

Dimensional structure and measurement invariance of the Interpersonal Reactivity Index (IRI) across gender

Beatriz Lucas-Molina¹, Alicia Pérez-Albéniz², Javier Ortuño-Sierra² and Eduardo Fonseca-Pedrero²

¹ Universitat de València and ² Universidad de La Rioja

Abstract

Background: The main goal of the present study was to examine the dimensional structure and measurement invariance of the Interpersonal Reactivity Index (IRI; Davis, 1980) across gender. **Method:** Two Spanish samples comprising 2,499 college students (71.2% women; M = 21.06 years) and 1,438 adults (57.8% women; M = 40.01 years) completed the IRI. **Results:** The study of the internal structure, using Exploratory Structural Equation Modeling (ESEM), revealed that the four-factor structure (Perspective Taking, Fantasy, Empathic Concern, and Personal Distress) fitted well to the college students' data; however, in the sample of adults all factorial models showed modest goodness-of-fit indexes. In addition, the results also supported the measurement invariance of the ESEM four-factor model across gender in college students. Women scored higher than men in all four subscales of the IRI, although differences were not significant for the PT scale. The reliability of the scores in this sample ranged from .72 to .79. **Conclusions:** The findings support the factorial validity of the IRI scores and suggest it is a useful instrument to measure self-reported empathy. Future studies should continue to examine the structure and measurement invariance of the IRI in adult populations and across cultures.

Keywords: Empathy, Interpersonal Reactivity Index (IRI), dimensional structure, measurement invariance, gender differences, college students, adults.

Resumen

Dimensionalidad e invarianza de la medida del Interpersonal Reactivity Index (IRI) en función del género. Antecedentes: el objetivo principal del presente estudio fue examinar la estructura dimensional y la invarianza de la medida en función del género del Índice de Reactividad Interpersonal (IRI, Davis, 1980). **Método:** dos muestras españolas de 2.499 estudiantes universitarios (71,2% mujeres, M = 21,06 años) y 1.438 adultos (57,8% mujeres, M = 40,01 años) completaron el IRI. **Resultados:** el análisis de la estructura interna, mediante modelos exploratorios de ecuaciones estructurales (ESEM), reveló que la estructura en cuatro factores (Toma de Perspectiva, Fantasía, Preocupación Empática y Malestar Personal) ajustaba bien en la muestra de universitarios. Sin embargo, en la muestra de adultos los modelos factoriales mostraron bajos índices de ajuste. Además, los resultados apoyaron la invarianza del modelo ESEM de cuatro factores según el género en los universitarios. Las mujeres puntuaron más alto que los hombres en las cuatro subescalas del IRI, aunque las diferencias no fueron significativas para la escala PT. La fiabilidad de las puntuaciones osciló entre 0,72 y 0,79. **Conclusiones:** los hallazgos apoyan la validez factorial de las puntuaciones del IRI y sugieren que es un instrumento útil para evaluar la empatía autoinformada. Futuros estudios deben continuar examinando su estructura e invarianza en poblaciones adultas y entre culturas.

Palabras clave: empatía, Interpersonal Reactivity Index (IRI), estructura dimensional, invarianza de la medida, diferencias de género, estudiantes universitarios, adultos.

Empathy is typically seen as a multidimensional construct that includes cognitive and affective components (e.g., Davis, 1980, 1983, 1996; Hoffman, 2000). According to this view, empathy refers to one's ability to recognize and identify what someone else is feeling (cognitive aspect) and to share that emotional state (affective aspect) in order to react properly to social situations. Indeed, empathy has been shown to be a key component of social interactions, as it promotes prosocial behavior while inhibiting aggressive behavior towards others (Hoffman, 2000; Jolliffe & Farrington, 2004).

The Interpersonal Reactivity Index (IRI; Davis, 1980) is one of the most widely-used self-report measures of dispositional empathy in adults. It is a tool made up of four separate subscales that assess the cognitive and affective aspects of empathy. The subscales Perspective Taking (PT; considering others' viewpoints) and Fantasy (FS; identifying with fictional characters in books and films) measure the cognitive components; whereas the affective components are measured by the Empathic Concern (EC; feelings of compassion and concern for others in need) and Personal Distress (PD; reaction of discomfort to others' distress) subscales.

Previous studies have offered support for the factorial validity of this four-factor model in American, Dutch, French, Spanish, and Swedish adults (Chrysikou & Thompson, 2015; Cliffordson, 2002; Gilet, Mella, Studer, Grünh, & Labouvie-Vief, 2013; Hawk et al., 2013; Pérez-Albéniz, De Paúl, Etxebarria, Montes, & Torres, 2003),

American, Chilean, and Spanish college students (Davis, 1983; Fernández, Dufey, & Kramp, 2011; Pérez-Albéniz et al., 2003) as well as in Dutch, Spanish and Swedish adolescents (Cliffordson, 2001; Escrivá, Frías, & Samper, 2004; Hawk et al., 2013). In addition, several authors have argued that these dimensions may yield a general empathy construct (Cliffordson, 2001). In this regard, some recent studies have found a second-order global empathy factor that accounts for the relationships between the IRI subscales in Dutch adolescents (Hawk et al., 2013) and Chilean college students (Fernández et al., 2011), as well as in Dutch and Swedish adults (Cliffordson, 2002; Hawk et al., 2013), indicating a hierarchical structure of the IRI. Hence, despite the extensive body of existing research on the subject, the underlying structure of the IRI scores remains unclear. Further analysis of its internal structure might, then, contribute to a better understanding of the structure of empathy using the IRI.

Due to this limitation and inconsistencies, and considering the emergence of new measurement models, such as Exploratory Structural Equation Modeling (ESEM) (Marsh, Morin, Parker, & Kaur, 2014), hypothesized models of the IRI need to be tested. For instance, an ESEM approach allows us to test less restrictive measurement models than those used in the traditional Confirmatory Factor Analysis (CFA) models (e.g., where all cross-loadings are constrained to zero). The ESEM model makes it possible to solve some of the problems associated with CFA, such as cases in which no satisfactory goodness-of-fit indexes are found or in which model modification (e.g., correlating error terms) is required (Marsh et al., 2014). All the factor loadings are estimated in the ESEM model, while specific restrictions are imposed on the parameters in CFA. This new measurement approach can contribute to a better understanding of the underlying factor structure of this measure of empathy.

Within the framework of empathy research, gender is one of the most frequently studied variables. Females have been widely found to be more empathic than males (Davis, 1983; Eisenberg & Fabes, 1990; Hoffman, 2000). Regarding the IRI, empirical research has generally shown that females score higher than males on all four subscales (Davis, 1983; De Corte et al., 2007; Hawk et al., 2013), although in some cases differences have not been found to be significant for the PT scale (Fernández et al., 2011; Gilet et al., 2013; Pérez-Albéniz et al., 2003).

Nevertheless, save for one piece of research (Hawk et al., 2013), none of these studies have examined an important aspect of scale measurement that has become a prerequisite for conducting cross-group comparisons (Byrne, 2012): establishing measurement invariance (MI). When comparisons between groups are made (e.g., by gender) it is typically assumed that the measurement instrument and the underlying psychological constructs behave similarly and have the same statistical significance in all the groups being compared. From a methodological standpoint, however, this assumption is untenable if measurement invariance is not tested beforehand. If the data do not hold MI, or if this invariance has not been tested, the validity of the inferences and interpretations drawn from the data could be completely erroneous or unfounded (Byrne, 2012). To the best of our knowledge, Hawk et al. (2013) have been the only researchers to examine whether the four-factor structure of the IRI and the global empathy structure were psychometrically invariant across gender. Specifically, they found both structures of the IRI to be equivalent for female and male adolescents in the Netherlands.

Within this research context, the main goal of the present study was to examine the dimensional structure and MI of the Interpersonal Reactivity Index (IRI; Davis, 1980) across gender in two large Spanish samples. The four-factor structure of the IRI was assessed using CFA and ESEM approach, whereas the second-order general empathy structure was tested using CFA. The MI of IRI scores across gender was also tested. We expected both factorial models to show an acceptable fit for these different samples. Moreover, we hypothesized that the factorial structure of the IRI would be equivalent across gender. In addition, we expected to find statistical differences in the mean scores of the IRI by gender. Finally, it is also hypothesized that IRI scores would show adequate internal consistency values. Addressing these issues might enable researchers to confidently include this measure in studies on empathy and its ties with a variety of psychological processes and outcomes, as well as to examine and compare the resulting patterns between genders.

Method

Participants

This study used two independent samples. Sample 1 consisted of 2,499 undergraduate students (719 men and 1,780 women) with a mean age of 21.06 years ($SD = 3.61$), from different degree programs (Computer Sciences, Psychology, Pedagogy, Law, Architecture, Chemistry, and Education) at two Spanish universities (University of the Basque Country and University of Santiago de Compostela). Sample 2 consisted of 1,438 adults (607 men and 831 women) with a mean age of 40.01 ($SD = 5.44$), recruited with the participation of 10 public and private schools in the Basque Country (Spain), so that the adults were part of the families of the children enrolled in these schools. Those participants who failed to respond to three or more items of the self-report were eliminated from the final sample.

Instruments

Interpersonal Reactivity Index (IRI; Davis, 1980). A Spanish version of the IRI (Pérez-Albéniz et al., 2003) was used. The IRI is a self-report questionnaire that measures perceived individual differences in the tendency to be empathic. It consists of 28 items on a five-point Likert-type response scale ranging from 0 (*Does not describe me well*) to 4 (*Describes me very well*). In the original Davis' version of the IRI, there are seven items to assess each of the four subscales: PT (e.g., "I try to look at everybody's side of disagreement before I make a decision"), EC (e.g., "I often have tender, concerned feelings for people less fortunate than I"), FS (e.g., "I really get involved with the feelings of the characters in a novel"), and PD (e.g., "In emergency situations, I feel apprehensive and ill-at-case"). The Spanish version of the IRI (Pérez-Albéniz et al., 2003) found that item 13 from the PD subscale ("When I see someone get hurt, I tend to remain calm") loaded better on the EC subscale. Thus, these subscales had six and eight items, respectively, rather than seven as in the original IRI version (Davis, 1980).

Procedure

In both samples, participants were informed about the voluntary nature of their participation and, after signing the consent form, were asked to complete anonymous questionnaires. They received

no type of incentive for taking part in the study. College students fulfilled the measurement instruments in a group session (10 to 50 students), during a standard hour-long class. Administration of the instruments was always under the supervision of a researcher. For the sample of adults, those who agreed to participate in the study were asked to take and complete the measures at home. Adults returned them to the school principal in sealed envelopes to ensure the privacy of their responses.

Data analysis

First of all, we conducted CFA and ESEM to examine the factor structure of the IRI in both samples. Due to the continuous nature and normal distribution of the data, the Maximum Likelihood procedure (ML) was used. Several fit indexes were computed: the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA) (and 90% confidence interval), and the standardized root mean residual (SRMR). The CFI and TLI values greater than .95 are preferred, values close to .90 are considered acceptable and values less than .08 for the SRMR indicates a good fit of the model. The RMSEA values should be under .08 for a reasonable fit, and under .05 for a good fit (Brown, 2006; Hu & Bentler, 1999).

Third, in order to study MI of the IRI across gender, we conducted successive multi-group CFAs. We tested three levels of group invariance, including configural, weak, and strong invariance. Configural invariance is supported by a good fit for an unconstrained multi-group baseline model in which factor loadings and intercepts vary between groups. Next step is to test weak invariance by comparing the baseline model with a model in which all factor loadings are simultaneously constrained across groups (Byrne, 2012; Cheung & Rensvold, 2002). We then established a strong invariance model, which contained cross-group equality constraints on all factor loadings and item intercepts.

In testing the invariance hypothesis, the change in CFI (Δ CFI) was used to determine whether the nested models were practically equivalent (Cheung & Rensvold, 2002). When Δ CFI is $\leq .01$ between two nested models, it is considered that all specified equal constraints are tenable. The confirmation of the equivalence of the intercepts permits comparison of the latent means in both groups.

Finally, Cronbach alpha was calculated as an estimation of the reliability of each IRI subscale for the student sample. The relatively few missing values in the data were replaced by regression-based estimates. IBM SPSS Statistics 22.0 (SPSS, 2013) and Mplus 7.0 (Muthén & Muthén, 2012) were used for data analysis.

Results

Sources of validity evidence based on the internal structure of the IRI

Following the Davis' model (1980, 1996), and guided by previous research on the IRI's dimensionality in Spanish college students and adults (Pérez-Albéniz et al., 2003), a four-factor model with the EC and PD scales containing eight and six items, respectively, was tested for both subsamples. As can be seen in Table 1, although the RMSEA and the SRMR indicated an acceptable fit of this model to the college students' data, the CFI and TLI suggested that caution should be exercised. For the adults' data, the fit of the four-factor model was less acceptable. In this sense, it is worth noting that some indexes, as it is the case of the CFI and TLI, can be affected and some authors have proposed the RMSEA as an even more relevant criterion of fit indexes (Marsh, Hau, & Wen, 2004). When ESEM model with four factors was tested for both samples, a better fit to the data was found, especially in the sample of college students. Factor correlations of the ESEM model were all statistically significant, ranging from .32 (FS and EC) to -.15 (PT and PD).

In addition, for this sample, the standardized factor loadings for the ESEM four-factor model were high and all statistically significant supporting the four-factor structure of Davis' model (1980, 1996) (see Table 2). Conversely, for the sample of adults, the standardized factor loadings patterns of the ESEM model, with the exception of the FS scale, differed from the established configuration of the IRI subscales.

A review of the modification indexes (MIs) and the expected parameter change values (EPC) of the ESEM four-factor model for the sample of college students showed four residual covariances between different pairs of items with the highest values. These pairs of items always loaded on the same factor (items 17 and 10, and items 27 and 24, on factor PD, items 12 and 7 on factor FS,

Table 1
Goodness-of-fit statistics resulting from the dimensional models tested

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
Four-factor model ¹							
College students	2,957.0	344	.823	.805	.055	.053, .057	.062
Adults	2,444.7	344	.729	.702	.065	.063, .068	.082
ESEM four-factor model							
College students	2,075.1	272	.894	.853	.052	.049, .054	.035
Adults	1,386.7	272	.881	.835	.053	.051, .056	.037
Four-factor second order model ²							
College students	3,043.1	346	.817	.800	.056	.054, .058	.067
Adults	2,491.3	346	.723	.697	.066	.063, .068	.085

Note: ¹ Pérez-Albéniz et al. (2003); ² Hawk et al. (2013); CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual

Table 2
Standardized factor loadings for ESEM four-factor model - College students

IRI Items	Factors			
	FS	EC	PT	PD
1	.317			
7	.427			
12	.504			
5	.659			
16	.715			
26	.736			
23	.803			
13		.392		
9		.407		
4		.427		
22		.449		
2		.538		
18		.541		
20		.560		
14		.605		
15			.307	
3			.337	
28			.550	
11			.572	
25			.612	
21			.677	
8			.697	
10				.340
17				.345
19				.393
6				.548
27				.758
24				.826

Note: Factor loadings under .30 have been omitted

and items 22 and 20 on factor EC). Such covariances can be due to item content overlap (Byrne, 2012), which appears to be the case here. When a second ESEM four-factor model was specified in which these parameters were freely estimated, fit indexes showed a considerable improvement, χ^2 ($df = 268$) = 1,544.3; $p < .001$; CFI = .925, TLI = .894, RMSEA = .044, 90% CI [.042, .046], SRMR = .030. Nonetheless, based on parsimony and methodological

criteria, as well as on recent criticisms on model specification (Heene, Hilbert, Freudenthaler, & Bühner, 2012), this model was not considered for further analysis.

In addition, a four-factor second order model was tested, for each sample, in which all IRI latent factors loaded onto a secondary global empathy factor. The goodness-of-fit indexes for this model were also inadequate (see Table 1), with CFIs below the standard cut-off, especially for the sample of adults.

Measurement invariance of the IRI scores across gender

Due to the poor fit of the previous CFA and ESEM models, the sample of adults was excluded from further analysis. Similarly, no additional analysis for the testing of MI across gender was conducted for the global empathy model. Thus, given that the four-factor ESEM model was the only one that evidenced a close to acceptable fit in the CFI and TLI goodness-of-fit indexes and acceptable in RMSEA and SRMR, we next assessed MI by gender for this model in the sample of college students. The results are shown in Table 4. The goodness-of-fit indexes obtained for both gender groups in RMSEA and SRMR were satisfactory, whereas CFI and TLI again were close to acceptable, for females, with a worst fit in the case of males. In the same way, the configural model in which no equality constraints were imposed showed an acceptable fit to the data attending to the indexes mentioned. Then we tested weak invariance for the two groups. The Δ CFI between the configural and weak models was $< .01$, indicating that the hypothesis of weak invariance was tenable. Subsequently, strong MI was tested, where the items' intercepts and factor loadings were constrained to be equal across groups. The Δ CFI between the constrained and the unconstrained models was below .01, indicating that strong invariance was supported according to the recommendations of Cheung & Rensvold (2002). Hence, overall the results support configural, weak, and strong invariance of the four-factor model across gender in the sample of college students, but again caution should be exercised as some goodness-of-fit indexes for the model were close to acceptable but still inadequate.

Test for latent means differences

Latent mean differences across gender in the sample of college students were estimated, fixing the latent mean values to zero in males (Marsh et al., 2014). For comparison between gender groups in the latent means, statistical significance was based on the z

Table 3
Goodness-of-fit indexes of measurement invariance across gender for ESEM four-factor model - College students

Model	χ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR	Δ CFI
<i>Gender</i>								
Males ($n=719$)	937.4	272	.862	.808	.058	.054, .062	.042	
Females ($n=1,780$)	1,434.4	272	.899	.860	.049	.047, .052	.035	
<i>Multiple group</i>								
Configural invariance	2,371.8	544	.888	.845	.052	.050, .054	.037	
Weak invariance	2,532.2	640	.884	.863	.049	.047, .051	.041	-.01
Strong invariance	2,679.2	664	.877	.860	.049	.047, .051	.044	-.01

Note: CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; SRMR = Standardized Root Mean Square Residual

statistic. Females scored higher than males in all four dimensions [FS (.336; $p < .001$), EC (.949; $p < .001$), PD (.461; $p < .001$)], although these differences were not statistically significant for the PT scale (.047; $p = .338$).

Reliability estimation of the IRI scores

To examine the reliability of the four subscales, we computed Cronbach's alpha coefficients. Scale score reliabilities of the IRI subscales were acceptable to good in the sample of college students: .72 for PD and EC, .74 for PT, and .79 for FS.

Discussion

The first goal of this study was to analyze the internal structure of the IRI in two Spanish samples to contribute to a better understanding of the nature of empathy construct using the IRI. To this end, we examined both the classic four-factor and the more recent hierarchical structure of the IRI among college students and adults. Our results indicated that the four-factor model proposed by Davis (1983, 1996) provided a modest fit to the data, especially to the adults' data. Other studies attempting to validate the factorial structure of the IRI through CFA approach in other languages, also reported a moderate fit of the four-factor model. Such is the case, for example, of the Chilean (Fernández et al., 2011), French (Gilet et al., 2013), and Spanish (Pérez-Albéniz et al., 2003) versions of the IRI. One might think of cultural differences, but such an assumption must be explored better in further studies. When new measurement models not tested in previous studies were explored, such as ESEM approach, a better fit to the data was found. In our study the four-factor model of the ESEM approach was the best model, but still questionable, attending to the goodness-of-fit indexes. These new measurement models performed may allow us to capture more in depth the complexity of empathy, as well as to solve several problems found using the CFA approach (e.g., restriction on the factor loadings) (Marsh et al., 2014). In addition, the four-factor structure had acceptable psychometric properties in the sample of college students, with acceptable to good scale score reliability, and adequate construct validity as highlighted by the correlations between dimensions. In addition, factor correlations were all statistically significant. Together, these results support Davis (1980, 1983, 1996) in that a valid measure of empathy should represent multiple constructs and that a multidimensional approach is needed in order to better understand and study empathy.

Regarding the four-factor second-order model, the goodness-of-fit indexes did not support the conclusion of previous studies (Cliffordson, 2001, 2002; Fernández et al., 2011; Hawk et al., 2013). Thus, although the second-order model seems to afford a more parsimonious description of the structure of the concept (Cliffordson, 2002), our study holds that empathy is a multidimensional construct made up of four different first-order factors.

The second aim of our study was to examine MI by gender on the four-factor and higher-order structure of the IRI, in order to further validate the measure and subsequently screen for gender

differences in empathy. This analysis was only conducted with the ESEM four-factor model for the sample of college students. Prior research has consistently reported higher scores for each dimension by females (Chrysikou & Thompson, 2015; Davis, 1983; De Corte et al., 2007; Hawk et al., 2013). However, with the exception of Hawk's (2013) work, these studies have not tested whether the IRI is psychometrically equivalent across gender. Consequently, if mean differences on latent empathy scores are found without testing MI, we cannot be sure that such results are due to true differences in the latent variable and not a measurement artifact. In our study, the results did support the MI of the ESEM four-factor model for college students. Thus, the number of factors and their loading patterns (configural invariance), the factor loadings (weak invariance), as well as the item intercepts (strong invariance), were the same for males and females college students. We then examined the latent means of the four subscales in the sample of college students and found that females obtained significantly higher scores than males on all subscales with the exception of the PT subscale. These results are very consistent with prior studies (Fernández et al., 2011; Gilet et al., 2013; Pérez-Albéniz et al., 2003) and thus, evidence of the construct validity of the IRI, although caution should be exercised, as the configural model for the female's and specifically the male's groups revealed some goodness-of-fit indexes close to acceptable but still inappropriate.

In conclusion, this study provides new evidence of the construct validity and equivalence between gender of the four-factor structure of the IRI for college students, making this instrument a useful tool for this population. It is worth noting that, to the best of our knowledge, previous studies have examined neither the structure of the IRI nor its MI by gender using the ESEM approach. In our study, the results obtained for the sample of adults, as well as for the global empathy model were inconclusive. Thus, further research would be necessary in order to clarify the structure and equivalence of the IRI among this population.

The findings of this study should be considered in light of several limitations. First, the IRI is a self-report measure for the assessment of empathy, and may be subject to social desirability biases. Thus, it would have been advantageous to include other measures of empathy or social desirability in order to investigate possible bias associated with self-report measures, as well as to test other aspects of the validity of the measure, such as discriminant validity. Second, the sample is limited to two convenience Spanish samples, which impacts the generalizability of the study results. Lastly, the cross-sectional nature of this study kept us from examining the stability and evolution of the four-factor and higher-order structures of the IRI over time.

Future longitudinal research might account for these limitations and test developmental trends in empathy and provide a basis for comparison with these cross-sectional results. Moreover, future research should consider studying the MI of the IRI across cultures, with a view to validating the comparability of its structure.

Acknowledgements

This work was partially supported by a grant of the Instituto de Estudios Riojanos (Resolución nº 56/2017).

References

- Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus. Basic concepts, applications, and programming*. New York: Routledge.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi: 10.1207/S15328007SEM0902_5
- Chrysikou, E. G., & Thompson, W. J. (2015). Assessing cognitive and affective empathy through the Interpersonal Reactivity Index: An argument against a two-factor model. *Assessment, 23*(6), 769-777. doi: 10.1177/1073191115599055
- Cliffordson, C. (2002). The hierarchical structure of empathy: Dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology, 43*, 49-59. doi: 10.1111/1467-9450.00268
- Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the Interpersonal Reactivity Index (IRI). *European Journal of Psychological Assessment, 17*, 36-47. doi: 10.1027//1015-5759.17.1.36
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology, 10*, 85, 1-17.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126. doi: 10.1037/0022-3514.44.1.113
- Davis, M. H. (1996). *Empathy. A social psychological approach*. Boulder: Westview Press.
- De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica, 47*, 235-260. doi: 10.5334/pb-47-4-235
- Eisenberg, N., & Fabes, R.A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion, 14*, 131-149. doi: 10.1007/BF00991640
- Escrivá, V., Frías, M. D., & Samper, P. (2004). La medida de la empatía: análisis del Interpersonal Reactivity Index [Measuring empathy: The Interpersonal Reactivity Index]. *Psicothema, 16*, 255-260.
- Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile. *European Journal of Psychological Assessment, 27*, 179-185. doi: 10.1027/1015-5759/a000065
- Gilet, A., Studer, J., Mella, N., Grün, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science, 45*, 42-48. doi: 10.1037/a0030425
- Hawk, S. T., Keijsers, L., Branje, S. J. T., Van der Graaf, J., De Wied, M., & Meeus, W. (2013). Examining the Interpersonal Reactivity Index (IRI) among early and late adolescents and their mothers. *Journal of Personality Assessment, 95*, 96-106. doi: 10.1080/00223891.2012.696080
- Heene, M., Hilbert, S., Freudenthaler, H. H., & Bühner, M. (2012). Sensitivity of SEM fit indexes with respect to violations of uncorrelated errors. *Structural Equation Modeling, 19*, 36-50. doi: 10.1080/10705511.2012.634710
- Hoffman, M. L. (2000). *Empathy and moral development. Implications for caring and justice*. Cambridge: University Press.
- Hu, L., & Bentler, P.M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 76-99). London: Sage.
- Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior, 9*, 441-476. doi: 10.1016/j.avb.2003.03.001
- Marsh, H.W., Hau, K.T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341. doi: 10.1207/s15328007sem1103_2
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110. doi: 10.1146/annurev-clinpsy-032813-153700
- Muthén, L.K., & Muthén, B.O. (2012). *MPlus: Statistical analysis with latent variables. User's guide*. Seven Edition. Los Angeles, CA: Muthén & Muthén.
- Pérez-Albéniz, A., De Paúl, J., Etxebarria, J., Montes, M. P., & Torres, E. (2003). Adaptación de Interpersonal Reactivity Index (IRI) al español [Spanish adaptation of the Interpersonal Reactivity Index]. *Psicothema, 15*, 267-272.
- SPSS (2013). *IBM SPSS Statistics for Windows. Version 22.0*. Armonk, NY: IBM Corp.