



# Measuring Empathy: A Meta-analytic Factor Analysis with Structural Equation Models (MASEM) of the Interpersonal Reactivity Index (IRI)

Giulia Raimondi<sup>1</sup> · Michela Balsamo<sup>2</sup> · Sjoerd J.H. Ebisch<sup>3</sup> · Massimo Continisio<sup>1</sup> · David Lester<sup>4</sup> · Aristide Saggino<sup>5</sup> · Marco Innamorati<sup>1</sup>

Accepted: 11 October 2023 / Published online: 25 October 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Empathy is an important social construct that has been defined in many ways by different authors, resulting in development of several questionnaires. The Interpersonal Reactivity Index (IRI) is one of the most used self-report scales to measure empathy in children, adolescents, and adults. However, studies have reported contradictory results about its factor structure. Therefore, the aim of the current study was to assess the dimensionality of the IRI through a meta-analytic structural equation modeling approach (MASEM). Eleven studies (total  $n=9470$ ) were included in the MASEM. The meta-analytic confirmatory factor analyses (CFAs) provided support for four of the tested models. A comparison of these models showed that the four-factor model proposed by Lucas-Molina et al. (2017) had the best fit. Overall, this MASEM suggests that the IRI provides a multidimensional, rather than a unidimensional, measurement of the empathy construct.

**Keywords** Empathy · Interpersonal Reactivity Index · Structural Equation Modelling · MASEM · Confirmatory Factor Analysis

## Introduction

The word “empathy” comes from the Greek word *empathia* (εμπάθεια), which refers to the physical experience of feeling intense emotions. However, the modern term “empathy” was coined by Titchener (1909) in the process of translating the German word *Einfühlung*, used by Theodor Lipps in his esthetic theory (Preston & De Waal, 2002). Both Titchener and Lipps thought that empathy occurs through a mechanism of inner imitation. Specifically, seeing another

person’s emotional state would cue the observer’s imitation process to mimic the other person’s emotion in themselves, resulting in a better understanding of the other’s mind.

Nowadays, many definitions of empathy have been proposed, which has resulted in the absence of a commonly shared definition of the construct (Cuff et al., 2016; Decety & Lamm, 2006; Elliott et al., 2011). Nevertheless, researchers agree in considering empathy as a social ability allowing the individual to understand and share emotional state of others, while keeping awareness of self-other distinction (Cohen & Strayer, 1996; Decety & Jackson, 2004).

The majority of authors and researchers also agree that empathy should be considered as a multidimensional construct, while disagreeing on the dimensions to include. According to one common proposal, empathy is composed of two dimensions, namely cognitive and affective empathy (Clark, 2010; Decety & Lamm, 2006; Mann & Barnett, 2013; Oliveira-Silva & Gonçalves, 2011). Cognitive empathy is defined as the ability to rationally understand the other person’s point of view and emotional states, whilst affective empathy is defined as the experience of sharing the other person’s emotional state and “feeling it” ourselves.

---

✉ Giulia Raimondi  
giulia.raimondi@unier.it

<sup>1</sup> Department of Human Sciences, European University of Rome, Rome, Italy

<sup>2</sup> Department of Psychological, Health, and Territorial Sciences, G. d’Annunzio University Chieti–Pescara, Chieti, Italy

<sup>3</sup> Department of Neuroscience, Imaging and Clinical Sciences, G. d’Annunzio University Chieti–Pescara, Chieti, Italy

<sup>4</sup> Stockton University, Galloway, NJ, USA

<sup>5</sup> Department of Medicine and Aging Sciences, G. d’Annunzio University Chieti–Pescara, Chieti, Italy

## The Davis Model of Empathy and the Interpersonal Reactivity Index

An important model of empathy was proposed by Davis (1980). According to Davis, empathy can be defined “as a set of constructs having to do with the responses of one individual to the experiences of another. These constructs specifically include the processes taking place within the observer and the affective and non-affective outcomes which result from those processes” (Davis, 1996; p. 12). In a typical situation where an individual is exposed to emotions of others, four related constructs are involved: (1) antecedents (i.e., characteristics of the individuals involved and especially the tendency of the observer to experience concern or distress to emotions of other); (2) processes (i.e., those processes influencing empathic outcomes; for example, mimicry or other more complex cognitive processes); (3) intrapersonal outcomes (i.e., cognitive, affective, and motivational responses to the empathic situation; for example, personal distress and empathic concern); and (4) interpersonal outcomes (i.e., behavioral responses directed to others; for example, helping behaviors or even aggression) (Davis, 1996). According to Davis (1980) a good scale assessing empathy should assess separately individual variations in cognitive and emotional empathy. Thus, it should assess differences in perspective-taking tendencies and emotional reactions typically experienced by the individual (Davis, 1980). To assess empathy according to his definition, Davis devised the Interpersonal Reactivity Index (IRI) (Davis, 1980), a 28-item self-report questionnaire assessing: (i) Perspective Taking, which refers to the ability to spontaneously take the other person’s point of view, anticipating how they are going to behave and react; (ii) Fantasy, the tendency of the person to strongly identify him/herself with characters in movies and/or books; (iii) Empathic Concern, which refers to the emotions that we experience which are cued by the other person’s emotional state, especially if they are going through unfortunate situations; and (iv) Personal Distress, the feeling of discomfort that a person experiences as result of seeing someone else going through negative situations. According to Davis (1980, 1983), despite the four dimensions/characteristics do not exhaust possible reactions to others’ emotions, the relationships among IRI’s subdimensions also support previous theories about the development of empathy. That is, greater perspective-taking is associated with greater empathic concern for others and less personal distress, while a tendency to fantasize about fictitious situations could influence empathic response to others (Davis, 1980, 1983). Although this model does not actually refer to Cognitive Empathy and Affective Empathy, its dimensions have been viewed in that model: Cognitive Empathy is

defined by Perspective Taking and Fantasy, whereas Affective Empathy is defined by Empathic Concern and Personal Distress (Gilet et al., 2013).

## Contradictory Results of the Factor Structure of the IRI

According to Davis (1980), a confirmatory factor analysis (CFA) supported the hypothesized four-factor model, although the fit indices were not reported in his study. To date, many international studies have evaluated the factor structure of the IRI, confirming its original four-factor structure (Albiero et al., 2006; Cliffordson, 2002; De Corte et al., 2007; Gilet et al., 2013; Himichi et al., 2017; Huang et al., 2012). Nevertheless, several other studies have suggested different factor models for the IRI: for example, hierarchical models (Cliffordson, 2002; Pulos et al., 2004), two-factor and one-factor models (Gilet et al., 2013), a three-factor model (Siu & Shek, 2005) and five-factor models (Ortiz et al., 2011).

Cliffordson (2002) confirmed the original four-factor model ( $\chi^2(318)=444.98$ ,  $p<0.001$ ; RMSEA=0.055, 90% CI=0.042–0.067). However, in order to better understand the theoretical nature of empathy, Cliffordson also tested a hierarchical model with a general second-order factor named empathy (i.e., EMP) which was loaded with all the specific dimensions (i.e., Perspective Taking, Fantasy, Empathic Concern and Personal Distress), which in turn were loaded with their specific items (Cliffordson, 2002). The hierarchical model had a good fit ( $\chi^2(320)=454.47$ ,  $p<0.001$ ; RMSEA=0.057, 90% CI=0.043–0.069) and supported the results of previous studies that described empathy as being hierarchically organized. These results were further confirmed by Hawk et al. (2013), who tested both the models identified by Davis (1980) and Cliffordson (2002) in two samples of adolescents, and reported a good fit in both groups for the original 4-factor model as well as for the hierarchical model (4-factor model: early adolescents, RMSEA=0.04; CFI=0.95; SRMR=0.04;  $\chi^2(48)=78.688$ ,  $p<0.01$ ; late adolescents, RMSEA=0.06; CFI=0.95; SRMR=0.05;  $\chi^2(48)=94.196$ ,  $p<0.001$ ) (hierarchical model: early adolescents, RMSEA=0.06; CFI=0.92; SRMR=0.06;  $\chi^2(48)=106.319$ ,  $p<0.001$ ; late adolescents, RMSEA=0.06; CFI=0.95; SRMR=0.05;  $\chi^2(48)=99.011$ ,  $p<0.001$ ). From a theoretical point of view, these results seem to suggest the presence of a broader and unitary construct of empathy which encompasses the four dimensions suggested by (Davis, 1980) rather than the presence of two separate components as the basis of the factor structure of the IRI (i.e., Cognitive Empathy and Affective Empathy).

A hierarchical model was also tested by Pulos et al. (2004), although the authors did not report fit indices for the

model. Their analyses suggested both the presence of the four original first-order dimensions and of two second-order factors, namely “General Empathy” and “Emotional Control”. The authors constructed a bifactor model in order to assess which dimension loaded on the second order factors. The results indicated that the General Empathy factor was loaded with the Perspective Taking, Fantasy and Empathic Concern subscales, while the Emotional Control factor was loaded with the Personal Distress subscale and partially loaded by some of the items of the Perspective Taking subscale. Nevertheless, this hierarchical model with two second-order factors was later found to have a poor fit in a new study by Chrysikou and Thompson (2016) (RMSEA=0.11; CFI=0.95; TLI=0.94). The authors also tested the original four-factor model proposed by Davis (1980), which also had suboptimal fit to the data (RMSEA=0.11, 90% CI=0.10–0.11; CFI=0.95; TLI=0.95).

Siu and Sheck (2005) suggested a 3-factor model (RMSEA=0.06; CFI=0.69; GFI=0.87; AGFI=0.86), with the 3 factors named fantasy, personal distress and empathy. However, it is relevant to note that the fantasy and personal distress scales did not perfectly resemble the original dimensions, and the empathy scale was not a perfect merge of the remaining Perspective Taking and Empathic Concern scales.

Lastly, Ortiz et al. (2011) reported findings for a 5-factor model ( $\chi^2(336)=832.43$ ,  $p<0.001$ ; RMSEA=0.06, 90% CI=0.04–0.07; GFI=0.93; AGFI=0.92). The authors did not find a perfect overlap with the original dimensions, and the original dimension Personal Distress was split into two different factors, namely Positive Personal Distress and Negative Personal Distress. Moreover, two of the five first-order factors loaded on a common second-order factor.

In summary, these results indicate a low agreement on the factor structure of the IRI. Furthermore, no supporting evidence for a two-factor structure of the IRI in terms of Cognitive Empathy and Affective Empathy dimensions has been reported. Since the IRI is one of the most often used scales to assess empathy, a better understanding of its dimensionality could help researchers to obtain more stable results and to be able to compare their results. Therefore, the aim of the current meta-analytic study is to evaluate the factor structure of the IRI through a Meta-Analytic factor analysis with Structural Equation Modeling (MASEM) (Cheung, 2008; Cheung & Chan, 2005).

Evaluating the dimensionality of the IRI with a MASEM approach can allow us to identify with more certainty the ‘correct’ numbers of dimensions of the IRI, compared to the evaluation of its dimensionality through single studies (Jak, 2015). Indeed, single studies usually tend to assess the dimensionality of questionnaires on relatively small samples, which could affect the fit of the model. In contrast,

in a MASEM, the number of participants of each study is summed, resulting in a large final sample associated with the pooled correlation (or covariance) matrix, so that stronger inferences can be made about the dimensionality of the IRI (Jak, 2015). Finally, taking into account the categorical nature of the rating scale of the IRI items, a robust estimator for categorical variables (e.g., unweighted least square) can be used to assess its dimensionality, since different estimators used in single studies can affect the model’s fit to the data and generate contradictory results.

## Methods

### Search Strategy and Coding Method

Publications from January 1980 to July 2023 reporting on the factor structure of the IRI were searched through the following databases: PubMed, PsycArticles, PsycINFO and Google Scholar. The following Boolean expression was entered into the databases: (TI “interpersonal reactivity index” OR TI “IRI” OR TI “empathy”) AND (TX “factor analysis” OR TX “confirmatory factor analysis” OR TX “exploratory factor analysis” OR TX “psychometric” OR TX “correlation” OR TX “principal component analysis” OR TX “dimensionality” OR TX “item analysis” OR TX “factor structure”). A total of 161 potentially usable studies were identified. Additional studies collected from the references of the identified articles were added. First, titles and abstracts were examined for relevance of the topic of the study and duplicates were removed. Subsequently, a full-text review of the studies was made. Only the articles that met the following inclusion criteria were used for the analyses: (i) psychometric studies that administered the 28-item version of the IRI (i.e., studies which either used a brief form of the IRI or deleted one or more items from the analyses were excluded); (ii) studies reporting either the inter-item correlation matrix of the IRI or the factor pattern matrices; (iii) studies conducted on samples from the general population; (iv) validation studies of the IRI (i.e., validation of the translated versions of the IRI in other languages); and (v) grey literature (i.e., conference paper, dissertation). In order to collect the studies that could not be found, the authors were contacted via e-mails. Seven studies were collected and went through the screening selection process, whilst the other 7 studies still could not be found (i.e., six authors did not reply and one author had difficulties retrieving the article and data). If studies conducted analyses and reported factor loadings for more than one sample, then each sample was considered separately from the others and entered in the analyses. For studies that tested multiple models, models with a reduced number of items were not taken into account.

Moreover, if studies reported multiple factor solutions for the same sample, the factor loading matrix with the highest number of factors was used for the analyses. The final number of studies included in the MASEM was 15 studies, comprising 17 samples. The detailed flow diagram of selection of studies is reported in Fig. 1.

The coding protocol used and the computer syntax for analyzing the data were the one used by Schroeders et al. (2021), and were adapted based-off the number of the items

of the IRI. Factor loadings of the 28 items of the IRI and Pearson correlations among the dimensions of the IRI were coded. Then studies (i.e., name of the first author, publication year and country), samples (i.e., sample size, mean age and percentage of female participants), and factor model information (i.e., Confirmatory Factor Analysis or Exploratory Factor Analysis, extraction method, numbers of extracted factors and rotation method) were reported in the coding sheet.

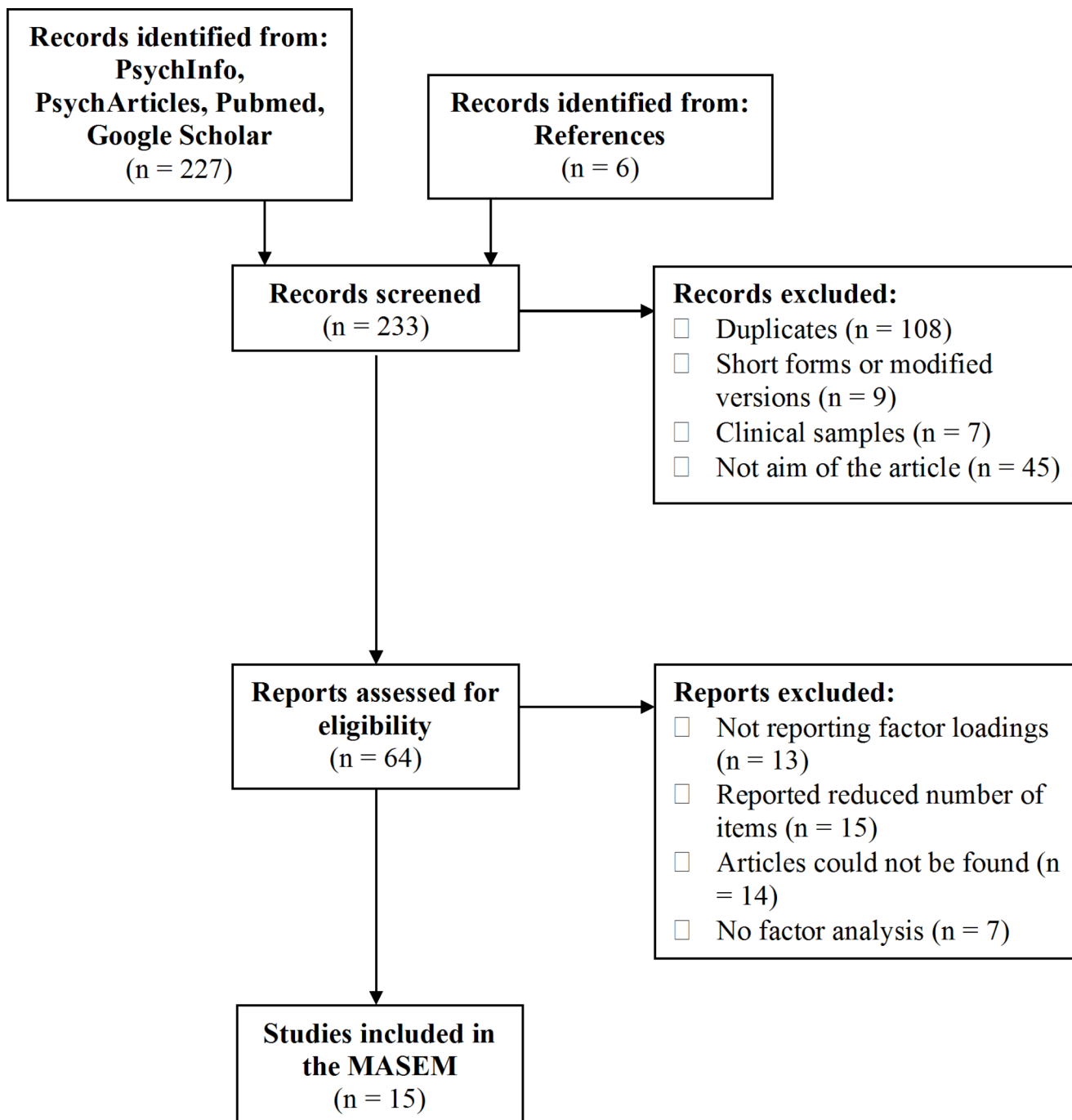


Fig. 1 Flow diagram of study selection. Abbreviation: IRI=Interpersonal Reactivity Index

Finally, the current study was approved by the Ethics Committee of the European University of Rome, Rome, Italy.

## MASEM Analysis

The MASEM was conducted with the metaSEM package (Cheung, 2021) for R studio (R Core Team., 2021) (version 4.1.1), which relies on OpenMx (Neale et al., 2016). To perform a MASEM, correlation (or covariance) matrices are needed. However, since correlation (or covariance) matrices are rarely reported in factor studies, we used the procedure described by Gnambs and Staufenbiel (2016). According to this procedure it is possible to reproduce model-implied inter-item correlation matrices from factors loadings and the factors correlations matrices. Following this procedure, when studies did not report factor loadings or correlations among factors, a value of 0 was implied. Then, a pooled correlation matrix was obtained using a fixed-effect method (FEM), which assumes that the population effect size is the same for each sample and that differences between effect sizes are due to sampling error (Cheung, 2014; Egger et al., 1997), and which is considered the most appropriate methodology when only a few samples are included in the analyses (Borenstein et al., 2009). Moreover, the random-effect model reported issues with the model estimation, which is probably due to the complexity of the model. Therefore, data (i.e., factor loadings and correlation among factors) were extracted from the 15 primary studies included and used to compute specific reproduced model-implied inter-item correlation matrices, which in turn were used to create the reproduced pooled correlation matrix.

In the second stage of the MASEM, the reproduced pooled correlation matrix was used to test multiple factor analytic models with Confirmatory Factor Analysis. Before testing different competing factor models, the Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) test were performed to assess whether the reproduced pooled matrix was suitable for factor analysis. Adequacy of the correlation matrix is suggested by a significant Bartlett's test ( $p < 0.05$ ) and a KMO index  $> 0.70$ . The following indices were used to assess all models' fit: (1) the Root Mean Square Error of Approximation (RMSEA), with values below 0.05 indicating evidence of absolute fit, values between 0.05 and 0.08 indicating adequacy of the model, and values above or equal to 0.10 indicating poor fit of the model (Browne & Cudeck, 1992; Hu & Bentler, 1999); (2) the Tucker Lewis Index (TLI), with values  $\geq 0.95$  indicating good fit of the model and values of 0.90 and higher an acceptable fit; (3) the Comparative Fit Index (CFI), with values  $\geq 0.95$  indicating good model fit and values of 0.90 and higher an acceptable fit; (4) the Standardized Root Mean square Residual (SRMR),

with values  $< 0.08$  indicating good fit (Yu, 2002); and (5) the chi-square ( $\chi^2$ ) test, with  $p$ -values greater than 0.05 indicating an adequate fit to the data. However, the chi square test is sensitive to sample size, and so  $p$  values might become significant for large samples (Schumacker & Lomax, 2012). Taking into account the categorical nature of the variables, a robust estimator (Weighted Least Square Mean and Variance adjusted, WLSMV) was used to test the different factor models. If more than one model showed good fit, a model-comparison was performed by using the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978), where smaller values indicate better model fit and McDonald's  $\omega$  (McDonald, 1999) was calculated as a reliability measure of the model(s) that reported the best fit.

## Results

### Studies Characteristics

The MASEM included a total of 15 studies and 17 samples. The total sample size was 11,381 participants, with a female percentage of 57.27% for the sample. The mean age of the sample was 24.12 years (SD = 10.51 years) (see Table 1 for studies and samples characteristics).

Firstly, fit information of all models tested in the 15 studies included were extracted and evaluated (see Table 2 for the description of all the models). The original four-factor model (Davis, 1980) was tested in 13 studies (Albiero et al., 2006; Cliffordson, 2002; De Corte et al., 2007; Gilet et al., 2013; Himichi et al., 2017; Huang et al., 2012; Lucas-Molina et al., 2017; Manarte & Andrade, 2020; Muchotrigro et al., 2022; Nomura & Akai, 2012; Ortiz et al., 2011; Pulos et al., 2004; Siu & Shek, 2005), reporting poor fit results in all of them, except in Ortiz et al. (2011). Besides the original model, 9 other models of the dimensionality of the IRI were identified in the studies included in the analyses (Cliffordson, 2002; Gilet et al., 2013; Holgado Tello et al., 2013; Lucas-Molina et al., 2017; Ortiz et al., 2011; Pulos et al., 2004; Siu & Shek, 2005). Specifically, a hierarchical model (Cliffordson, 2002), a one factor model (Gilet et al., 2013), a two-first-order factor model (Gilet et al., 2013), a five-first-order factor model (Holgado Tello et al., 2013), a four-first-order factor model (Lucas-Molina et al., 2017), a five-first-order factor model (Ortiz et al., 2011), a five-first-order factor model with a second-order factor (Ortiz et al., 2011), a hierarchical model with two second-order factors (Pulos et al., 2004), and a three-first-order factor model (Siu & Shek, 2005). (See figures in the supplemental material for the graphical representation of all the models). Even though some of these models did not indicate an adequate fit in

**Table 1** Studies and samples' characteristics

First author and year	Number of samples	Sample(s) size	Population	Mean age	Female percentage	Country	EFA/CFA	Rotation method	Number of factors extracted
Davis, 1980	2	• 579	University students	-	100%	America	CFA	-	4
	-	• 582		-	0%				
Albiero, 2006	1	• 828	Adolescents	14.75	52.77%	Italy	CFA	-	4
Cliffordson, 2002	1	• 125	University students	28.9	83.46%	Sweden	CFA	-	4 first-order factors and 1 s-order factor
De Corte, 2007	1	• 651	General population	24.48	54%	Belgium	CFA	Oblique	4
Gilet, 2013	1	• 322	General population	49.5	59%	France	CFA	-	4
Himichi, 2017	1	• 416	General population	39.61	50%	Japan	CFA	-	4
Holgado Tello, 2013	2	• 345	Children and adolescents	18.82	100%	Spain	CFA	-	5
		• 337			0%				
Huang, 2012	1	• 320	Teachers	30.52	66.25%	China	CFA	-	4
Lucas-Molina, 2017	1	• 3937	University students	30.53	66.31%	Spain	CFA	-	4
Manarte, 2018	1	• 275	Adults	25.8	50.18%	America	CFA	-	4
Muchotrigo, 2022	1	• 859	University students	20.36	77.30%	Perù	CFA	-	4
Nomura, 2012	1	• 95	University students	19.5	44.21%	Japan	CFA	-	4
Ortiz, 2011	1	• 721	Children and adolescents	12.82	47.90%	Spain	CFA	-	5
Pulos, 2004	1	• 409	University students	22.09	69.92%	America	CFA	Oblique	4
Siu, 2005	1	• 580	Adolescents and university students	17.4	52.24%	China	CFA	Oblique	3

**Note.** EFA, Exploratory Factor Analysis; CFA, Confirmatory Factor Analysis

“-”, information not reported

the primary studies (Gilet et al., 2013; Lucas-Molina et al., 2017; Siu & Shek, 2005) they were also tested in the second stage of the MASEM.

### Meta-analytic Confirmatory Factor Analysis

The pooled correlation matrix was adequate for factor analysis (Bartlett's test of sphericity = 42122.91;  $df=378$ ;  $p < 0.001$ ; KMO = 0.81).

Every model assessed in the studies included in the current work (see models figures in supplemental material) was tested through a Confirmatory Factor Analysis in the second stage of the MASEM (see Table 3 for the fit results of the competing models). The Confirmatory Factor Analyses demonstrated that, among the eight models tested, four models had a good fit: the original four-factor model (Model-1) (Davis, 1980), the four-factor model with a second-order global factor (Model-2) (Cliffordson, 2002), the four-factor model with item #13 loading on the Empathic Concern dimension (instead of Personal Distress) (Model-6) (Lucas-Molina et al., 2017) and the four-factor model with two second-order factors (Model-9) (Pulos et al., 2004). Although the other four models reported optimal values of RMSEA and SRMR, the CFI and TLI were below their cut-off values. Therefore, these models had to be rejected. A comparison among the four models that reported a good fit to the data was made

using the AIC and BIC indices. According to these indices, the model tested by Lucas-Molina et al. (2017) had the best fit. Finally,  $\omega$  values indicated a higher reliability of model-6 (Perspective Taking:  $\omega = 0.68$ , Fantasy:  $\omega = 0.73$ , Empathic Concern:  $\omega = 0.62$ , Personal Distress:  $\omega = 0.68$ ). Moreover, item #13 had a significantly higher loading on the Empathic Concern dimension ( $=0.21$ ) compared to its loading on the Personal Distress dimension ( $=0.04$ ), as the original model hypothesized (see Table 4).

### Discussion

The aim of the current work was to evaluate the factor structure of the IRI, one of the most used self-report measures of empathy. The Confirmatory Factor Analyses performed in the second stage of the MASEM indicated four models with the best fit to the data. These results partially supported the original four-factor model proposed by Davis (1980). However, the comparison between these models revealed that the model proposed by Lucas-Molina et al. (2017) described the data best and had the highest reliability values (i.e., McDonald's  $\omega$ ). In the original model (Davis, 1980), item #13 loaded on the Personal Distress dimension. However, when analyzing the content of the item (i.e., “When I see someone get hurt, I tend to remain calm”), it seems conceptually

**Table 2** Models' characteristics of the primary studies

First author and year	Number of models tested	Models description	$\chi^2$ or $\chi^2/df$	RMSEA (95% CI)	Models fit		
					CFI	SRMR	AIC
Davis, 1980	1	Four first-order factors*	-	-	-	-	-
Albiero, 2006	2	Original model	838.10 p < 0.001	0.04 (-)	CFI = 0.86	0.06	-
		Cliffordson's model	657.76 p < 0.001	0.04 (-)	CFI = 0.88	0.06	-
Cliffordson, 2002	2	Original model	454.47 p < 0.001	0.057 (0.043–0.069)	-	-	-
		Original model with a second-order factor	444.98 p < 0.001	0.055 (0.042–0.067)	-	-	-
De Corte, 2007	1	Original model	-	0.06 (0.05–0.06)	CFI = 0.86	0.06	1219.06
Gilet, 2013	3	One factor model	1897.21 p < 0.01	0.12 (0.11–0.13)	CFI = 0.33	0.14	-
		Original model	789 p < 0.01	0.065 (0.06–0.07)	CFI = 0.81	0.07	-
		Two first-order factors model	1734.45 p < 0.01	0.11 (0.11–0.12)	CFI = 0.40	0.14	-
Himichi, 2017	1	Original model	-	0.064 (-)	CFI = 0.83	-	-
Holgado Tello, 2013	1	Five first-order factors	1851.08 p < 0.001	0.07 (0.067–0.073)	CFI = 1 GFI = 0.91 NFI = 1 NNFI = 1.10	-	-
Huang, 2012	1	Original model	-	-	-	-	-
Lucas-Molina, 2017	2	Four first-order factors with item #13 loading on the EC dimension	2,975	0.55 (0.53–0.57)	CFI = 0.82 TLI = 0.80	0.62	-
		Original model	3,043	0.56 (0.54–0.58)	CFI = 0.81 TLI = 0.80	0.067	-
Manarte, 2018	1	Original model	2.13	0.06 (-)	CFI = 0.77 AGFI = 0.81	-	-
Muchotrigo, 2022	1	Original model	3895.29	0.13 (0.12–0.13)	TLI = 0.47	0.13	-
Nomura, 2012	1	Original model	-	-	GFI = 0.97 AGFI = 0.96	-	-
Ortiz, 2011	3	Five first-order factors with two cross-loading items	1033.70 p < 0.001	0.075 (0.070–0.080)	GFI = 0.92 AGFI = 0.91	-	-
		Five first-order factors with two cross-loading items and a second order factor for only two dimensions	832.43 p < 0.001	0.064 (0.049–0.070)	GFI = 0.93 AGFI = 0.92	-	-
		Original model	4947.47 p < 0.001	0.136 (0.13–0.14)	GFI = 0.68 AGFI = 0.61	-	-
Pulos, 2004	2	Original model	-	-	-	-	-
		Original model with two second-order factors	-	-	-	-	-
Siu, 2005	2	Original model	2471.30	0.06 (-)	CFI = 0.65 GFI = 0.88 AGFI = 0.85	-	-
		Three-factor model	2471.30	0.06 (-)	CFI = 0.69 GFI = 0.87 AGFI = 0.86	-	-

**Note.**  $\chi^2$ , chi squared; RMSEA, Root Mean Square Error Approximation; CFI, Comparative Fit Index; TLI, Tucker Lewis Index; SRMR, Standardized Root Mean Squared Residual; AIC, Akaike Information Criterion; EC, Empathic Concern

\* Original model

“-”, value not reported

**Table 3** Results of the competing models tested in the second stage of the MASEM

Model	$\chi^2$	<i>df</i>	RMSEA (95%CI)	CFI TLI	SRMR	AIC BIC	McDonald's $\omega$			
							PT	F	EC	PD
<b>Model-1</b>	2108.37	344	0.021 (0.020–0.022)	0.93	0.02	1420.37	0.68	0.73	0.60	0.64
• Davis (1980)				0.93		-1104.48				
• Albiero (2006)										
• Cliffordson (2002)										
• De Corte (2007)										
• Gilet (2013)										
• Himichi (2017)										
• Huang (2012)										
• Lucas-Molina (2017)										
• Manarte (2018)										
• Muchotrigo (2022)										
• Nomura (2012)										
• Ortiz (2011)										
• Pulos (2014)										
• Siu (2005)										
<b>Model-2</b>	2204.72	346	0.021 (0.020–0.022)	0.93 0.92	0.02	1512.72 -1026.81				
• Cliffordson (2002)										
<b>Model-3</b>	8919.36	349	0.046 (0.045–0.047)	0.68 0.66	0.08	8221.77 5660.22				
• Gilet (2013)										
<b>Model-4</b>	13391.33	350	0.05 (-)	0.52 0.48	0.10	12691.33 10122.44				
• Gilet (2013)										
<b>Model-5</b>	2115.25	338	0.021 (0.020–0.022)	0.92 0.93	0.03	1439.25 -1041.56				
• Holgado Tello (2013)										
<b>Model-6</b>	1761.08	344	0.019 (0.018–0.020)	0.95 0.94	0.02	1073.08 -1451.76	0.68	0.73	0.62	0.68
• Lucas-Molina (2017)										
<b>Model-7</b>	2152.43	346	0.021 (0.020–0.022)	0.89 0.87	0.03	1472.43 -1023.06				
• Ortiz (2011)										
<b>Model-8</b>	6352.40	347	0.039 (0.038–0.039)	0.78 0.76	0.06	5658.40 3111.52				
• Siu (2005)										
<b>Model-9</b>	2138.82	345	0.021 (0.020–0.022)	0.93 0.93	0.03	1450.82 -1074.03	0.68	0.73	0.60	0.64
• Pulos (2004)										

**Note.**  $\chi^2$ , chi squared; *df*, degrees of freedom; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; TLI, Tucker Lewis Index; SRMR, Standardized Root Mean Squared Residual; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion;  $\omega$ , McDonald's omega

“-”, values could not be computed

more suitable for the Empathic Concern dimension, instead of the Personal Distress dimension. Moreover, the factor loading of item #13 is significantly higher on the Empathic Concern dimension compared to the Personal Distress dimension, which resulted in a higher reliability for both the Empathic Concern and Personal Distress dimensions, compared to the reliabilities for those dimensions of the original model. Therefore, the current results provided support that the best fitting model for the IRI is the model reported by Lucas-Molina et al. (2017). However, even if Model-6 (Lucas-Molina et al., 2017) had the best fit, the subscale reliabilities were low. This result is in line with previous studies reporting low reliability coefficients for each IRI subscales (Cliffordson, 2001; Manarte & Andrade, 2020; Nomura & Akai, 2012; Ortiz et al., 2011), suggesting that a revision of the item is needed. Moreover, administering a measure with low reliability scores means that a great part of the observed item variance is due to measurement error, therefore the interpretation of results needs caution (Coaley,

2010). However, since other studies reported adequate reliability coefficients ( $> 0.75$ ) for the IRI subscales (De Corte et al., 2007; Gilet et al., 2013; Pulos et al., 2004), perhaps cultural differences could affect factor loadings and, consequently, reliability scores, which can also explain why other studies reported a different factor structure of the IRI.

Many authors have theorized that empathy is not only a multidimensional construct, but its structure might also be hierarchical. In fact, the hierarchical model (Cliffordson, 2002) provided an adequate fit to the data in the current study, however not the best. Therefore, the hypothesis that empathy, as measured with the IRI, could be considered hierarchical is still to be demonstrated. Moreover, in line with empathy as a multidimensional construct, findings of neuroimaging studies linked different brain regions to distinct aspects of empathy (Molenberghs et al., 2016; Schurz et al., 2014). For instance, whereas the orbital-frontal cortex and frontal gyrus are implicated in the recognition and the experience of other people's emotions, the right prefrontal

**Table 4** Standardized factor loadings and correlation between factors of the models proposed by Davis (1980) and Lucas-Moline et al. (2017)

Items	Davis model				Lucas-Molina model			
	PT	F	EC	PD	PT	F	EC	PD
PT3	0.17				0.17			
PT8	0.65				0.65			
PT11	0.57				0.57			
PT15	0.17				0.17			
PT21	0.62				0.62			
PT25	0.54				0.54			
PT28	0.54				0.53			
F1		0.34				0.34		
F5		0.62				0.62		
F7		0.21				0.21		
F12		0.29				0.30		
F16		0.71				0.71		
F23		0.76				0.76		
F26		0.65				0.65		
EC2			0.44				0.44	
EC4			0.30				0.34	
EC9			0.44				0.44	
EC14			0.41				0.46	
EC18			0.37				0.41	
EC20			0.46				0.46	
EC22			0.46				0.46	
PD6				0.57				0.57
PD10				0.44				0.44
PD13				0.04			0.21	
PD17				0.47				0.47
PD19				0.17				0.17
PD24				0.71				0.71
PD27				0.63				0.64
PT	1				1			
F	0.09	1			0.09	1		
EC	0.23	0.19	1		0.21	0.17	1	
PD	-0.01	0.13	0.11	1	-0.01	0.13	0.10	1

**Note.** PT, Perspective Taking; F, Fantasy; EC, Empathic Concern; PD, Personal Distress

cortex is implicated in the ability of perspective taking (Hillis, 2014), as defined by Davis (1980). Therefore, taken together these results suggest that a unidimensional view of empathy and, consequently, of the factor structure of the IRI is not adequate.

The model hypothesized by Gilet et al. (2013), with two first-order factor namely Cognitive Empathy (defined by the items of Perspective Taking and Fantasy) and Affective Empathy (defined by the items of Empathic Concern and Personal Distress), did not show good fit, which supported the hypothesis that empathy, as measured with the IRI, could be considered a multi-dimensional construct rather than a bi-dimensional construct. Supporting these results, several studies demonstrated specific empathy deficits associated with different psychopathological conditions, such as autism (Smith, 2009), bipolar disorder (Shamay-Tsoory et al., 2009), alcohol addiction (Clark et al., 2007). For example, findings in individuals with eating disorders have shown

that individuals with anorexia nervosa are characterized by low abilities in fantasizing, but not in perspective taking, as measured with the IRI (Gaggero et al., 2023). From a theoretical perspective this perfectly reflects Davis' theory of empathy (1980), according to which a good scale assessing empathy should assess individuals' differences in cognitive and emotional empathy, separately, and highlights that the impairments of intrapersonal and interpersonal aspects of empathy are different across different psychopathological conditions.

Possible explanations of why, in the primary studies included in the current work, different factor models reported good fit may be due to the fact that smaller samples can affect factor loadings (Hirschfeld et al., 2014), which in turn affect the model's specification (Heene et al., 2011). Moreover, factor solutions might also depend on the different estimation methods used (e.g., maximum likelihood, ML; weighted least square mean and variance, WLSMV),

which affect the model's error estimates (DiStefano & Morgan, 2014; Li, 2016; Suh, 2015).

The current study has several strengths. First, the up-to-date statistical approach used allowed us to make stronger inferences about the dimensionality of the IRI, taking into account the categorical nature of the analyzed variables. Second, by including multiple samples, the proportions of females and males were more balanced and, consequently, the results were less affected by gender differences. Some limitations of this study also need to be taken into account. First, this approach should tend to give reasonably accurate meta-analytic factor loadings. However, reproducing the model-implied inter-item correlation matrices implies losing a certain portion of the original inter-item correlation matrices variance. Therefore, the newly-generated model fit indices could be biased in the direction of better fit because the fit is being computed in relation to these model-derived approximations to the real correlation matrix and not to the real correlation matrix itself. Second, we did not investigate the accuracy of the factors in clinical samples. Future studies should investigate whether the model reported by this MASEM analysis is also supported for clinical populations.

## Conclusion

The current MASEM on the IRI indicated that the four-factor model reported by Lucas-Molina et al. (2017) had the best fit and reliability among the tested models. Future studies are needed to investigate the accuracy of the dimensionality of the IRI as indicated by the MASEM, across different psychiatric diagnoses and conditions, in order to be able to compare results among different groups. Moreover, the association between the suggested dimensionality of the IRI and other psychological constructs needs to be investigated as well to provide additional validity for the questionnaire.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10862-023-10098-w>.

**Author Contributions** Conceptualization, G.R.; Data curation, G.R. and M.I.; Formal Analysis, G.R.; Methodology, G.R. and M.I.; Results interpretation: G.R., M.B., S.E., D.L., and M.I.; Writing – Original Draft Preparation, G.R., M.B., S.E., M.C., D.L., A.S., and M.I.; All authors contributed to the Writing, Reviewing, Editing and Approval of the final version of the manuscript.

**Data Availability** The information needed to reproduce all of the reported results are not openly accessible. The data is available on request from the author(s). The information needed to reproduce all of the reported methodology is not openly accessible. The material is available on request from the author(s).

## Declarations

**Open Science** We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models.

**Preregistration of Studies and Analysis Plans** This study was not pre-registered.

## References

- Akaike, H. (1974). A New look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Albiero, P., Ingoglia, S., & Lo Coco, A. (2006). Contributo all'adattamento italiano dell'interpersonal reactivity index. *Testing Psicometria Metodologia*, *13*(2), 107–125.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Fixed-effect Versus Random-effects models. *Introduction to Meta-Analysis*. John Wiley & Sons.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Cheung, M. W. L. (2008). A model for integrating Fixed-, Random-, and mixed-effects Meta-analyses into structural equation modeling. *Psychological Methods*, *13*(3), 182–202. <https://doi.org/10.1037/a0013163>
- Cheung, M. W. L. (2014). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods*, *46*(1), 29–40. <https://doi.org/10.3758/s13428-013-0361-y>
- Cheung, M. W. L. (2021). Package "metaSEM." <https://doi.org/10.3389/fpsyg.2014.01521%3E.License>
- Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40–64. <https://doi.org/10.1037/1082-989X.10.1.40>
- Chryssikou, E. G., & Thompson, W. J. (2016). Assessing Cognitive and Affective Empathy through the interpersonal reactivity index: An argument against a two-factor model. *Assessment*, *23*(6), 769–777. <https://doi.org/10.1177/1073191115599055>
- Clark, A. J. (2010). Empathy and Sympathy: Therapeutic distinctions in Counseling. *Journal of Mental Health Counseling*, *32*(2), 95–101.
- Clark, U. S., Oscar-Berman, M., Shagrin, B., & Pencina, M. (2007). Alcoholism and judgments of affective stimuli. *Neuropsychology*, *21*(3), 346–362. <https://doi.org/10.1037/0894-4105.21.3.346>
- Alcoholism
- Cliffordson, C. (2001). Parent's judgements and students' self-judgements of empathy: The structure of empathy and agreement of judgements based on the interpersonal reactivity index (IRI). *European Journal of Psychological Assessment*, *17*, 36–47.
- Cliffordson, C. (2002). The hierarchical structure of empathy: Dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology*, *43*(1), 49–59. <https://doi.org/10.1111/1467-9450.00268>
- Coaley, K. (2010). Chapter 5: The Importance of Reliability. In *An Introduction to Psychological Assessment and Psychometrics*.
- Cohen, D., & Strayer, J. (1996). Empathy in conduct-disordered and comparison youth. *Developmental Psychology*, *32*, 988–998.
- Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. <https://doi.org/10.1177/1754073914558466>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Journal of Personality and Social Psychology*, *10*(January), 85. [http://www.uv.es/~friasnav/Davis\\_1980.pdf](http://www.uv.es/~friasnav/Davis_1980.pdf)

- 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/10.1145/234611.234613>
- Davis, M. H. (1996). In Boulder (Ed.), *Empathy. A social psychological approach*. 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(4), 235–260. <https://doi.org/10.5334/pb-47-4-235>
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100. <https://doi.org/>
- Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *TheScientificWorldJournal*, 6, 1146–1163. <https://doi.org/10.1100/tsw.2006.221>
- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for Ordinal Data. *Structural Equation Modeling*, 21(3), 425–438. <https://doi.org/10.1080/10705511.2014.915373>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. In J. Norcross (Ed.), *Psychotherapy relationships that work* (2nd ed.), pp. 132–152. Oxford University Press.
- Gaggero, G., Luminet, O., Vermeulen, N., De Timary, P., Nandrin, J. L., Goffinet, S., Dereau, J., Shankland, R., Dassonneville, C., & Grynberg, D. (2023). A multidimensional examination of affective and cognitive empathy in Anorexia Nervosa. *European Eating Disorders Review*, November 2022, 505–519. <https://doi.org/10.1002/erv.2980>
- Gilet, A. L., Mella, N., Studer, J., Griehn, 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(1), 42–48. <https://doi.org/10.1037/a0030425>
- Gnambs, T., & Staufenbiel, T. (2016). Parameter accuracy in the meta-analyses of factor structures. *Research Synthesis Methods*, 7, 168–186. <https://doi.org/10.1002/jrsm.1190>
- Hawk, S. T., Keijsers, L., Branje, S. J. T., Van Der Graaff, 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(1), 96–106. <https://doi.org/10.1080/00223891.2012.696080>
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking Misfit in Confirmatory Factor Analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 319–336. <https://doi.org/10.1037/a0024917>
- Hillis, A. E. (2014). Inability to empathize: Brain lesions that disrupt sharing and understanding another's emotions. *Brain*, 137(4), 981–997. <https://doi.org/10.1093/brain/awt317>
- Himichi, T., Osanai, H., Goto, T., Fujita, H., Kawamura, Y., Davis, M. H., & Nomura, M. (2017). Development of a Japanese version of the interpersonal reactivity index. *Methodological Advancement*, 88(1/4), 61–71.
- Hirschfeld, G., von Brachel, R., & Thielsch, M. (2014). Selecting items for big five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality*, 53, 54–63. <https://doi.org/10.1016/j.jrp.2014.08.003>
- Holgado Tello, F. P., Egido, D., Carrasco Ortiz, B., M. A., & Del Gandara, B., M. V (2013). Interpersonal reactivity index: Analysis of invariance and gender differences in Spanish youths. *Child Psychiatry and Human Development*, 44(2), 320–333. <https://doi.org/10.1007/s10578-012-0327-9>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, X., Li, W., Sun, B., Chen, H., & Davis, M. H. (2012). The validation of the interpersonal reactivity index for Chinese teachers from primary and middle schools. *Journal of Psychoeducational Assessment*, 30(2), 194–204. <https://doi.org/10.1177/0734282911410588>
- Jak, S. (2015). Meta-analytic structural equation modeling. *Springer International Publishing*. <https://doi.org/10.1007/978-3-319-27174-3>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Lucas-Molina, B., Pérez-Albéniz, A., Ortuño-Sierra, J., & Fonseca-Pedrero, E. (2017). Dimensionalidad E invarianza de la medida del interpersonal reactivity index (IRI) en función del género. *Psicothema*, 29(4), 590–595. <https://doi.org/10.7334/psicothema2017.19>
- Manarte, L., & Andrade, A. R. (2020). Comparative Analysis and Validation of the Portuguese Version of the interpersonal reactivity index. *Psilogos*, 16(1), 45–59.
- Mann, R. E., & Barnett, G. D. (2013). Victim Empathy intervention with sexual offenders: Rehabilitation, punishment, or Correctional Quackery? *Sexual Abuse: Journal of Research and Treatment*, 25(3), 282–301. <https://doi.org/10.1177/1079063212455669>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, 65, 276–291. <https://doi.org/10.1016/j.neubiorev.2016.03.020>
- Muchotrigo, M. G., Correa-Rojas, J., Medina, E. M., & Milera, K. M. (2022). Validity and factorial invariance of the interpersonal reactivity index in Peruvian university students. *Ciencias Psicológicas*, 16(2), 1–15. <https://doi.org/10.22235/cp.v16i2.2810>
- Neale, M. C., Hunter, M. D., Pritkin, J., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81(2), 535–549. <https://doi.org/10.1007/s11336-014-9435-8>. OpenMx
- Nomura, K., & Akai, S. (2012). Empathy with fictional stories: Reconsideration of the Fantasy scale of the interpersonal reactivity index. *Psychological Reports*, 110(1), 304–314. <https://doi.org/10.2466/02.07.09.11.PR0.110.1.304-314>
- Oliveira-Silva, P., & Gonçalves, Ó. F. (2011). Responding empathically: A question of heart, not a question of skin. *Applied Psychophysiology Biofeedback*, 36(3), 201–207. <https://doi.org/10.1007/s10484-011-9161-2>
- Ortiz, M. Á. C., Egido, B. D., García, M. I. B., Tello, F. P. H., & del Barrio Gándara, M. V. (2011). Propiedades psicométricas del interpersonal reactivity index (IRI) en población infantil y adolescente española. *Psicothema*, 23(4), 824–831.
- Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1–20.
- Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the interpersonal reactivity index. *Social Behavior and Personality*, 32(4), 355–360. <https://doi.org/10.2224/sbp.2004.32.4.355>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Schroeders, U., Kubera, F., & Gnambs, T. (2021). The structure of the Toronto Alexithymia Scale (TAS-20): A

- meta-analytic confirmatory factor analysis. *Assessment*. <https://doi.org/10.1177/10731911211033894>
- Schumacker, R., & Lomax, R. (2012). A Beginner's Guide to Structural Equation Modeling, 3rd edn. In *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (Vol. 175, Issue 3). [https://doi.org/10.1111/j.1467-985x.2012.01045\\_12.x](https://doi.org/10.1111/j.1467-985x.2012.01045_12.x)
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.
- Shamay-Tsoory, S., Harari, H., Szepsenwol, O., & Levkovitz, Y. (2009). Neuropsychological evidence for impaired cognitive empathy in euthymic BD. *Journal of Neuropsychiatry and Clinical Neurosciences*, *21*, 59–67.
- Siu, A. M. H., & Shek, D. T. L. (2005). Validation of the interpersonal reactivity index in a Chinese context. *Research on Social Work Practice*, *15*(2), 118–126. <https://doi.org/10.1177/1049731504270384>
- Smith, A. (2009). The empathy imbalance hypothesis of autism: A theoretical approach to cognitive and emotional empathy in autistic development. *The Psychological Record*, *59*(3), 489–510.
- Suh, Y. (2015). The performance of maximum likelihood and weighted least Square Mean and Variance Adjusted estimators in Testing Differential Item Functioning with Nonnormal Trait distributions. *Structural Equation Modeling*, *22*(4), 568–580. <https://doi.org/10.1080/10705511.2014.937669>
- Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-process*. The MacMillan Company.
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.