Development and validation of a scale for measuring intercultural empathy

Desarrollo y validación de una escala de medida de la empatía intercultural

González-González, Hugo; Álvarez-Castillo, José-Luis & Fernández-Caminero, Gemma
Universidad de Córdoba

Resumen
La investigación y la práctica relacionadas con el uso de estrategias empáticas en entornos educativos multiculturales se está revelando como una de las más eficaces vías para la lucha contra el sesgo intergrupal. El objeto del presente estudio es diseñar un instrumento de medición de la empatía intercultural en su componente afectivo, específico para la población española y hacia la población de origen marroquí, que permita diagnosticar este rasgo o capacidad antes de aplicar estrategias educativas en entornos multiculturales. Con este propósito se utilizaron dos muestras de estudiantes universitarios para realizar una validación cruzada y, mediante análisis factorial exploratorio (utilizando el software Factor) y confirmatorio (SPSS-AMOS), se obtuvo una estructura de 3 factores: preocupación empática, comprensión y afecto. Los índices de bondad de ajuste, así como la validez y fiabilidad revelada por los análisis multidimensionales, demuestran que el modelo es óptimo. Finalmente se discute la conveniencia de diseñar instrumentos específicos en los que se tenga en consideración el perfil de las poblaciones objeto de estudio en ámbitos multiculturales.

Palabras clave:
Empatía; comunicación intercultural; validación; sesgo intergrupal

Abstract
The research and practice connected to the use of empathic strategies in multicultural educational settings is emerging as one of the most effective ways to combat intergroup bias. The aim of this study is to design an instrument to measure the specific intercultural empathy in its affective component for the Spanish population towards the population of Moroccan origin. Thus, this instrument makes possible the measurement of this feature or capability before implementing educational strategies in multicultural contexts. Therefore, two samples of University students were used to carry out a cross validity study through exploratory (Factor 9.20) and confirmatory factor analysis (SPSS-AMOS). Then, a 3-factor structure was obtained: empathic concern, understanding and affection. The goodness of fit indices, together with the validity and reliability revealed by multidimensional analysis prove that the model is optimal. Eventually, it is discussed the desirability of designing specific instruments that take into account the profile of the populations under this study.

Keywords:
Empathy; intercultural communication; validity; intergroup bias

Fecha de recepción
21 de Diciembre 2015
Fecha de aprobación
27 de Noviembre 2015
Fecha de publicación
30 de Noviembre 2015

Reception Date
2014 December 21
Approval Date
2015 November 27
Publication Date:
2015 November 30

Autor de contacto / Corresponding author
González-González, Hugo. Departamento de Educación. Universidad de Córdoba. Av. San Alerto Magno s/n 14074 - Córdoba.(Spain) hugo.gonzalez@uco.es
As far as the empathy is concerned, the literature about the topic considers an important diversity in relation to the delimitation of this construct (Fernández-Pinto, López-Pérez & Márquez, 2008; Gerdes, Segal & Lietz, 2010). The empathy has been defined by Hoffman (1975, 1981, 1982, 1983) and Strayer & Eisenberg (1987) as the affective experience of other person’s feelings. Nevertheless, the revised studies show that it has to do with a construct that can admit different approaches from a variety of perspectives which have similar and complementary components that can be particularly useful in a study from different points of view, that is, from a psychological, sociological, intercultural, pathological perspective or taking into account the gender, religion, etc. (Albiero & Matricardi, 2013; Barnett & Mann, 2013; Fernández-Pinto et al., 2008; Gerdes et al., 2010; Hardy, Walker, Rackham & Olsen, 2012; Harris & Picchioni, 2013; Wang et al., 2003).

In spite of the huge variety of approaches, a certain degree of consensus has been achieved in relation to the determination of some behavioral correlate of the empathy. In recent years, its importance in relation to the pro-social attitude of people has been highlighted (Belacchi & Farina, 2012; Butrus & Witenberg, 2013; Hodges, Clark & Myers, 2011; Shen, Carlo & Knight, 2013; Welp & Brown, 2014), as well as its role played in social conflicts (Barnett & Mann, 2013; Sanmartín, Carbonell & Banos, 2011; Zembylas, 2013). In a review concerning the different studies about this issue, Eisenberg (2000) considers the relevance of the empathy in the moral development, understood as an emotional response that comes from the comprehension of the other person’s situation, with the effect of experiencing the other person’s similar feelings. Therefore, the empathic response includes the capacity to understand the other person and to put oneself in the other person’s shoes, using observation, verbal information or other type of information approachable from the memory (perspective-taking), by covering the affective reaction produced when an emotional condition is shared. This can generate sadness, discomfort or anxiety. Empathy, understood in this way, would play a central role in the pro-social attitude of people (Eisenberg, 2000). Recently, Li, Mai and Liu (2014) have checked the research about the empathy and other social fields: emotion, mind theory or “mentalizing” and moral judgments. But, in this occasion, they have done so from a perspective based on the running of the neurological level and on the interconnection between the different areas of the human brain. In this research, it is evident how the experimentation with neuro-image has contributed to demonstrate that we use our memories and mental associations of past experiences as the pillars to understand the emotions and the cognitive conditions of the others.

If we go beyond and also consider the link between empathy and prejudice, social exclusion and intergroup explicit and implicit attitudes (Albiero & Matricardi, 2013; Li, Mai & Liu, 2014; Shih, Stotzer, & Gutierrez, 2013; Shi, Trahan, Wang & Stotzer, 2009), we will understand better the reasoning used to stand up for the research and practice related to the use of empathic strategies in multicultural educational environments (Belacchi y Farina, 2012; Numata, 2013) as well as the use of different programmes of intercultural education with the goal of increasing the empathy (Peek & Park, 2013; Todd, Bodenhausen & Galinsky, 2012). In conclusion, the empathy is a key concept in the establishment of social relationships. This is why it is perceived as an educational need in the intercultural contexts of our schools and society.

Therefore, given the socio-educational relevance of the empathy, we miss the existence of easy and high-quality tools, valid and reliable, aimed at facilitating its measurement in research and intercultural educational settings. The present study emerges from the identification of this need. Its main objective is to contribute to the
diagnosis and evaluation of empathy, together with the facilitator effect on the planning of educational interventions which can be more faithful to the intercultural approach in diverse environments.

**Measurement of empathy**

The common empathy has been evaluated by means of different types of instruments. So, in their review, Fernández-Pinto and his colleagues (2008) analyze the following typology of measures: some questionnaires have a unified perspective (IRI, EQ, TECA); others are focused on the affective dimension (QMEE, BEES, MEE) and others are centered around the cognitive component, whose instruments used in these measurements are older than the techniques used in the first two categories (DRTIE in 1949 & EM in 1969). Concerning their structures, all these instruments are composed of a number of sub-scales that varies between 2 and 7.

In the most recent ones -IRI, EQ y TECA-, it is observed that the analyzed dimensions adopt similar names, starting with the index of interpersonal reactivity (IRI) (Davis, 1980, 1983). According to Mestre Escrivá, Frías Navarro, and Samper García (2004), this index is one of the most used. Its 4 sub-scales contain cognitive and emotional factors, that is, perspective taking (PT), fantasy (FS), empathic concern (EC) and personal discomfort (PD). The empathy quotient (EQ) by Baron-Cohen & Wheelwright (2004) also includes the multidimensional perspective with two sub-scales called Cognitive and Emotional Reactivity. Another one has to be added, named Social Abilities. To conclude, it has also been developed the Test of Cognitive and Affective Empathy (TECA) (López-Pérez, Fernández-Pinto & Abad, 2008) from an integrating perspective addressed to the Spanish speaking population. It takes into account, inside the cognitive sub-scales, the Emotional Comprehension. This sub-scale tries to fuse both spheres, the cognitive and the emotional.

To sum up, it is observed in the instruments a tendency towards the measurement of both dimensions, affective and cognitive, in contrast to the previous tests, which seemed to be focused only on one of these. Nevertheless, it is obvious that there is no agreement concerning the measurement of empathy. This is connected to the issue about the delimitation of the construct (Fernández-Pinto et al., 2008; Gerdes et al., 2010).

On the basis of this premise and considering the field of practice on which our research focuses, that is, intercultural education, it is important to face the design of a high-quality measure of empathy to be used in multicultural contexts, since it is well-known its predictive capacity related to the stereotypes and prejudices in a field of cultural diversity (González, 2011; González, Álvarez & Fernández, 2012; Shih, Stotzer & Gutierrez, 2013). Considering the importance of the different components of intergroup bias (stereotypes are centered around beliefs, while prejudices are more concerned with the affective side), as well as their relationship with the strategies for the control of bias, tested in previous studies, and the characteristics of the intergroup bias in the Spanish population, (Álvarez, 2005; González, 2011; Álvarez, Palmero & Jiménez, 2011; Álvarez, Jiménez, Palmero & González, 2014), it is necessary an approximation oriented towards the affective component of empathy and its measurement in multicultural contexts, where the bias towards the different out-groups is unequal.

**Intercultural empathy**

Although the empathy has been the object of study of numerous studies from different perspectives, when it is analyzed from a cultural or ethnic approach, we find that this field of research is barely investigated (Green, 1998; Dyche & Zayas, 2001; Rasoal, Eklund, Hansen, 2011). Here, the construct has not been labelled or operationalized in the same way.

Perhaps Ridley and Lingle (1996) are the first researchers who used and defined the concept of cultural empathy. This construct would exceed the concept of general empathy,
including comprehension and the acceptance of the other person’s culture. These authors state that the culture establishes differences, which are usually outstanding regarding values and expectations, so that the empathic response involves the mutual comprehension in relation to these cultural differences. In the same way, it is also shown that the same empathy is associated to cultural differences. That is to say, the level of emotional adjustment and the level of empathic interest for other person will vary depending on the own culture (Trommsdorf, Friedlmeyer & Mayer, 2007). This would also validate the concept of cultural empathy. However, in spite of the potentiality of this kind of contextualized empathy, the research about it has been scarce (Howe, 2013).

Wang et al. (2003), who are aware of the importance of the cultural and ethnic components, developed the concept of ethnocultural empathy, which is similar to cultural empathy. This concept is also related to the concepts of cultural competence and cross-cultural empathy (Dyche & Zayas, 2001; Green, 1998; Wang et al., 2003). Wang and his colleagues understood the ethnocultural empathy as the empathy expressed towards members of cultural, ethnic or racial groups which are different from one's own. Advances in this kind of empathy would involve the reduction of intolerance, discrimination and conflicts, and, at the same time, understanding and mutual respect would progress, regardless the cultural or ethnic membership. It is a more difficult capacity to develop than the interpersonal empathy, because it means assuming the perspective of a person perceived as an out-group member. Therefore, the other person must be regarded in his/her cultural context to be able to adopt his/her perspective—this will be the first distinguishing feature of the ethnocultural empathy. Secondly, the ethnocultural empathy implies the self-control of one's own prejudices towards a person belonging to a different cultural or ethnic group. Finally, the third relevant feature of this type of empathic capacity is that this one is not independent from the previous experience that a person has of the other culture.

Taking this concept as a starting point, Wang et al. (2003) developed a scale of ethnocultural empathy whose validity and reliability are still under scrutiny. They identified the following factors: Empathic Feeling and Expression, Empathic Perspective Taking, Acceptance of Cultural Differences, and Empathic Awareness. Albiero and Matricardi (2013), apart from confirming the appropriate validity and reliability of the scale designed by Wang et al. (2003) by means of multivariate analysis methods, they found out a moderate association with the scale of general empathy and a strong negative relationship with the measurement of prejudice. These types of attitudinal links support the associations verified in other studies between general empathy and prejudice attitudes (Batson et al., 1997; Shih et al., 2009, 2013).

Having this tool as a referent and considering the peculiarity of the Spanish context, the validation of a technique able to assess specifically the intercultural empathy was tackled. This instrument had to be used in different educational contexts, and it had to be quickly implemented and easily interpreted. From an intercultural point of view, it is obvious that measures have to be associated to the socio-cultural contexts where they are used, because they can produce outstanding differences in the levels of empathy and in the predictive capacity of this variable (Howe, 2013). This is the reason why the Moroccan group has been chosen as the target of empathy in the tool to be designed. The Moroccan group is the most numerous non-European minority in Spain. Besides, they are considered as the typical example of the migratory phenomenon by the majority of the Spanish population. At the same time, migratory phenomenon is associated to stereotypes and prejudices. When the Centre of Sociological Investigations (CIS, 2014) asked Spanish citizens which group of immigrants came to their minds first,
Moroccans were mentioned. This group was the one who inspired less friendliness. They were only surpassed by the Romanian immigrants.

So, taking this situation into account, together with the important relationship between empathy and prejudice, a method used to carry out an instrument to measure the empathy towards the Moroccan population is going to be presented right below. The affective component will be especially emphasized.

**Method**

**Sample**

The survey-based research included two samples with the aim of carrying out a cross-validation; the first one was used to implement an exploratory study and the second sample to implement a confirmatory one. 821 university students of education degrees (degrees in Social Education and Psychopedagogy, that were about to expire, and degrees in Primary Education and Early Childhood Education) participated in both studies. The participants were selected among the available students who attended lectures. The samples, both exploratory and confirmatory, belonged on a similar average to the provinces of Córdoba (48.3%) and Burgos (51.7%). The two cities were chosen because they represented different cultural contexts and they have an unequal representation of the Moroccan population, what can make a difference in intercultural empathy. Particularly, the Moroccan population represents 0.35% of the population in Córdoba and 0.75% in Burgos (INE, 2015). However, significant differences could not be confirmed between groups in any of the dimensions of intercultural empathy, as it is observed later in table 1 on data collected from the initial total sample. This invariance facilitated the total analysis of the data.

The first sample, which is pilot, was used to test the index of discrimination of the items and to analyze the factorial structure of the instrument. It was composed of 250 students of education degrees from the Universities of Córdoba and Burgos. The 73.2% of them were women and the 26.8% were men. The average age of this group was 22 years old, with a typical deviation of 4.46.

On the other hand, the sample of participants involved in the confirmatory study consisted of 571 people, 69.9% were women and 30.1% were men. The average age was 22.2 years old, with a typical deviation of 4.81.

**Instrument**

The instrument was designed *ad hoc* according to the inferences derived from the theoretical frame and following the instructions by Zhou, Valiente and Eisenberg (2003) in their revision about the different methods used to assess empathy (see Batson, 1987). So, it was measured the evaluative component of the bias towards the members of a specific collective, who were identified by means of images which were presented on a computer screen.

The Direct RT software by Empirisoft was the one chosen to collect responses to 20 critical items, preceded by some instructions and questions about socio-demographic issues: age, gender and some other items aimed at excluding the participants of the analysis in case of a positive answer, that is, the existence of a relationship between a participant and an immigrant person in any context: family, neighbours, partners, classmates, colleagues. Besides, the frequency and length of these contacts were also taken into account.

In the instrument, the students were asked, in the 20 items, to evaluate the degree in which several emotions were experimented by them, while they observed a photo on a computer. The images portrayed faces that were prototypical of the exogroup members (Moroccan immigrants). These 20 images were selected from the prototypical scoring that 4 judges gave to 100 photographs taken from a database of Moroccan immigrants’ faces. These photographs had been previously filtered to unify the format, that is, their size, colour, brightness and perspective. The judges, 2 teachers from Córdoba University and 2...
teachers from Burgos University, were chosen due to their experience in the field of intercultural education. They were asked to give a distinctiveness score of the people shown in the photos in relation to the category "Moroccan" in a 5-point rating scale (A = identical appearance to a Moroccan person; B = similar appearance; C = vague appearance; D = a slightly different appearance to a Moroccan person; E = appearance totally different from a Moroccan person). The photos which did not achieve the total agreement of the 4 judges in the A level were rejected and, among the photos which fulfilled this requirement, those of best quality and definition were selected.

The self-assessed emotions were five: being moved, sympathy, tenderness, affection and compassion (see Batson, 1991; Vescio, Sechrist & Paolucci, 2003). Each participant was shown four images for each one of the emotions asked to self-assess. Ratings of the emotional intensity experimented by the participant were made on 7-point Likert scales (1= nothing at all, and 7=extremely). The order of appearance of the images was at random for each participant, with the aim of controlling possible biases due to the position of the image in the item sequence.

![Image](image.png)

Figure 1. Example of one of the test items.

**Data analysis**

According to their answers in the questionnaire, 8 participants were dismissed in the exploratory analysis because they were emotionally involved with the immigrant collective. The following step was to normalize the sample. Firstly, the outliers detected by means of the Mahalanobis test were eliminated. It was carried out with the help of AMOS 22.0 software. Finally, the exploratory sample had 211 participants, 154 women and 57 men, with a mean age of 21.80 and a standard deviation of 4.27.

During the exploratory analysis, the SPSS 22.0 was used to calculate descriptive statistics, as well as to analyze the properties of the items. It was verified that the parameters of univariate and multivariate normality of the distribution were appropriate (Byrne, 2012; González, Abad & Lévy, 2006; Kline, 2011). Next, it was carried out the exploratory factor analysis by means of the FACTOR programme (Lorenzo-Seva & Ferrando, 2006). It allowed us to do the analysis with polychoric correlation matrices, using PA-MBS as the procedure to determine the number of factors, MRFA as the method to extract the factors, and Promin as the method of rotation.

Concerning the confirmatory factor analysis, first of all, the 17 participants of the second sample who had an affective link with the immigrants were dismissed. Similarly, the outliers of the sample were eliminated by
means of the Mahalanobis test (AMOS 22.0). Consequently, the number of participants was reduced to 482, 149 men and 333 women, with a mean age of 22.08 and a standard deviation of 4.76. Models of structural equations with the programme AMOS 22.0 were used. It was evaluated the adjustment of the model by means of the statistics of goodness of fit that the majority of authors stand up for. According to Byrne (2001), these are the following: the test $\chi^2$ and the reason $\chi^2$/degrees of freedom, the Comparative Fit Index (CFI) (Bentler, 1990), the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI), the Tucker-Lewis Index (TLI), the Root Mean Square Residual (RMR/RMSE), the Root Mean Square Error of Approximation (RMSEA), the Akaike Information Criterion (AIC) and the Expected Cross-Validation Index (ECVI).

Then, the reliability of each of the dimensions was obtained by means of the Cronbach's alpha coefficient. Furthermore, in order to analyse the validity and reliability of the instrument, the indexes recommended in the literature were calculated using AMOS 22.0 software: Composite Reliability (CR), Maximum Shared Variance (MSV), and Average Shared Variance (ASV). With the help of them, the reliability, convergent validity and discriminant validity were established.

Finally, using SPSS and AMOS 22.0, it was carried out on the whole sample a research about gender differences in the dimensions of the validated questionnaire.

**Results**

Firstly, as it is observed in table 1, means point out average levels of empathy, with an inclination towards middle-high levels in the five initial factors. Moreover, it is verified that the participants from the two geographical origins, in spite of the assumed contextual differences, experienced the five emphatic emotions with an equivalent intensity. Similarly, the measurements in the emotional reactions did not differ in the samples of both studies (exploratory and confirmatory).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Sample/ residence</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t(819)</th>
<th>p</th>
<th>Sample/ study</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t(819)</th>
<th>p</th>
<th>Global mean</th>
<th>Global standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathy</td>
<td>Burgos</td>
<td>5.099</td>
<td>1.105</td>
<td>0.330</td>
<td>.741</td>
<td>Exploratory</td>
<td>5.158</td>
<td>1.155</td>
<td>1.147</td>
<td>.252</td>
<td>5.087</td>
<td>1.167</td>
</tr>
<tr>
<td></td>
<td>Cordoba</td>
<td>5.072</td>
<td>1.241</td>
<td></td>
<td></td>
<td>Confirmatory</td>
<td>5.056</td>
<td>1.172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassion</td>
<td>Burgos</td>
<td>4.618</td>
<td>1.216</td>
<td>-1.042</td>
<td>.298</td>
<td>Exploratory</td>
<td>4.598</td>
<td>1.298</td>
<td>-0.929</td>
<td>.353</td>
<td>4.661</td>
<td>1.300</td>
</tr>
<tr>
<td></td>
<td>Cordoba</td>
<td>4.714</td>
<td>1.396</td>
<td></td>
<td></td>
<td>Confirmatory</td>
<td>4.689</td>
<td>1.301</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being moved</td>
<td>Burgos</td>
<td>4.648</td>
<td>1.208</td>
<td>-1.644</td>
<td>.100</td>
<td>Exploratory</td>
<td>4.698</td>
<td>1.175</td>
<td>-1.190</td>
<td>.849</td>
<td>4.710</td>
<td>1.205</td>
</tr>
<tr>
<td></td>
<td>Cordoba</td>
<td>4.786</td>
<td>1.198</td>
<td></td>
<td></td>
<td>Confirmatory</td>
<td>4.715</td>
<td>1.219</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenderness</td>
<td>Burgos</td>
<td>4.622</td>
<td>1.284</td>
<td>-1.258</td>
<td>.209</td>
<td>Exploratory</td>
<td>4.767</td>
<td>1.287</td>
<td>1.308</td>
<td>.191</td>
<td>4.675</td>
<td>1.332</td>
</tr>
<tr>
<td></td>
<td>Cordoba</td>
<td>4.739</td>
<td>1.387</td>
<td></td>
<td></td>
<td>Confirmatory</td>
<td>4.634</td>
<td>1.350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>Burgos</td>
<td>4.647</td>
<td>1.243</td>
<td>-1.518</td>
<td>.129</td>
<td>Exploratory</td>
<td>4.748</td>
<td>1.191</td>
<td>.635</td>
<td>.526</td>
<td>4.706</td>
<td>1.240</td>
</tr>
<tr>
<td></td>
<td>Cordoba</td>
<td>4.779</td>
<td>1.233</td>
<td></td>
<td></td>
<td>Confirmatory</td>
<td>4.688</td>
<td>1.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By means of the exploratory factor analysis it was verified that, in relation to the structure, the model of 5 factors showed appropriate values (tables 2 and 3). The saturation of items in the theoretical factors widely exceeded the threshold of .30. Besides, the discrimination indexes of the items (by means of the corrected item-total correlation, obtained with ViSta-CITA) and the estimated reliability of the factors (table 4) were of a high size. On the other hand, the correlation between factors (none of them reach .70)
allows us to anticipate that the factor structure is adequate. So, the inventory was finally constituted as it was initially planned: 20 items corresponding to 5 factors with 4 item each.

**Table 2.** Factor weights of the items in each factor and corrected total-item correlations

<table>
<thead>
<tr>
<th>Item</th>
<th>Sympathy</th>
<th>Compassion</th>
<th>Being moved</th>
<th>Tenderness</th>
<th>Affection</th>
<th>Corrected Item-Total correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.635</td>
<td>-.004</td>
<td>-.051</td>
<td>-.080</td>
<td>.184</td>
<td>.550</td>
</tr>
<tr>
<td>2</td>
<td>.732</td>
<td>.009</td>
<td>-.005</td>
<td>.012</td>
<td>.079</td>
<td>.603</td>
</tr>
<tr>
<td>3</td>
<td>.900</td>
<td>.014</td>
<td>-.042</td>
<td>.028</td>
<td>-.118</td>
<td>.551</td>
</tr>
<tr>
<td>4</td>
<td>.718</td>
<td>-.182</td>
<td>.020</td>
<td>.019</td>
<td>.166</td>
<td>.548</td>
</tr>
<tr>
<td>5</td>
<td>.147</td>
<td>.758</td>
<td>.071</td>
<td>-.147</td>
<td>-.095</td>
<td>.565</td>
</tr>
<tr>
<td>6</td>
<td>.016</td>
<td>.682</td>
<td>.131</td>
<td>-.079</td>
<td>.065</td>
<td>.650</td>
</tr>
<tr>
<td>7</td>
<td>.020</td>
<td>.773</td>
<td>-.039</td>
<td>.047</td>
<td>.033</td>
<td>.665</td>
</tr>
<tr>
<td>8</td>
<td>-.065</td>
<td>.752</td>
<td>-.112</td>
<td>.224</td>
<td>.008</td>
<td>.611</td>
</tr>
<tr>
<td>9</td>
<td>.043</td>
<td>.062</td>
<td>-.039</td>
<td>.034</td>
<td>.630</td>
<td>.565</td>
</tr>
<tr>
<td>10</td>
<td>.063</td>
<td>.044</td>
<td>-.103</td>
<td>-.059</td>
<td>.812</td>
<td>.571</td>
</tr>
<tr>
<td>11</td>
<td>-.048</td>
<td>-.066</td>
<td>.119</td>
<td>.022</td>
<td>.699</td>
<td>.591</td>
</tr>
<tr>
<td>12</td>
<td>.021</td>
<td>-.050</td>
<td>.080</td>
<td>.060</td>
<td>.569</td>
<td>.558</td>
</tr>
<tr>
<td>13</td>
<td>-.169</td>
<td>-.004</td>
<td>-.081</td>
<td>.740</td>
<td>.221</td>
<td>.482</td>
</tr>
<tr>
<td>14</td>
<td>.025</td>
<td>-.079</td>
<td>.091</td>
<td>.774</td>
<td>-.094</td>
<td>.566</td>
</tr>
<tr>
<td>15</td>
<td>.102</td>
<td>.012</td>
<td>.025</td>
<td>.674</td>
<td>-.048</td>
<td>.529</td>
</tr>
<tr>
<td>16</td>
<td>.111</td>
<td>.098</td>
<td>.016</td>
<td>.573</td>
<td>-.062</td>
<td>.529</td>
</tr>
<tr>
<td>17</td>
<td>.037</td>
<td>-.063</td>
<td>.819</td>
<td>.041</td>
<td>.083</td>
<td>.620</td>
</tr>
<tr>
<td>18</td>
<td>.053</td>
<td>.105</td>
<td>.687</td>
<td>-.122</td>
<td>.070</td>
<td>.643</td>
</tr>
<tr>
<td>19</td>
<td>-.073</td>
<td>.051</td>
<td>.662</td>
<td>-.053</td>
<td>.157</td>
<td>.641</td>
</tr>
<tr>
<td>20</td>
<td>-.049</td>
<td>-.078</td>
<td>.701</td>
<td>.174</td>
<td>-.090</td>
<td>.629</td>
</tr>
</tbody>
</table>

**Table 3.** Between-factor correlation matrix

<table>
<thead>
<tr>
<th>Factor</th>
<th>Sympathy</th>
<th>Compassion</th>
<th>Being moved</th>
<th>Tenderness</th>
<th>Affection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Sympathy</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Compass</td>
<td>.476</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Being moved</td>
<td>.564</td>
<td>.652</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Tenderness</td>
<td>.599</td>
<td>.439</td>
<td>.576</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>5 Affection</td>
<td>.599</td>
<td>.495</td>
<td>.659</td>
<td>.607</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The 5 factors were able to explain together the 84.48% of the total variance. It was achieved a residual mean of -0.002 and a variance of 0.001. The Root Mean Square Residual (RMSR) of the model was 0.028, which is considerably below of the average/mean value expected of the RMSR (0.069) for an acceptable model, according to Kelley's criterion (Kelley, 1935; see also Harman, 1962).

However, after starting the confirmatory factor analysis on the data of the second sample, the indexes of modification showed the existence of covariance between errors associated to items belonging to different factors. This circumstance made the model to be reformulated, as it can be observed in figure...
2, evolving from the 5 initial factors to only 3 (three of the initial factors were summarized in a more general one, "Empathic Concern"). On the other hand, after testing the covariance matrices of the standardized residuals, it was observed that all the values were under 2.58. Only one intersection of items (Tenderness A x Sympathy D) was near the mentioned threshold, with a value of 2.54. So, the items Sympathy D and Tenderness B were eliminated. The last one had two values which were inferior to the one mentioned above, but these values were also close to the established threshold. This data is another indicator of a good model adjustment (Joreskog & Sorbom, 1988; Byrne, 2001).

![Figure 2. Model of second-order factor structure of the questionnaire of Spanish-Moroccan intercultural empathy and standardized estimations of the regression weights](image)

The reliability coefficients of the factors and the test, as it is shown in table 4, hardly suffered any changes. Nevertheless, the statistics corresponding to the exploratory analysis were calculated again to confirm the adaptation of the new model with the three factors. However, this time a model with three factors was forced, generating results which were equally favourable in the explained variance, index of discrimination of the items,
estimated reliability of the factors and the test as a whole. Although the total variance explained by the factorial techniques was reduced to 74.84%, the reliability of each factor was kept beyond .80 (table 4) and the validation of the model, by using structural analysis, turned out to be successful, as it can be seen in table 5.

Table 4. Reliability coefficients of the factors and the test as a whole, in the models of 5, 4 and 3 factors

<table>
<thead>
<tr>
<th>5 factors Nº de items</th>
<th>4 factors Nº of items</th>
<th>3 factors Nº of items</th>
<th>α</th>
<th>α</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affection 4 0.809</td>
<td>Affection 4 0.809</td>
<td>Affection 4 0.870</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sympathy 4 0.823</td>
<td>Sympathy 4 0.823</td>
<td>Sympathy 3 0.910</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenderness 4 0.776</td>
<td>Tenderness 4 0.776</td>
<td>Empathic Concern 11 0.864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassion 4 0.860</td>
<td>Empathic Concern 8 0.895</td>
<td>Total 18 0.923</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being moved 4 0.839</td>
<td>Total 20 0.922</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total 20 0.922</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Adjustment indexes for the three-factor model

<table>
<thead>
<tr>
<th></th>
<th>χ²</th>
<th>df</th>
<th>p</th>
<th>X²/df</th>
<th>GFI</th>
<th>CFI</th>
<th>NNFI (TLI)</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>Superior (90%)</th>
<th>Inferior (90%)</th>
<th>PCLOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole sample (n=482)</td>
<td>470.28</td>
<td>303</td>
<td>.000</td>
<td>1.55</td>
<td>.950</td>
<td>.978</td>
<td>.967</td>
<td>.0361</td>
<td>.024</td>
<td>.028</td>
<td>.020</td>
<td>1.000</td>
</tr>
<tr>
<td>Men (n=149)</td>
<td>154.80</td>
<td>101</td>
<td>.000</td>
<td>1.53</td>
<td>.902</td>
<td>.960</td>
<td>.939</td>
<td>.0672</td>
<td>.060</td>
<td>.078</td>
<td>.040</td>
<td>.187</td>
</tr>
<tr>
<td>Women (n=333)</td>
<td>145.53</td>
<td>101</td>
<td>.002</td>
<td>1.44</td>
<td>.954</td>
<td>.982</td>
<td>.973</td>
<td>.0351</td>
<td>.036</td>
<td>.049</td>
<td>.022</td>
<td>.962</td>
</tr>
</tbody>
</table>

The resultant model fulfils the required adjustments and it improves the validity indexes achieved with the tested models of five or four factors. Specifically, in table 6, the validity and reliability coefficients which come from the regression analysis of the standardized weights and the correlations achieved with AMOS 22.0. are introduced. The results can be regarded as fairly acceptable, specially, taking into account the reduced number of items that each factor has.

Concerning the debate about the multi-dimensionality or uni-dimensionality of empathy, it is accepted that the answers to the questionnaire are described in a more appropriate way by a factorial and hierarchic structure. That is to say, three of the factors of first order (Compassion, Being moved and Tenderness) are explained by a unique second-order factor, which has been called "Empathic Concern", as it is regarded in the model presented in figure 2.
Table 6. Validity and reliability coefficients of the models of 5, 4 and 3 factors

<table>
<thead>
<tr>
<th></th>
<th>CR</th>
<th>AV</th>
<th>MSV</th>
<th>ASV</th>
<th>Reliability</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathy</td>
<td>0.82</td>
<td>0.60</td>
<td>0.37</td>
<td>0.35</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Affection</td>
<td>0.82</td>
<td>0.54</td>
<td>0.45</td>
<td>0.40</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Compassion</td>
<td>0.83</td>
<td>0.56</td>
<td>0.82</td>
<td>0.46</td>
<td>✓</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Being moved</td>
<td>0.82</td>
<td>0.52</td>
<td>0.82</td>
<td>0.50</td>
<td>✓</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Tenderness</td>
<td>0.71</td>
<td>0.45</td>
<td>0.43</td>
<td>0.37</td>
<td>✓</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Model of 5 factors

| Sympathy   | 0.80 | 0.57| 0.42 | 0.38 | ✓           |                      | ✓                     |
| Affection  | 0.83 | 0.55| 0.47 | 0.43 | ✓           |                      | ✓                     |
| Empathic Concern | 0.93 | 0.87| 0.47 | 0.41 | ✓           |                      | X                     |
| Tenderness | 0.74 | 0.48| 0.45 | 0.43 | ✓           |                      | X                     |

Model of 4 factors

| Sympathy   | 0.82 | 0.60| 0.45 | 0.40 | ✓           |                      | ✓                     |
| Affection  | 0.83 | 0.54| 0.52 | 0.44 | ✓           |                      | ✓                     |
| Empathic Concern | 0.89 | 0.72| 0.52 | 0.49 | ✓           |                      | ✓                     |

Model of 3 factors

With regard to the gender differences, in table 7, it is evidenced that the women get scores significantly greater than men in the sub-scale of Sympathy and in Empathic Concern.

Table 7. Gender differences in the dimensions of the questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Sympathy</td>
<td>4.85</td>
<td>5.19</td>
<td>1.22</td>
<td>1.24</td>
</tr>
<tr>
<td>Affection*</td>
<td>4.61</td>
<td>4.81</td>
<td>1.38</td>
<td>1.20</td>
</tr>
<tr>
<td>Empathic Concern</td>
<td>4.34</td>
<td>4.85</td>
<td>1.13</td>
<td>1.11</td>
</tr>
</tbody>
</table>

*In the Affection factor, equality of variance is not assumed in both groups (F = 5.339, p = .021)

Discussion

Against other studies which try to validate a general instrument in a specific population and that they do not consider the specific idiosyncrasy of the context where these studies are applied, this research has validated a measure of intercultural empathy, specifically designed to evaluate this variable in the Spanish population in relation to a minority but relevant social group, as it is the Moroccan one (CIS, 2014). On this point, our results show a solid validity and reliability, and a clear factorial and hierarchic structure which have allowed us to gather dimensions that are very close together from a conceptual point of view.

The strategy used to design this validated measure, which comes from the observations detailed in the theoretical frame of this study, is based on the necessity of building tools able to measure the emotional dimension of the bias, by means of the use, in the same instrument, of images or other elements that allow the participant to contextualize his/her feelings towards the out-group. In this sense, the designed measure represents an important advance with regard to other previous instruments (i.e, Wang et al., 2003; see also
Albiero & Matricardi, 2013), since it is shorter and includes images as vivid stimuli that facilitate semi-automatic responses about members of a minority group against whom the majority group is prejudiced.

With regard to the explained variance, our instrument achieves the 74.84% with only 18 items, while the exploratory analysis of the instrument by Wang and his colleagues maintains in its final version, 31 out of the 62 tested items, with an explained variance of 47%. On the other hand, the questionnaire published in 2003 provided evidence about reliability and validation based on test-retest, Cronbach's alpha and a concurrent and discriminating validity by means of a correlational analysis with other instruments, without providing information about the Composite Reliability (CR), Average Variance Extracted (AVE), Maximum Shared Variance (MSV) and Average Shared Variance (ASV). On this point, it is important to highlight that it is highly advisable when carrying out a factor analysis to calculate the validity and reliability using good quality statistics. Otherwise, even when a satisfactory adjustment of the model is achieved, the validity and reliability of the structure of factors and their items will not be demonstrated. So, methodological variations can be found in the base of the differences between the factor structures produced in the research by Wang et al. (2003) and ours. Besides, it cannot be dismissed the modulator nature of the cultural context from which the samples of both investigations come (Trommsdorff et al., 2007). That is to say, the adjective "intercultural" applied to empathy would make reference not only to the group who is the target of the emotional reaction, but also to the interaction of the cultural features of the majority group together with the target minority group. Here lies the distinctiveness of the measured construct.

Concerning the gender, it is important to underline that some differences have been found in two out of the three dimensions; sympathy and empathic concern (the last one is composed of three first-order factors: compassion, being moved, and tenderness). These differences show that women have higher levels of intercultural empathy. However, in the third dimension, affection, the null hypothesis could not be rejected. Although the data confirm this difference between genders (Hoffman, 1977; Lennon & Eisenberg, 1987; Rueckert, Branch & Doan, 2011), it is interesting to notice that Eisenberg and Lennon (1983) anticipated that, in the measurement of empathy, this type of differences depended on the design of the questionnaire. Nevertheless, these authors affirmed that the instruments that used techniques based on images or stories showed less gender differences, while those questionnaires based on self-reports, established fairly important differences between men and women, being women the most empathic ones. However, our research would also support the existence of differences in the first type of instruments.

Apart from other issues related to the design of the research or the questionnaire, there are some studies that, although they confirm the unimportant gender differences in empathy (Block, 1979; Maccoby & Jacklin, 1974), Rueckert et al. (2011), contend that it is a strong phenomenon in the literature, showing that women have higher levels of empathy. According to these authors, the difference in gender can be due to the emotional empathy, and it is evident thanks to the higher levels of emotional reactivity in women. Levels of empathy in women are more affected by their relationship with the other person. While, in their study, women showed higher levels of empathy towards the group of "friends" than the men, levels of empathy in women were meaningfully lower towards people considered as "enemies". Taking this into account, it is necessary to consider the fact that the group of women perceive the Moroccan out-group less threatening than the Spanish group of men does.

On the other hand, concerning the issue of gender, it is also important to highlight the evolutionary perspective. Van der Graaff et al.
(2014) examined in their conclusions two theories that justify gender differences in levels of empathy: the biological ones based on the different neurological maturations or hormonal differences which are consequences of the physical maturation, and the social ones, due to the expectations resulting from the different gender roles. According to the results achieved by these authors, this difference is noticeable during the adolescence period, and as our data confirm, it would be maintained and consolidated after this period.

In conclusion, both the instrument obtained and the resulting evidence are really interesting for future studies oriented towards the measurement of empathy in multicultural contexts. If it is considered the association between empathy and prejudice (Batson et al., 1997; Shih et al., 2009, 2013), it is important to bear them in mind in relation to the design of educational strategies aimed at the reduction or control of intergroup bias in diverse settings.

References


González-González, Hugo; Álvarez -Castillo, José -Luis & Fernández -Caminero, Gemma (2015). Development and validation of a scale for measuring intercultural empathy. RELIEVE, 21 (2), art. 3. DOI: http://dx.doi.org/10.7203/relieve.21.2.7841


**ACKNOWLEDGMENT**

This research comes from the conclusions obtained in the projects with references SEJ2005-00331 and BU030A08, which confirm the importance of demarcating in a better way the construct of empathy when it is measured in contexts of cultural diversity. This new conception is being applied in a R+D project, which is in progress. It is entitled "Service-Learning and Educational Innovation in Spanish Universities. A programme centered in the improvement of the academic performance and the social capital of the students" (Ref. EDU 2013-41687-R). The first two authors of the article participate in the three projects mentioned.
González-González, Hugo; Álvarez-Castillo, José-Luis & Fernández-Caminero, Gemma (2015). Development and validation of a scale for measuring intercultural empathy. RELIEVE, 21 (2), art. 3. DOI: http://dx.doi.org/10.7203/relieve.21.2.7841

Authors

González-González, Hugo (hugo.gonzalez@uco.es).
Assistant Professor of the Department of Education at the University of Cordoba (UCO). He is the corresponding author for this article. Member of the research group "Education, Diversity and Society" of the Andalusian Research Plan. Coordinator of Theory and History of the Education Department in the Faculty of Education at the UCO. Research areas: inclusive and intercultural education, moral education, social cognition and prejudice reduction strategies.
Postal address: Departamento de Educación. Universidad de Córdoba. Av. San Alerto Magno s/n 14074 - Córdoba.

Álvarez-Castillo, José-Luis (jlalvarez@uco.es).
Associate Professor of the Department of Education at the University of Cordoba. Director of the Doctoral Program in Education. Coordinator of the research group "Education, Diversity and Society" of the Andalusian Research Plan. It was until recently director of Teacher Training and Teaching Innovation at the University of Cordoba. Research areas: inclusive and intercultural education, social cognition and prejudice reduction strategies.
Postal address: Department of Education. University of Cordoba. Av. San Alerto Magno s/n 14074 - Cordoba (Spain).

Fernández-Caminero, Gemma (gemma.fernandez@uco.es).
Master in Inclusive Education. Professor of the Department of Education at the University of Cordoba (UCO). Member of the research group "Education, Diversity and Society" of the Andalusian Research Plan. Coordinator of the Degree of Social Education at the Faculty of Education Sciences of the UCO. Research areas: inclusive and intercultural education, social education, education for coexistence.
Postal address: Department of Education. University of Cordoba. Av. San Alerto Magno s/n 14074 - Cordoba (Spain).