que la versión reducida del IRI es una medida válida y fiable en universitarios peruanos.

*Palabras clave:* análisis factorial; propiedades psicométricas; confiabilidad y validez; empatía; universitarios peruanos

**Resumo**

O Índice de Reatividade Interpessoal (IRI) é uma das medidas de autorrelato mais utilizadas em jovens e adultos para mensurar a empatia, e considera aspectos cognitivos e afetivos em sua composição. Nesta pesquisa apresenta-se a validade estrutural, a invariância e a confiabilidade da medida em estudantes universitários peruanos. Foram selecionados intencionalmente 859 estudantes de psicologia, entre 17 e 39 anos, de duas universidades da região metropolitana de Lima. A análise fatorial confirmatória realizada em três modelos: Modelo 1, replica a estrutura original, Modelo 2, analisa a estrutura original, mas foram excluídos os itens redigidos negativamente, e o Modelo 3, se excluíram os itens inversos e os que apresentavam variância de construto irrelevante (item 9). Este último modelo é o que apresenta melhor ajuste (CFI = .924; RMSEA = .0091; SRMR = 0.058). Sobre este modelo foram estabelecidas as invariâncias configuracional, métrica, escalar e residual segundo a idade. Adicionalmente, foi testado um modelo de segunda ordem que demonstrou a presença dos fatores latentes Empatia Cognitiva e Reações Emocionais, este modelo apresenta um ajuste adequado (CFI = .957; RMSEA = .078; SRMR = 0.068). Conclui-se que a versão reduzida do IRI é uma medida válida e confiável em estudantes universitários peruanos.

*Palavras-chave:* análise fatorial; propriedades psicométricas; confiabilidade e validade; empatia; estudantes universitários peruanos

Empathy is an emotional and cognitive response coming from the understanding of another person's condition (Eisenberg, 2000); it allows coping with vital demands (Cañero et al., 2019), and motivates cooperation (Zaki, 2018) and morality (De Waal, 2010). It has been studied from three perspectives: cognitive, emotional, and integrative (Lorente, 2014; Palma, 2013). The integrative view of empathy gained strength in the 1980s (Parra et al., 2012). An important representative of this perspective is Davis (1980) who defines empathy as a reaction to the observed experience and suggests the presence of different factors involved in the construct.

Based on this theoretical structure, Davis (1980) designed the Interpersonal Reactivity Index (IRI), which constitutes one of the most widely used self-report measures in youth and adults to measure this attribute (Jordan et al., 2016; Lucas-Molina et al., 2017). The IRI is made up of four subscales that measure two cognitive components: (a) Perspective-Taking, which constitutes the tendency to adopt the other's point of view, and (b) Fantasy, considered as the tendency to identify with the feelings and actions of fictional characters, and two affective components: (c) Empathic Concern, referring to feelings of concern for people in need, and (d) Personal Distress, referring to feelings of personal anxiety in the face of others' distress.

This study used the Spanish adaptation of Mestre et al. (2004), that analyzed the validity of the IRI from the relationship with other constructs with which it correlated positively such as prosocial behavior and prosocial reasoning styles, and negatively with aggressive behavior and emotional instability. In terms of reliability, it yielded the following values: Fantasy .70, Empathic Concern .65, Personal Distress .64, and Perspective-Taking .56.

The IRI has been adapted and validated in different contexts. In Asia (Siu & Shek, 2005), a reliability index ranging from .65 to .70 was obtained for its dimensions; in addition to defining three factors through factor analysis. Wang et al. (2020) note that researchers often use IRI scores flexibly in their studies based on different empathy constructs. Thus, some assume it as a general construct and sum the four dimensions of the IRI to obtain an empathy score (Sun et al., 2018), although without empirical evidence to justify such a process (Dueber & Toland, 2021). In addition, researchers who recognize empathy as a dualistic construct combine the dimensions of Empathic Concern and Personal Distress into an affective empathy factor and the dimensions of Perspective-Taking and Fantasy into a cognitive empathy factor (Fan & Hu, 2017). In other cases, they combine the dimensions of Empathic Concern and Personal Distress to represent affective empathy, but only Perspective-Taking as cognitive empathy (He & Zhu, 2016). They have also been found to only consider the Empathic Concern and Perspective-Taking dimensions as representing affective empathy and cognitive empathy, respectively (Luo et al., 2013).

Regarding the above, in the United States, Pulos et al. (2004) found that, if a higher order empathy scale of the IRI is desired, the Empathic Concern, Perspective-Taking, and Fantasy subscales are useful. Thus, the Personal Distress dimension is left out, due to its low correlation with the other dimensions. Recently, the Personal Distress dimension is deficient in construct validity (Murphy et al., 2018). In Canada, in a study conducted with young adults, they considered the score of the Perspective-Taking and Empathic Concern subscales as the overall score (Nicol & Rounding, 2013). In Mexico, Ahuatzin et al. (2019) found the existence of four factors, although some items were shifted from one dimension to another, as found in the study by Pérez-Albéniz et al. (2003).

In the Netherlands, the psychometric properties of the IRI were examined using a Confirmatory Factor Analysis (CFA) and the results revealed that there is a need for psychometric tailoring of the scores in terms of the factor structure. However, overall, the internal consistency and the factor structure of the scores of the Dutch version suggest that it is a useful instrument (De Corte et al., 2007).

In Latin America, there are several instrumental studies in different countries. In Colombia, Bernal et al. (2015), determined the psychometric properties of the IRI and concluded that they should eliminate five items that had in common being negatively worded, to improve the reliability indices of each of the subscales. Similarly, García-Barrera et al. (2017), examined the internal structure of the IRI in ex-combatants and found that the negative items presented low factor loadings.

In Argentina, Richaud de Minzi (2008) validated the instrument in a child population by verifying the internal structure through exploratory factor analyses carried out consecutively in independent samples, which showed that the IRI retains its basic internal structure, although an unstable behavior of the negatively worded items was observed; an Alpha of .70 was found for the total version. For their part, Müller et al. (2015) analyzed the factor structure using a CFA with a sample of adults and obtained a four-dimensional model adjusted to the theoretical model, with adequate reliability for

each dimension. In Chile, Fernández et al. (2011) applied the CFA demonstrating a good fit of the model in line with Davis' (1980) proposal, with adequate internal consistency.

In Peru (Diaz et al., 2015), its internal structure did not demonstrate solvency and in terms of reliability it obtained alpha coefficients equal to .41 for the Perspective-Taking dimension, .49 for Fantasy, .48 in Empathic Concern, and .52 in the Personal Distress dimension; these results suggest that negatively worded items could influence responses due to cultural and even linguistic differences.

One of the least explored psychometric properties of the IRI is its age invariance. In this sense, establishing the invariance of the measure implies that the structure, items, measures, and measurement errors are similar in both groups. That is, it would only be possible to interpret the scores of the instrument in the same way in different groups (Byrne, 2008). Thus, given that this study was conducted on a sample of psychology students, it is expected that in the early stages of their training no variations are established concerning empathy; however, exposure to activities and knowledge specific to the program may generate changes in the attribute as they advance in their training, which is why age has been considered as a criterion, this characteristic is related to the academic cycle of university students.

The measurement of empathy in psychology students is important since the mastery of empathic ability is an essential element in the training of professionals (Hernández, 2019). In addition, the IRI constitutes a widely used tool in the assessment of empathy, due to its multidimensional nature (Müller et al., 2015). In this sense, the study aimed to provide evidence of the factorial structure of the IRI. The existence of a general latent factor underlying the dimensions of this scale was verified, the invariance according to age was analyzed, and, finally, the reliability by internal consistency.

## Method

### Participants

A total of 859 psychology students were purposively selected from two universities in Lima, 31.8 % were studying at a public university and 68.2 % at a private university. The 77.3 % were females with an average age of 20.36 ($SD = 2.69$) and 22.7 % were males with an average age of 21.29 ($SD = 2.64$), the ages of the total sample ranged from 17 to 29 years ($M = 20.57$; $SD = 2.70$). For the distribution of the participants in the age groups, the concept of emerging adults was taken into consideration, whose ages range from 14 to 29 years; in this case, the sample was subdivided into two groups of emerging adults, the first group from 17 to 20 years (55.50 %) where there is a predominance of students from first cycles, with characteristics more oriented towards adolescence, and a second group from 21 to 29 years (44.50 %), where there is a predominance of students from intermediate and final cycles, whose characteristics were more oriented towards adulthood. As for their distribution by year of study, 26.4 % were first-year students, 14.7 % were second-year students, 23.1 % were third-year students, 17.8 % were fourth-year students, and 18 % were fifth-year students.

### Instrument

The Interpersonal Reactivity Index (IRI; Davis, 1983) is a measure that explores cognitive and affective empathy, using a Likert-type response format with five options: *It does not describe me well* (1), *It describes me a little* (2), *It describes me well* (3), *It describes me fairly well* (4), and *It describes me very well* (5). It consists of 28 items that allow measuring individual differences of the empathy construct through the following four subscales (7 items each): Perspective-Taking (PT) and Fantasy (F) (the cognitive

component) and Empathic Concern (EC) and Personal Distress (PD) (the emotional component). For this investigation, we used the Spanish adaptation of Mestre et al. (2004), who analyzed the evidence of validity based on discriminant analysis, finding canonical correlations equal to .436 between the subscales of the IRI that allow a correct classification of males in 70.2 % and 69.7 % in females; the authors point out that both sex and age introduce biases in the measurement of empathy. The validity of other constructs was verified with prosocial behavior (.312), aggressiveness (-.171), and hedonistic reasoning (-.142). The internal consistency of the dimensions is partially acceptable PT (.56), F (.70), EC (.64) and PD (.68).

**Procedure**

The instrument was administered in April and May 2019, divided into different groups, and taken collectively. The recommendations and regulations for the application of tests proposed by the International Test Commission (2001) were considered to minimize the variance irrelevant to the construct prone to occur during the administration of psychological tests. The participants completed the tests and signed the informed consent form detailing aspects related to their participation, emphasizing the voluntary nature and freedom to participate in the study, thus it was communicated that the assessment would be anonymous and that the results would be confidential.

**Data Analysis**

First, the missing data were analyzed, and when it was found to be absent, the outliers were verified. Then, the absolute frequency of the response options was explored to identify that all of them were functional, following the criteria of Linacre (2002), who states that frequencies lower than 10 can be problematic. Next, the descriptive measures of the items were analyzed considering the mean ($M$), standard deviation ($SD$), skewness ($g_1$), and kurtosis ($g_2$), through the latter, it was possible to evaluate the distribution of the items considering $\pm 2.0$ as the marginal value of normality (Reed & Wu, 1974).

The internal structure was analyzed using confirmatory factor analysis, testing 3 models: Model 1, which includes all 28 items, Model 2, which excludes unstable items, and Model 3, which excludes all inverse items. The Weighted Least Square Mean and Variance Adjusted (WLSMV) method was used due to the ordinal nature of the data. The chi-square between degrees of freedom, Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) were checked both with values below .08; the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) whose acceptable values are above .90 (Kline, 2016; Xia & Yang, 2018). Additionally, the Average Variance Extracted (AVE) was calculated with satisfactory values around .50. Moreover, the invariance of the measure was explored by considering deltas less than .01 for $\Delta$CFI and $\Delta$RMSEA (Byrne, 2008).

Finally, the categorical omega coefficient ($\omega$) was analyzed with its confidence intervals, because of random error, it is advisable to establish a range to estimate the real value of reliability (Ventura-León, 2018). The Compositive Reliability (CR) was calculated (Zumbo et al., 2007). The analyses were performed with IBM SPSS, version 25, and RStudio version 3.3.2 (RStudio Team, 2015), using the Lavaan package (Rosseel, 2012).

## Results

Table 1 shows the absolute frequencies of the response options for each of the IRI items, showing that the response options are functional, except for item 27, in which the response option *It describes me very well* has a low frequency of responses. However, this did not affect the calculations performed.

**Table 1**

*Descriptive statistics of the IRI*

| | Absolute frequencies | | | | | Descriptive measures | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Min | Max | M | SD | g1 | g2 |
| iri1 | 103 | 191 | 260 | 185 | 120 | 1 | 5 | 3.03 | 1.22 | 0.00 | -0.89 |
| iri2 | 45 | 108 | 261 | 312 | 133 | 1 | 5 | 3.44 | 1.06 | -0.44 | -0.32 |
| iri3 | 29 | 73 | 220 | 326 | 211 | 1 | 5 | 3.72 | 1.03 | -0.62 | -0.07 |
| iri4 | 34 | 83 | 193 | 252 | 297 | 1 | 5 | 3.81 | 1.13 | -0.68 | -0.36 |
| iri5 | 223 | 235 | 214 | 129 | 58 | 1 | 5 | 2.49 | 1.22 | 0.40 | -0.81 |
| iri6 | 121 | 289 | 295 | 126 | 28 | 1 | 5 | 2.59 | 1.01 | 0.24 | -0.43 |
| iri7 | 54 | 179 | 294 | 222 | 110 | 1 | 5 | 3.18 | 1.10 | -0.06 | -0.67 |
| iri8 | 21 | 52 | 239 | 356 | 191 | 1 | 5 | 3.75 | 0.95 | -0.60 | 0.21 |
| iri9 | 18 | 75 | 261 | 340 | 165 | 1 | 5 | 3.65 | 0.96 | -0.44 | -0.13 |
| iri10 | 91 | 291 | 297 | 127 | 53 | 1 | 5 | 2.72 | 1.04 | 0.34 | -0.35 |
| iri11 | 15 | 54 | 228 | 359 | 203 | 1 | 5 | 3.79 | 0.93 | -0.56 | 0.09 |
| iri12 | 30 | 88 | 234 | 292 | 215 | 1 | 5 | 3.67 | 1.07 | -0.51 | -0.36 |
| iri13 | 71 | 185 | 229 | 210 | 164 | 1 | 5 | 3.25 | 1.22 | -0.12 | -0.98 |
| iri14 | 28 | 86 | 189 | 291 | 265 | 1 | 5 | 3.79 | 1.09 | -0.66 | -0.29 |
| iri15 | 49 | 94 | 260 | 271 | 185 | 1 | 5 | 3.52 | 1.11 | -0.45 | -0.42 |
| iri16 | 177 | 222 | 245 | 146 | 69 | 1 | 5 | 2.66 | 1.21 | 0.25 | -0.85 |
| iri17 | 130 | 254 | 275 | 156 | 44 | 1 | 5 | 2.69 | 1.09 | 0.18 | -0.66 |
| iri18 | 27 | 58 | 127 | 190 | 457 | 1 | 5 | 4.15 | 1.10 | -1.18 | 0.50 |
| iri19 | 85 | 268 | 356 | 111 | 39 | 1 | 5 | 2.71 | 0.97 | 0.25 | -0.09 |
| iri20 | 74 | 208 | 297 | 205 | 75 | 1 | 5 | 3.00 | 1.08 | 0.01 | -0.64 |
| iri21 | 15 | 55 | 279 | 338 | 172 | 1 | 5 | 3.69 | 0.92 | -0.40 | -0.04 |
| iri22 | 58 | 129 | 295 | 211 | 166 | 1 | 5 | 3.35 | 1.15 | -0.21 | -0.68 |
| iri23 | 96 | 190 | 285 | 182 | 106 | 1 | 5 | 3.01 | 1,17 | 0.02 | -0.78 |
| iri24 | 236 | 329 | 224 | 55 | 15 | 1 | 5 | 2.17 | 0.96 | 0.57 | -0.07 |
| iri25 | 56 | 146 | 320 | 256 | 81 | 1 | 5 | 3.19 | 1.04 | -0.22 | -0.40 |
| iri26 | 62 | 166 | 279 | 226 | 126 | 1 | 5 | 3.22 | 1.13 | -0.13 | -0.72 |
| iri27 | 335 | 315 | 159 | 42 | 8 | 1 | 5 | 1.92 | 0.92 | 0.82 | 0.18 |
| iri28 | 29 | 102 | 251 | 297 | 180 | 1 | 5 | 3.58 | 1.05 | -0.41 | -0.42 |

*Note*. *M*: Mean, *SD*: Standard deviation, *g1*: Coefficient of skewness, *g2*: Coefficient of kurtosis.

### Evidence for Validity Based on the Internal Structure

Table 2 shows the three models evaluated. The first model corresponds to the original proposal of four factors. This model presents unstable factor loadings in the negatively worded items ($< .40$), it also presents AVE values ranging between .20 and .36 and its goodness-of-fit indices are inadequate (CFI = .521; RMSEA = .133; SRMR = .139). For this reason, these items were not considered for the third model, which obtained adequate factor loadings and acceptable goodness of fit indices (CFI = .924; RMSEA = .073; SRMR = .058). These results show that the reduced version of the IRI is a measure that maintains its original structure, but improves when the inverse items (7, 15, 7, 12, 4, 14, 18, 13, and 19) and redundant statements (item 9) are excluded (see Table 3).

**Table 2**

*Confirmatory factor analysis of the IRI*

| | Model 1 | | | | | Model 2 | | | | | Model 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | F1 | F2 | F3 | F4 | $\Lambda^2$ | F1 | F2 | F3 | F4 | $\Lambda^2$ | F1 | F2 | F3 | F4 | $\Lambda^2$ |
| iri8 | .65 | | | | 0.43 | .60 | | | | 0.35 | .59 | | | | 0.35 |
| iri11 | .74 | | | | 0.54 | .71 | | | | 0.50 | .71 | | | | 0.50 |
| iri21 | .69 | | | | 0.47 | .64 | | | | 0.40 | .64 | | | | 0.41 |
| iri25 | .65 | | | | 0.43 | .64 | | | | 0.40 | .65 | | | | 0.42 |
| iri28 | .65 | | | | 0.42 | .65 | | | | 0.42 | .65 | | | | 0.42 |
| iri3 | .08 | | | | 0.01 | - | | | | - | - | | | | - |
| iri15 | .19 | | | | 0.04 | - | | | | - | - | | | | - |
| iri1 | | .62 | | | 0.38 | | .66 | | | 0.43 | | .66 | | | 0.43 |
| iri5 | | .65 | | | 0.42 | | .63 | | | 0.39 | | .63 | | | 0.40 |
| iri16 | | .72 | | | 0.51 | | .75 | | | 0.57 | | .75 | | | 0.57 |
| iri23 | | .78 | | | 0.60 | | .77 | | | 0.59 | | .77 | | | 0.59 |
| iri26 | | .73 | | | 0.54 | | .71 | | | 0.50 | | .71 | | | 0.50 |
| iri7 | | -.03 | | | 0.00 | | - | | | - | | - | | | - |
| iri12 | | .06 | | | 0.00 | | - | | | - | | - | | | - |
| iri2 | | | .53 | | 0.28 | | | .56 | | 0.31 | | | .56 | | 0.32 |
| iri9 | | | .51 | | 0.26 | | | .54 | | 0.28 | | | - | | - |
| iri20 | | | .64 | | 0.40 | | | .61 | | 0.37 | | | .65 | | 0.42 |
| iri22 | | | .65 | | 0.42 | | | .70 | | 0.49 | | | .73 | | 0.53 |
| iri4 | | | .03 | | 0.00 | | | - | | - | | | - | | - |
| iri14 | | | .09 | | 0.01 | | | - | | - | | | - | | - |
| iri18 | | | .08 | | 0.01 | | | - | | - | | | - | | - |
| iri6 | | | | .66 | 0.43 | | | | .60 | 0.35 | | | | .60 | 0.35 |
| iri10 | | | | .71 | 0.50 | | | | .70 | 0.48 | | | | .69 | 0.48 |
| iri17 | | | | .78 | 0.61 | | | | .77 | 0.59 | | | | .77 | 0.59 |
| iri24 | | | | .73 | 0.54 | | | | .69 | 0.47 | | | | .70 | 0.48 |
| iri27 | | | | .62 | 0.39 | | | | .60 | 0.36 | | | | .61 | 0.37 |
| iri13 | | | | .18 | 0.03 | | | | - | - | | | | - | - |
| iri19 | | | | .12 | 0.01 | | | | - | - | | | | - | - |
| F1 | - | | | | | - | | | | | | | | | |
| F2 | .26 | - | | | | .32 | - | | | | .32 | | | | |
| F3 | .64 | .65 | - | | | .63 | .63 | - | | | .47 | .63 | | | |
| F4 | -.15 | .52 | .52 | - | | .04 | .54 | .58 | - | | .04 | .54 | .62 | | |
| AVE | 0.33 | 0.35 | 0.20 | 0.36 | | 0.42 | 0.50 | 0.29 | 0.46 | | 0.42 | 0.50 | 0.42 | 0.46 | |
| CR | 0.74 | 0.73 | 0.53 | 0.76 | | 0.78 | 0.83 | 0.70 | 0.81 | | 0.78 | 0.83 | 0.69 | 0.81 | |

*Note.* $\Lambda^2$ = Determinant; AVE: Average Variance Extracted, CR: Composite Reliability.

**Table 3**

*Goodness of Fit Indices*

| | SB-$\chi2$(*df*) | IFC | TLI | RMSEA [CI 90%] | SMRM | WRMR |
|---|---|---|---|---|---|---|
| Model 1 | 3895,295(344) | 0.521 | 0.474 | .133 [.129-.137] | 0.139 | 3.316 |
| Model 2 | 1042,545(146) | 0.890 | 0.871 | .085 [.080-.089] | 0.068 | 1.920 |
| Model 3 | 714,377(129) | 0.924 | 0.910 | .073 [.068-.078] | 0.058 | 1.633 |

*Note.* SB $\chi^2$ = Satorra-Bentler chi-square; *df:* degrees of freedom.

**Factorial invariance of the IRI according to age**

The invariance of the reduced version of the IRI (model 3) was analyzed according to the age of the participants, differentiating between two groups, one under 21 years of age and the other older than 21 years of age. It can be seen that the RMSEA values and their variations are acceptable ($\Delta$RMSEA < .01), as are the variations in SRMR ($\Delta$SRMR < .01). As for, changes in CFI, its adequacy was demonstrated ($\Delta$CFI < .01). These findings demonstrate that the invariance of the measure has been established in

both groups. In Table 4, it is observed that the values of the fit indices are satisfactory for the configural, metric, scalar, and strict invariance, the delta of the metric variation is in the margins suggested by the literature (Byrne, 2008).

**Table 4**

*Differentials of the IRI Adjustment Indices*

| Invariance | χ2 | *df* | IFC | RMSEA | SRMR | Δ *df* | Δ CFI | Δ RMSEA | Δ SRMR |
|---|---|---|---|---|---|---|---|---|---|
| Configural | 857.311 | 258.000 | 0.919 | 0.074 | 0.066 | | | | |
| Metrics | 938.165 | 322.000 | 0.916 | 0.067 | 0.066 | 64.000 | -0.002 | -0.007 | 0.001 |
| Scalar | 1146.619 | 326.000 | 0.889 | 0.077 | 0.066 | 4.000 | -0.028 | 0.010 | 0.000 |
| Strict | 1169.047 | 344.000 | 0.888 | 0.075 | 0.068 | 18.000 | -0.001 | -0.002 | 0.001 |

Figure 1 presents a second-order CFA where the latent variables on which the original dimensions of the IRI are grouped are established. The DWLS method was used, and a satisfactory fit was obtained (CFI = .957; RMSEA = .078 [.073-.083]; SRMR = 0.068; WRMR = 1.902. Through this proposed modeling, the structure of two latent factors (Cognitive Empathy and Emotional Empathy) that contain the specific factors (Cognitive Empathy: Perspective-Taking and Fantasy, and Emotional Empathy: Empathic Concern and Personal Distress) described in the original theoretical model of the IRI is confirmed.

**Internal Consistency of the IRI**

Internal consistency was analyzed through the categorical omega coefficient with their respective confidence intervals. The results show that the dimensions of the IRI are consistent in terms of their scores, for the Perspective-Taking dimension an omega of .757 [.719-.784] was obtained, in Fantasy an omega equal to .802[.776-.825] was found, and the Empathic Concern dimension reached an omega equal to .653[.606-.689], and the Personal Distress dimension an omega of .782 [.753-.806].

**Figure 1**
*Second-Order Confirmatory Factor Analysis of the IRI*
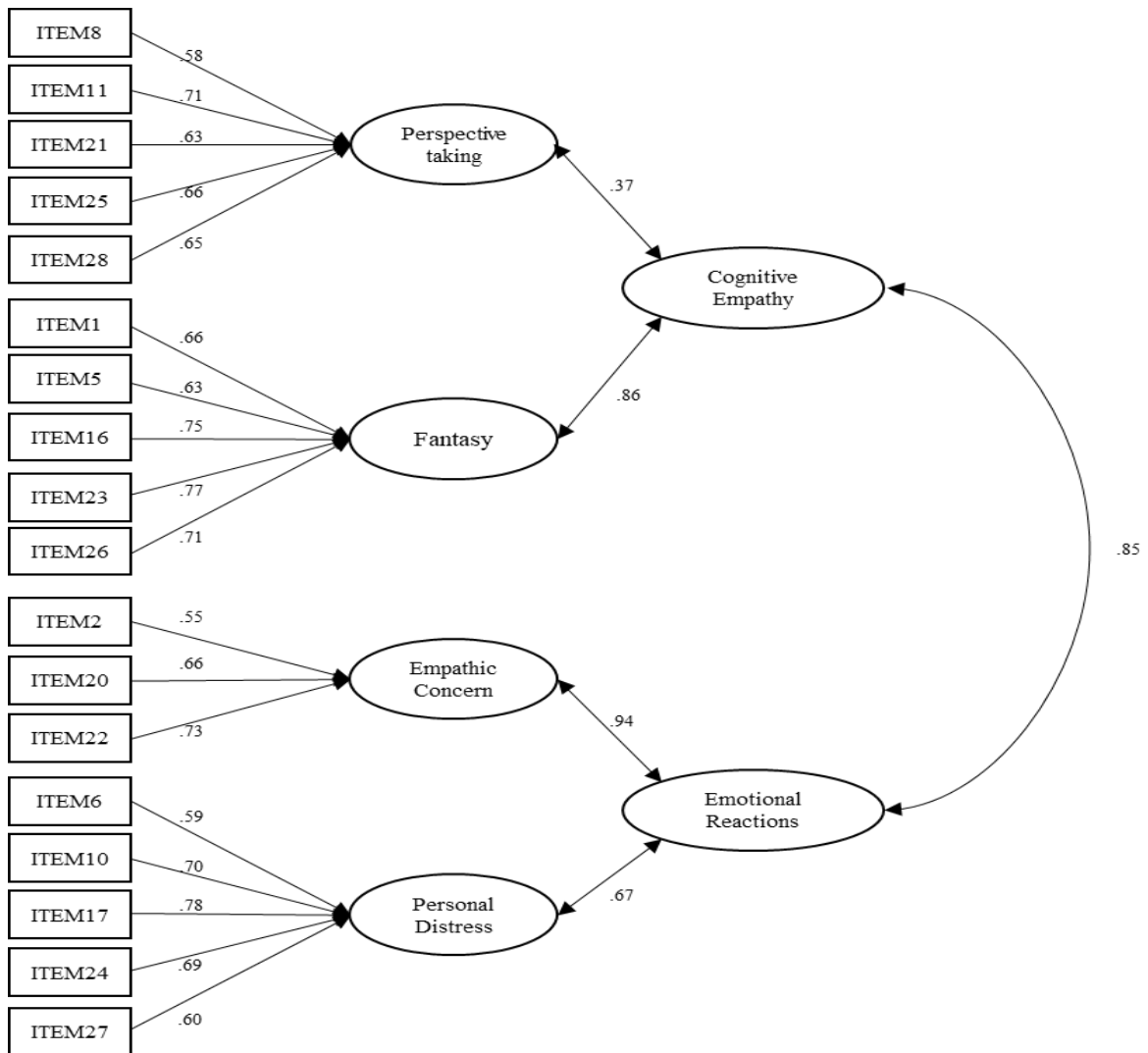


## Discussion

The findings indicate that the evidence based on the structural validity of the instrument is appropriate after the elimination of the inverse items since it presents a structure consistent with the original theoretical model. In addition, a favorable increase in the reliability indexes is observed. With this, it is demonstrated that negatively worded items have an important effect on validity and reliability (Suárez-Álvarez et al., 2018; Tomás et al., 2012), which coincides with that reported by Bernal et al. (2015) in a Colombian sample. Thus, similar dimensions were found; however, the presence of negative items or those that possess a directionality opposite to the logic of the construct (Weijters & Baumgartner, 2012) affect the structure of the IRI.

Although some authors such as Nunnally (1978) suggest the need to incorporate negative items to control acquiescence; that is, the tendency to answer oriented to one extreme regardless of the content (Hidalgo-Rasmussen & Gonzáles-Betanzos, 2019), which generates bias and could produce response patterns (Van Sonderen et al., 2013). In addition, the presence of reverse items can confuse respondents (Podsakoff et al., 2003),

this phenomenon is known as the method effect (Abad et al., 2016; Danner et., 2015); and can produce biases in the measurement of an attribute (Tomás et. al, 2013).

Regarding the above, Benson and Hocevar (1985 as cited in Weems et al., 2003) point out that it is complex to determine the evidence based on the construct since the items are organized more by a semantic aspect than by the theoretical construct and should therefore be done with caution. Other studies reveal that the combination of positive and negative items fails to reduce the acquiescence bias (Sauro & Lewis, 2011), which could be characterized in samples with low educational or cognitive levels (Meisenberg & Williams, 2008), which is why they cannot establish differences in positively and negatively worded items (Solis, 2015); although it could also evidence laziness or indifference at the time of answering a test (Solis, 2015), which is also demonstrated in the inconsistency between items and could lead to measurement error.

As found in this study, negatively or inversely worded items were, apparently, not understood or as pointed out by Suárez-Álvarez et al. (2018) could have been misinterpreted. Therefore, some authors suggest using positively worded items (DeVellis, 2003). Although this depends on the culture and educational level (Tomás et al., 2012); thus, it has been found that the recognition of negative items in the United States is more feasible than in Asian countries (Wong et al., 2003). Similarly, the study by Marín et al. (2002) showed inconsistencies when combining positive and negative items.

Finally, after the elimination of the negative or inverse items, the evidence supports the validity based on the internal structure and consistency of the IRI scores, in line with the theoretical model proposed by Davis (1980). A substantial increase is seen in all goodness-of-fit indices when these items are excluded, where AVE values also approach a satisfactory level, and factor loadings reach satisfactory values (Kline, 2016). This growth further extends to the reliability measures. It is concluded that the Interpersonal Reactivity Index is a valid and reliable measure of the empathy construct in university students in Lima.

A second-order model was also analyzed to identify the theoretical congruence of the IRI concerning the original proposal. The findings were favorable, identifying that the IRI in its reduced version retains the theoretical structure initially proposed by Davis (1983). These results coincide with other studies (Mestre et al., 2004; Mestre et al., 1999). These findings have a very important practical implication: it is possible to obtain a total score of the dimensions that compose the IRI and this score can be interpreted as the presence of the empathy attribute.

Based on the results obtained, the invariance of the measure is established. The practical implications of these findings are that the construct is understood in the same way by university students in initial and final cycles. Likewise, the equivalence of the construct is established in both groups; therefore, it is possible to use the instrument to make comparisons in similar samples and under the same criteria (Byrne, 2008). This implies that training in their program and knowledge do not affect the self-perception of this attribute, nor its measurement with this scale.

Regarding the evidence of reliability, the findings show the consistency of the IRI scores which, despite the reduction of negatively worded items, it is observed that all dimensions reached acceptable omega coefficients. These results coincide with other studies in which the stability of the IRI is also noted (Fernández et al., 2011; Müller et al., 2015). However, it is necessary to mention that although the Empathic Concern dimension obtained an omega slightly below .70, which in turn is lower than that obtained in the other dimensions, this does not disqualify the use of the measure because as pointed out by Oviedo and Campos-Arias (2005), these turn out to be referential values. Furthermore, according to Cortina (1993), its use is pertinent to the extent that no other

_____

more efficient measure exists. In line with this, these results are substantially better than the values reported by Mestre et al. (2004) in the validation carried out on the Spanish population.

One limitation is that the participants were not selected by probability sampling. This limits the generalizability of the results. It was also not possible to calculate the stability of the IRI, this property is necessary to the internal consistency since it denotes the presence of the attribute and its susceptibility to be modified (Correa-Rojas, 2021).

It is suggested to replicate the study in different urban groups from different cities and different programs, as it has been found that empathy levels may vary as a function of gender, program, and academic degree (Kidron et al., 2018).

## Funding

## References

Abad, F. J., Sorrel, M. A., Garcia, L. F., & Aluja, A. (2016). Modeling general, specific, and method variance in personality measures: results for ZKA-PQ and NEO-PI-R. *Assessment*, *25*(8), 959-977. https://doi.org/10.1177/1073191116667547

Ahuatzin, A., Martínez-Velázquez, E., García, G., & Vázquez-Moreno, A. (2019). Propiedades psicométricas del Interpersonal Reactivity Index (IRI) en mexicanos universitarios. *Revista Iberoamericana de Psicología, 12*(1), 111-122.

Bernal, E. L., Cetina, M. J. & Romero, L. F. (2015). *Propiedades psicométricas del Interpersonal Reactivity Index (IRI) diseñado por Davis (1980), en una muestra de residentes de la ciudad de Bogotá* [Unpublished Bachelor's Thesis]. Universidad Piloto de Colombia.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema, 20*(4), 872-882.

Cañero, M., Mónaco, G., & Montoya, I. (2019). La inteligencia emocional y la empatía como factores predictores del bienestar subjetivo en estudiantes universitarios. *European Journal of Investigation in Health, Psychology and Education, 9*(1), 19-29. https://dialnet.unirioja.es/servlet/articulo?codigo=7109640

Correa-Rojas, J. (2021). Coeficiente de Correlación Intraclase: Aplicaciones para estimar la estabilidad temporal de un instrumento de medida. *Ciencias Psicológicas*, *15*(2), e–2318. https://doi.org/10.22235/cp.v15i2.2318

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. https://doi.org/10.1037/0021-9010.78.1.98

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality, 57,* 119-130. https://doi.org/10.1016/j.jrp.2015.05.004

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Davis, M. H. (1983). Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. https://doi.org/10.1037/0022-3514.44.1.113

De Corte, K., Buysse, A., Verhosfstadt, L., & Roeyers, H. (2007). Measuring empathic tendencies: Reliability and validation of the dutch versión of the Interpersonal Reactivity Index, *Psychologica Belgica, 47*(4), 235-260. https://doi.org/10.5334/pb-47-4-235

De Waal, F. (2010). *The age of empathy: Nature's lessons for a kinder society*. Random House.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Sage Publications.

Díaz, B., Rodríguez, K., & Santa Cruz, H. (2015). Propiedades del Índice de Reactividad Interpersonal en alumnos del ISTP San Luis de Trujillo. *Revista JANG, 2*(1), 2-14.

Dueber, D. M. & Toland, M. D. (2021). A bifactor approach to subscore assessment. *Psychological Methods.* https://doi.org/10.1037/met0000459

Eisenberg, N. (2000). Emotion, regulation and moral development. *Annual Review of Psychology, 51*, 665-697. https://doi.org/10.1146/annurev.psych.51.1.665

Fan, M. & Hu, Y. (2017). Status and related factors of empathy in adolescents. *Chinese Mental Health Journal*, *31*(11), 879-884. https://doi.org/10.3969/j.issn.1000-6729.2017.11.009

Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile. Empathy in a different cultural context. *European Journal of Psychological Assessment, 27*(3), 179-185. https://doi.org/10.1027/1015-5759/a000065

García-Barrera, M. A., Karr, J. E., Trujillo-Orrego, N., Trujillo-Orrego, S., & Pineda, D. A. (2017). Evaluating empathy in Colombian ex-combatants: examination of the internal structure of the Interpersonal Reactivity Index (IRI) in Spanish. *Psychological Assessment, 29*(1), 116-122. https://doi.org/10.1037/pas0000331

He, N. & Zhu, Y. (2016). Self-love and other-love: research on the relationships among narcissism, empathy and implicit altruism. *Acta Psychologica Sinica*, *48*, 199-210. https://doi.org/10.3724/SP.J.1041.2016.00199

Hernández, R. M. (2019). La empatía: una necesidad en la educación superior. *Educación Médica*, *20*(Sup. 1), 190. https://doi.org/10.1016/j.edumed.2017.07.026

Hidalgo-Rasmussen, C. A. & Gonzáles-Betanzos, F. (2019). El tratamiento de la Aquiescencia y la Estructura Factorial de la Escala Breve de Resiliencia (BRS) en estudiantes universitarios mexicanos y chilenos. *Anales de Psicología, 35*(1), 26-32. http://dx.doi.org/10.6018/analesps.35.1.297781

International Test Commission. (2001). *Guidelines on Test Use: Spanish Version*. https://www.intestcom.org/files/guideline_test_use.pdf

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct? *Emotion, 16*(8), 1107-1116. https://doi.org/10.1037/emo0000228

Kidron, R., Kaganovsky, L., & Baron-Cohen, S. (2018). Empathizing-systemizing cognitive styles: Effects of sex and academic degree. *PLoS ONE, 13*(3). https://doi.org/10.1371/journal.pone.0194515

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Linacre, J. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Lorente, S. (2014). Efecto de la competencia social, la empatía y la conducta prosocial en adolescentes [Unpublished doctoral dissertation]. Universitat de València.

Lucas-Molina, B., Perez-Albeniz, A., Ortuño, J., & Fonseca, P. (2017). Dimensional structure and measurement invariance of the Interpersonal Reactivity Index (IRI) across gender. *Psicothema, 29*(4), 590-595. https://doi.org/10.7334/psicothema2017.19

Luo, X., Zheng, D., Liu, W., & Liu, Y. (2013). The relationship between higher vocational school students' empathy and forgiveness: the mediating effect of trait anger. *Chinese Journal of Special Education, 5,* 88-92.

Marín, G., Gamba, R. & Marin, B. (1992). Extreme Response Style and Acquiescence among Hispanics: The Role of Acculturation and Education, *Journal of Cross-Cultural Psychology, 23*(4), 498-509. https://doi.org/10.1177/0022022192234006

Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*(7), 1539-1550. https://doi.org/10.1016/j.paid.2008.01.010

Mestre, V., Pérez Delgado, E., Frías, D., & Samper, P. (1999). Instrumentos de evaluación de la empatía. In E. Pérez Delgado & V. Mestre (Coords.), *Psicología moral y crecimiento personal* (pp. 181-190). Ariel.

Mestre, V.; Frías; M. D. & Samper, P. (2004). La medida de la empatía: análisis del Interpersonal Reactivity Index. *Psicothema, 16*(2), 255-260.

Müller, M., Ungaretti, J., & Etchezahar, E. (2015). Evaluación multidimensional de la empatía: adaptación del Interpersonal Reactivity Index (IRI) al contexto argentino. *Revista de Investigación en Psicología Social, 3*(1), 42-53.

Murphy, B. A., Costello, T. H., Watts, A. L., Cheong, Y. F., Berg, J. M., & Lilienfeld, S. O. (2018). Strengths and weaknesses of two empathy measures: a comparison of the measurement precision, construct validity, and incremental validity of two multidimensional indices. *Assessment, 27,* 246-260. https://doi.org/10.1177/1073191118777636

Nicol, A. A. & Rounding, K. (2013). Alienation and empathy as mediators of the relation between social dominance orientation, right-wing authoritarianism and expressions of racism and sexism. *Personality and Individual Differences*, 55, 294-299. https://doi.org/10.1016/j.paid.2013.03.009

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGrawHill.

Oviedo, H. & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, *34*(4), 572-580.

Palma, J. (2013). *Componentes cognitivos y afectivos en la configuración de la personalidad* [Unpublished doctoral dissertation]. Universitat de València.

Parra, A. M., Carvajal, N. M., & Moreno, J. H. (2012). *Acoso escolar y empatía en un grupo de adolescentes con trastornos disocial de la conducta* [Graduation project]. Universidad de La Sabana.

Pérez-Albéniz, A., De Paúl, J., Etxeberría, J., Paz, M., & Torres, E. (2003). Adaptación de Interpersonal Reactivity Index (IRI) al español. *Psicothema*, *15*(2), 267-272. https://www.psicothema.com/pdf/1056.pdf

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. http://dx.doi.org/10.1037/0021-9010.88.5.879

Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality, 32*, 355-360. https://doi.org/10.2224/sbp.2004.32.4.355

Reed, A. H. & Wu, G. T. (1974). Estimation of a transformation method for estimation of normal range. *Clinical Chemistry, 20*(5), 576-581.

Richaud de Minzi, M. C. (2008). Estudio del IRI de Davis en población infantil argentina. *Revista de Investigación en Psicología, 11*(1), 101-115. https://doi.org/10.15381/rinvp.v11i1.3880

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.v048.i02

RStudio Team. (2015). *RStudio: Integrated Development for R*. http://www.rstudio.com/

Sauro, J. & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 2215-2224), Association for Computing Machinery, Nueva York. https://doi.org/10.1145/1978942.1979266

Siu, A. & Shek, D. (2005). Validation of the Interpersonal Reactivity Index in a Chinese Context. *Research on Social Work Practice, 15*(2), 118-126. http://dx.doi.org/10.1177/1049731504270384

Solis, M. (2015). The dilemma of combining positive and negative items in scales. *Psicothema, 27*(2), 192-199. http://dx.doi.org/10.7334/psicothema2014.266

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: a questionable practice. *Psicothema*, *30*(2), 149-158. https://doi.org/10.7334/psicothema2018.33

Sun, J., Liu, P., & Li, D. (2018). Attentional characteristics of high-empathy people in processing emotional stimuli and evidence of eye movements. *Journal of Psychological Science, 41*, 1084-1089.

Tomás, J. M., Galiana, L., Hontangas, P., Oliver, A., & Sancho, P. (2013). Evidencia acumulada sobre los efectos de método asociado a ítems invertidos. *Psicológica, 34,* 365-381.

Tomás, J. M., Sancho, S., Oliver, A., Galiana, L., & Melendez, J. (2012). Efecto de método asociado a ítems invertidos vs ítems en negativo. *Revista Mexicana de Psicología, 29*(2)*,* 105-115.

Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Correction: ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. *PLoS ONE, 8*(9). https://doi.org/10.1371/annotation/af78b324-7b44-4f89-b932-e851fe04a8e5

Ventura-León, J. (2018). Intervalos de confianza para coeficiente omega: propuesta para el cálculo. *Adicciones, 30*(1). https://doi.org/10.20882/adicciones.962

Wang, Y., Li, Y., Xiao, W., Fu, Y., & Jie, J. (2020). Investigation on the rationality of the extant ways of scoring the Interpersonal Reactivity Index based on confirmatory factor analysis. *Frontiers in Psychology, 11*, 1086. https://doi.org/10.3389/fpsyg.2020.01086

Weems, G., Onwuegbuzie, A., Schreiber, J. & Eggers, S. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. *Assessment & Evaluation in Higher Education, 28*(6), 587-607. https://doi.org/10.1080/0260293032000130234

Weijters, B. & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: a review. *Journal of Marketing Research, 49*(5), 737-747. https://doi.org/10.1509/jmr.11.0368

Wong, N., Rindfleisch, A., & Burroghs, J. E. (2003). Do reversed worded items confound measures in cross cultural consumer research? The case of material values scale. *Journal of Consumer Research, 30*(1), 72-91. http://dx.doi.org/10.1086/374697

Xia, Y. & Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51,* 409-428. http://dx.doi.org/10.3758/s13428-018-1055-2

Zaki, J. (2018). Empathy is a moral force. In K. G. Gray, J. (Ed.), *The atlas of moral psychology* (pp. 49-58). Guilford Press.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal version of coefficients Alpha and Theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21-29.

**How to cite:** Grimaldo Muchotrigo, M., Correa-Rojas, J., Manzanares Medina, E., & Macavilca Milera, K. (2022). Validity and factorial invariance of the Interpersonal Reactivity Index in Peruvian university students. *Ciencias Psicológicas*, *16*(2), e-2810. https://doi.org/10.22235/cp.v16i2.2810

**Authors' participation:** a) Conception and design of the work; b) Data acquisition; c) Analysis and interpretation of data; d) Writing of the manuscript; e) Critical review of the manuscript.
M. G. M. has contributed in a, b, d; J. C. R. in c, d, e; E. M. M. in b, d, e; K. M. M. in a, b, d.

**Scientific editor in charge:** Dra. Cecilia Cracco.