Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model

Evangelia G. Chrysikou[1] and W. Jake Thompson[1]

Abstract

One aspect of higher order social cognition is empathy, a psychological construct comprising a cognitive (recognizing emotions) and an affective (responding to emotions) component. The complex nature of empathy complicates the accurate measurement of these components. The most widely used measure of empathy is the Interpersonal Reactivity Index (IRI). However, the factor structure of the IRI as it is predominantly used in the psychological literature differs from Davis's original four-factor model in that it arbitrarily combines the subscales to form two factors: cognitive and affective empathy. This two-factor model of the IRI, although popular, has yet to be examined for psychometric support. In the current study, we examine, for the first time, the validity of this alternative model. A confirmatory factor analysis showed poor model fit for this two-factor structure. Additional analyses offered support for the original four-factor model, as well as a hierarchical model for the scale. In line with previous findings, females scored higher on the IRI than males. Our findings indicate that the IRI, as it is currently used in the literature, does not accurately measure cognitive and affective empathy and highlight the advantages of using the original four-factor structure of the scale for empathy assessments.

Keywords

cognitive and affective empathy, interpersonal reactivity index, factor structure, confirmatory factor analysis

The ability for empathy is considered among the building blocks of successful interpersonal relationships. Past research has shown that empathy is a complex and multi-dimensional construct (Davis, 1983; Lietz et al., 2011; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011; Zoll & Enz, 2005) that involves the ability to interpret correctly the emotions of others, as well as have the correct emotional response in a given situation. These abilities are considered to make up the two main components of the empathic process: cognitive and affective empathy. More specifically, cognitive empathy has been generally conceptualized as involving conscious emotional processing such as mentalizing behaviors, perspective taking, imagination, and emotion recognition (see Smith, 2006, for a review). Affective empathy, on the other hand, is thought to constitute largely unconscious processes involving the sharing of emotions, such as personal distress, affective responsiveness, and emotional contagion, and can be characterized as one's "gut reaction" to emotional stimuli (Hooker, Verosky, Germine, Knight, & D'Esposito, 2010). Due to its importance for building meaningful interpersonal relationships, recent research has focused on defining and assessing empathy, as well as identifying potential empathy impairments associated with personality disorders and various forms of psychiatric illness.

The most common psychometric tool for measuring an individual's empathy is the Interpersonal Reactivity Index (IRI; Davis, 1980). This questionnaire was originally validated as a multidimensional measure and consists of four subscales that are thought to measure distinct aspects of empathy: Perspective Taking (the ability to shift to another's emotional perspective), Empathic Concern (feeling warmth or compassion for others), Fantasy (the ability to put oneself in a fictional situation), and Personal Distress (feeling fear or anxiety in response to seeing others in distress). Carey, Fox, and Spraggins (1988) later validated this structure by using principal components analysis to identify four factors that were in line with the scale structure originally reported by Davis (1980). Further psychometric studies (e.g., Cliffordson, 2001; Hawk et al., 2013; Pulos, Elison, &

[1]University of Kansas, Lawrence, KS, USA

Corresponding Author:
Evangelia G. Chrysikou, Department of Psychology, University of Kansas, 1415 Jayhawk Boulevard, 426 Fraser Hall, Lawrence, KS 66045, USA.
Email: lilachrysikou@ku.edu

Lennon, 2004) have offered support for a possible hierarchical structure of the scale. Under a hierarchical model, just as items of a scale can be combined together to form latent factors, these different factors can, in turn, be combined together to make up a second-order factor forming a hierarchy, thus increasing the model's explanatory power. Although Davis (1983) did not use a hierarchical model, he described the IRI as having a hierarchical structure with each factor representing a specific aspect of the more general empathy construct. Later studies hypothesized that the four latent factors of the IRI share a common, second-order factor. Specifically, both Cliffordson (2002) and Hawk et al. (2013) found support for the four-factor model of the scale and identified a general empathy second-order factor onto which all of the subscales loaded heavily. Pulos et al. (2004) also found support for the original structure as well as a second-order general empathy factor, but in these findings, the general empathy factor consisted of only the Perspective Taking, Fantasy, and Empathic Concern subscales, whereas Personal Distress constituted a factor on its own. Thus, past work on the structure of the IRI has confirmed the explanatory strength of the four-factor model, with the potential for a second-order, general empathy factor, encompassing three or four of the IRI subscales.

A substantial body of research (e.g., Davis & Franzoi, 1991; Rankin, Kramer, & Miller, 2005; Schutte et al., 2001; Tangney, 1991) has employed the four-factor model of the IRI to study empathy and its relationship to other psychological constructs or disorders. On the other hand, more recent work has focused on the distinction between cognitive and affective empathy and the potential differential contributions of each to abnormal behavior. However, although cognitive and affective empathy have been generally considered as the key aspects of empathy, neither construct has been particularly well defined in the literature and it is currently unclear which behaviors exactly would reflect either cognitive or affective empathy. Despite their importance of our understanding of empathy (Hooker et al., 2010; Smith, 2006), the lack of precise operational definitions for cognitive and affective empathy unavoidably introduces challenges for the assessment of these constructs. Nevertheless, although the IRI has only been validated for measuring empathy as a general construct, it has recently been adapted to measure cognitive and affective empathy, and, in turn, has been used to contribute to the operational definition of these aspects of empathy, despite the lack of psychometric support of this two-factor model of the scale from the quantitative literature.

In particular, it has become common practice in recent psychological studies on empathy to combine the Perspective Taking and the Fantasy subscales of the IRI into a single "Cognitive Empathy" factor, and the Empathic Concern and the Personal Distress subscales into a single "Affective Empathy" factor (e.g., Bock & Hosser, 2014;

Calabria, Cotelli, Adenzato, Zanetti, & Miniussi, 2009; Cusi, Macqueen, Spreng, & McKinnon, 2011; Dziobek et al., 2011; Harari, Shamay-Tsoory, Ravid, & Levkovitz, 2010; Hengartner et al., 2013; Hooker et al., 2010; Maurage et al., 2011; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009; Shamay-Tsoory, Shur, Harari, & Levkovitz, 2007; Shamay-Tsoory, Tomer, Goldsher, Berger, & Aharon-Peretz, 2004). This "cognitive–affective" split of the IRI has then been used to examine cognitive and affective empathy in the context of personality disorders, alcoholism, dementia, depression, recidivism, schizophrenia, as well as in neuroimaging studies aiming to identify the neural correlates of empathy and its subcomponents.

For instance, following this psychometrically arbitrary cognitive–affective empathy divide of the IRI as specified above, Harari et al. (2010) found that, compared with control subjects, individuals diagnosed with borderline personality disorder suffered from impaired cognitive, but not affective, empathy and this impairment was related to psychotic symptomatology. In a similar study, Hengartner et al. (2013) examined the relationship between 10 personality disorders identified in the *Diagnostic and Statistical Manual for Mental Disorders–Fourth edition* and empathy. They used a modified two-factor model of the IRI according to which the Empathic Concern and the Personal Distress subscales were combined as a measure of affective empathy, whereas the Fantasy subscale was omitted and the Perspective Taking subscale was used on its own as a measure of cognitive empathy. The authors found that a diagnosis of a personality disorder was related to a decline in affective empathy, but was unrelated to cognitive empathy. They argued that affective empathy was critical to personality disorders, in spite of the fact that when its two components—as measured by the IRI's Empathic Concern and Personal Distress subscales—were examined separately, the majority of the personality disorders studied were not characterized by scores significantly different from control subjects on either subscale. Indeed, scores on each subscale were in opposition with each other, such that most personality disorders examined were associated with either *low* scores on the Empathic Concern subscale or *high* scores on the Personal Distress subscale, but not low or high scores on both scales. Critically, paranoid and schizoid personality disorder, the only two disorders that were associated with significantly different scores from control subjects on both subscales, were marked by this reversed pattern according to which Empathic Concern scores were low, whereas Personal Distress scores were elevated. Not only did the scales move together, as would be expected if they represented the same construct, but they actually moved in opposite directions, strongly suggesting that they measure very different aspects of empathy.

Additionally, the two-factor model of the IRI has also been used to inform clinical decisions for a number of

psychopathologies. For example, using the two-factor model, Maurage et al. (2011) found that affective empathy was impaired in individuals suffering from alcoholism, whereas cognitive empathy remained intact. Thus, affective empathy, and its associated processes, might be an important symptom of the psychopathology associated with alcohol substance disorder and, thus, should be a focus of patient evaluation and treatment. Similarly, Cusi et al. (2011) found that major depressive disorder was associated with both decreased cognitive and affective empathy. However, when examined independently, only the scores on the Perspective Taking and the Empathic Concern subscales were significantly lower in depressed individuals compared with healthy controls. The second cognitive subscale (Fantasy) and the second affective subscale (Personal Distress) did not differ between depressed and control participants. Thus, the results did not align with the two-factor model of the IRI, as only one of the two subscales used to measure cognitive empathy and only one of the two subscales used to measure affective empathy differed significantly between depressed patients and healthy control subjects.[1] Despite this significant shortcoming in the measurement of cognitive and affective empathy through the IRI, however, these assessments under the two-factor model were considered critical toward guiding future research on the treatment and underlying processes involved in depression, such as investigating impaired empathy as a state rather than trait and examining the relationship between, empathy, interpersonal functioning, and social performance in a variety of roles (e.g., work, parent, spouse, etc.) as a potential treatment directive (for a similar study on recidivism in violent and nonviolent offenders, see Bock & Hosser, 2014).

In line with the practice espoused in behavioral studies as discussed above, the neuroscience community has, equally problematically, used this two-factor model of the IRI in work attempting to localize cognitive and affective empathy to specific brain regions. For example, Calabria et al. (2009) examined empathy and emotional processing in a case study of a patient diagnosed with semantic dementia, a neurodegenerative disorder that selectively affects left frontotemporal cortex and which is marked by the progressive loss of knowledge about the world (i.e., semantic memory). Using the two-factor model of the scale, Calabria et al. (2009) found that the patient showed deficits in cognitive but not affective empathy, compared with family members' assessments of her empathy levels before the onset of the dementia. These findings were then used to establish hypotheses about the likely functional–anatomical locus of affective empathy processes in the brain, given the patient's deficits and the response patterns on the IRI observed in patients with other kinds of neurodegenerative disorders.

Similarly, studies have sought to identify differences in scores on the two IRI factors between groups of patients with lesions in different brain regions (Shamay-Tsoory

et al., 2004; Shamay-Tsoory et al., 2009). These studies have reported a dissociation between the two factors, with impairments in cognitive empathy associated with lesions in the ventromedial prefrontal cortex, whereas impairments in affective empathy associated with lesions in the inferior frontal gyrus. In contrast to these findings, using a functional magnetic resonance imaging (fMRI) paradigm, Hooker et al. (2010) found that cognitive empathy was associated with activity in the inferior frontal gyrus and the superior temporal sulcus, whereas affective empathy was associated with activity in the precentral gyrus. In line with these results, Dziobek et al. (2011) used fMRI to relate cognitive empathy with activity in the superior temporal sulcus and the superior temporal gyrus; though, contrary to Hooker et al. (2010), affective empathy was linked with activity in medial insular cortex. Overall, much of the literature on the neural bases of empathy has employed this dichotomy between cognitive and affective empathy, as measured by the different subscales of the IRI, to identify regions sensitive to each empathy subcomponent.

Based on the findings reviewed above, it is evident that, although there has been no psychometric validation of the two-factor approach to the IRI, several studies have used this method to measure cognitive and affective empathy and relate them to a number of different psychological constructs or different brain systems. This practice is problematic for empathy research because (a) cognitive and affective empathy may reflect a theoretically meaningful division of empathy abilities (Davis, 1983; Lietz et al., 2011; Reniers et al., 2011; Zoll & Enz, 2005), yet it is currently unknown whether the IRI under the two-factor model provides a valid measure of these constructs; (b) if the two-factor model of the IRI has low construct validity, conclusions regarding empathic abilities and their neural bases may be incorrect or compromised; and (c) clinical decisions may be made on the wrong premises. For these reasons, it is critical for the clinical psychological and neuroscience literature on empathy that this two-factor approach to the IRI be reexamined for psychometric support. To achieve this objective, in the current study we used confirmatory factor analysis (CFA) to test whether the two-factor model of the IRI significantly fits the measure. If the two-factor model were to fit the IRI, we would expect to see acceptable model fit indices from the CFA, which would provide support to the validity of the model used in psychological studies of empathy. Poor model fit indices, however, would indicate that this model is not supported and that the latent constructs proposed are not actually valid under this measure. This would then call into question the conclusions of studies that made use of this two-factor model of the scale to examine cognitive and affective empathy. Based on previous research investigating the structure of the IRI, we hypothesized that the two-factor model would not show good fit to the IRI and that the ideal approach to using the scale would be a four-factor model

similar to the original model proposed by Davis (1980), which has received extensive prior psychometric support.

Method

Participants

Four hundred and thirty-five participants ($N = 435$, 247 female) completed the IRI through Amazon's Mechanical Turk (MTurk), after providing informed consent. Qualification requirements limited participation to individuals located in the United States who had an MTurk approval rating of over 50%. The participants ranged in age from 18 to 70 years, with a mean age of 33.17. Participants received $0.10 for approximately 3.5 minutes ($1.71 per hour), which is above the median hourly pay on MTurk of $1.38 (Horton & Chilton, 2010). Of the 435 participants who accepted the job, 18 participants completed only the demographic questionnaire and declined to complete the IRI. Thus, these individuals were removed from the analyses. The final sample consisted of 417 individuals ($N = 417$, $M_{age} = 33.17$). Of the sample, 59.0% identified as female and 41.0% as male. The ethnic makeup of the sample was 70.0% Caucasian, 10.3% African American, 9.4% Asian, 7.2% Hispanic/Latino, and 3.1% identified as "Other."

Materials

Interpersonal Reactivity Index (Davis, 1980). The IRI is a 28-item questionnaire that is measured on a 5-point Likert-type scale ranging from 0 (*does not describe me well*) to 4 (*describes me very well*). The questionnaire is divided into four subscales of seven items each. The psychometric properties of the scale were confirmed by Davis (1980), with the subscales showing acceptable test–retest reliability. Like most measures of empathy, females, on average, score higher on all subscales compared with males (Davis, 1983).

Procedure

Participants accepted the job (known as a "HIT") on MTurk. After accepting, they were directed to the survey, which was hosted on Qualtrics™. They provided informed consent and they were free to discontinue the study at any point. Participants completed a demographics questionnaire and the IRI. They were then given a completion code to be entered on MTurk to receive their payment.

Analytical Procedure

Confirmatory Factor Analysis. To investigate the validity of the two-factor structure of the IRI as it is currently used in the psychological literature, we conducted a categorical CFA by analyzing the polychoric correlations (i.e., the correlations of theorized latent variables that underlie ordinal variables) of the scale items using M*plus* version 7.11. The analysis used the robust diagonally weighted least squares estimation method (weighted least squares means and variance adjusted in M*plus*), which has been shown to have appropriate power and which performs well with sample sizes of $N > 200$ (Flora & Curran, 2004). Because latent variables were unmeasured, and, thus, were not associated with specific units of measurement, the variance of each factor was fixed at 1, placing the factors on a standard scale. Finally, we fit an oblique model in which the latent factors were allowed to correlate. To verify the validity of the original model on this data set, we also replicated previous studies on the structure of the IRI by estimating the original four-factor model (Davis, 1980), as well as the hierarchical model as suggested by Pulos et al. (2004). These analyses used the same estimation method (weighted least squares means and variance adjusted) and identification procedures that were used for the two-factor model. The latent factors in these models we also allowed to correlate with each other.

Results

The descriptive statistics for the IRI subscales are presented in Table 1. The CFA of the two-factor model failed to yield an acceptable model fit. The fit indices for this model, all suggest a poor fit for the model using standard cutoffs of >0.95 for the comparative fit index (CFI) and Tucker–Lewis index (TLI), and <0.05 for the root mean square error of approximation (RMSEA; Table 2; Hooper, Coughlan, & Mullen, 2008).[2] The correlation between the two factors was $r = .71$ ($p < .001$); however, this should be interpreted with caution considering the fit indices of the model. Given the poor model fit exhibited by the two-factor model, we aimed to replicate previous findings regarding the structure of the IRI to ensure this finding was not due to sample characteristics. A CFA of Davis's (1980) original structure showed much better fit than the two-factor model (see Table 2). The CFI of 0.96 and TLI of 0.95 are above or equal to the accepted cutoff of 0.95, indicating that the four-factor structure is an acceptable fit for these data. Although the RMSEA of 0.11 is higher than would normally be acceptable using a cutoff of 0.05, overall, the current CFA evidence supports the original four-factor structure of the IRI. The factor correlations for this model are presented in Table 3.

We further tested the hierarchical model as suggested by Pulos et al. (2004) in which the Perspective Taking, Fantasy, and Empathic Concern subscales are allowed to load onto a higher order empathy construct. This model fit similarly to the original four-factor structure, with an acceptable CFI and TLI and an RMSEA higher than the <0.05 cutoff (see Table 2). The correlation between the higher order factor and the Personal Distress factor was $r = -.13$ ($p = .01$). Thus, our analyses indicate that the original subscales should be scored as separate factors, although the precise relationship between the factors is unclear from this set of

Table 1. Descriptive Statistics for Each of the Interpersonal Reactivity Index Subscales.

	Perspective Taking	Fantasy	Empathic Concern	Personal Distress
Mean	17.98	18.41	19.86	11.83
Median	18.00	18.00	20.00	12.00
Standard deviation	4.96	5.61	5.16	5.47
Minimum	3.00	0.00	3.00	0.00
Maximum	28.00	28.00	28.00	28.00

Table 2. Model Fit Indices for Various Latent Structures of the Interpersonal Reactivity Index in Past Studies and the Present Study.

	Previous studies					
	Model	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
Davis (1980)	Four-factor (original)	Model fit indices not reported				
Cliffordson (2002)	Hierarchical (all subscales)	—	—	0.057	[0.043, 0.069]	—
Pulos et al. (2004)	Hierarchical (PT, FS, and EC)	Model fit indices not reported				
Hawk et al. (2013)	Hierarchical (all subscales)	0.922-0.953	—	0.063-0.065	—	0.057-0.067

	Current study					
	Model	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
Cognitive–affective split (PT and FS; EC and PD)	Two-factor	0.568	0.532	0.182	[0.178, 0.186]	—
Per Davis (1980)	Four-factor (original)	0.955	0.950	0.110	[0.106, 0.115]	—
Per Pulos et al. (2004)	Hierarchical (PT, FS, and EC)	0.950	0.945	0.116	[0.112, 0.121]	—

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CI = confidence interval; PT = Perspective Taking; FS = Fantasy; EC = Empathic Concern; PD = Personal Distress. Davis (1980) and Pulos et al. (2004) determined model fit by the pattern of factor loadings, with each item loading strongly onto one factor indicating good fit. The suggested cutoffs for the fit indices as suggested by Hooper et al. (2008) are >0.95 for the CFI and TLI; <0.05 for the RMSEA; <0.08 for the SRMR.

Table 3. Factor Correlations of the Original Four-Factor Model.

	Perspective Taking	Fantasy	Empathic Concern	Personal Distress
Perspective Taking	1.00			
Fantasy	.25***	1.00		
Empathic Concern	.66***	.31***	1.00	
Personal Distress	−.20***	.13**	−.09	1.00

*p < .05. **p < .01. ***p < .001.

data. A comparison of the model fit indices attained from these analyses with Davis's (1980) original four-factor model and the hierarchical models of Cliffordson (2002), Hawk et al. (2013), and Pulos et al. (2004) can be found in Table 2.

To examine whether our sample was consistent with other samples that have been used in empathy research, we last examined the original four subscales for gender differences. In the original validation of the IRI, Davis observed significantly higher scores for females compared with males (Davis, 1980, 1983). To control for Type I error, a corrected *p* value of .0125 was used for the four *t* tests. Using this adjusted *p* value, the gender difference findings were replicated in this study, with females scoring higher on

all subscales (*p* values ranged from < .001 to .006; Cohen's *d* ranged from −0.28 to −0.57; see Table 4). Together with the outcomes of the CFA on the original, four-factor and the hierarchical models, these results suggest that the sample used in this study is analogous to those used in past examinations of the factor structure of the IRI in the literature, this offering additional validation of the present findings.

Discussion

Several recent studies in psychology and neuroscience have used the IRI as a measure of cognitive and affective empathy with the intention to examine their relationship with other psychological constructs and disorders. However, a

Table 4. *t* Tests of Gender Differences on the Interpersonal Reactivity Index Subscales.

Subscale	Mean score: males	Mean score: females	*t*-Test results	Cohen's *d*
Perspective Taking	17.19	18.52	$t(375.56) = -2.75, p = .006$	−0.28
Empathic Concern	18.25	20.98	$t(382.26) = -5.62, p < .001$	−0.57
Fantasy	17.17	19.27	$t(374.15) = -3.85, p < .001$	−0.40
Personal Distress	10.91	12.48	$t(380.3) = -2.94, p = .003$	−0.30

question that remains is whether this two-factor model that is widely used in the literature accurately represents the underlying structure of the IRI. Here, we examined for the first time the validity of this two-factor structure of the questionnaire by conducting a CFA to determine if this model provides an acceptable representation of the latent structure of the scale. The CFA showed unacceptable model fit for the widely popular two-factor approach to the IRI. Notably, all of the fit indices were below the normally accepted values. Thus, our data do not support this type of underlying structure of the scale; instead, our analyses confirm the four-factor model originally proposed by Davis (1980) and suggest that the best practice for scoring the scale is to obtain four separate scores for the IRI, reflective of participant performance on each of its four subscales. Our findings strongly support the conclusion that the two-factor model of the IRI, as identified and used in the literature, does not provide a valid measure of cognitive and affective empathy.

These findings and the lack of support for the two-factor model may be attributed, in part, to an inherent bias of the questionnaire toward cognitive empathy. That is, items that are considered to capture affective empathy in the two-factor model (e.g., "In emergency situations, I feel apprehensive and ill-at-ease") require the individual to use cognitive empathy to put herself in a situation before responding. Essentially, in this scale, cognitive empathy acts as a gatekeeper to the accurate measurement of affective empathy. As a result, a diminished ability for cognitive empathy may substantially influence one's responses on affective empathy items, thus resulting in scores that misrepresent one's affective empathy abilities. Similarly, normal cognitive empathy could pose problems for measuring affective empathy deficits through this questionnaire. For example, when putting oneself in a specific situation, the participant may provide a response they know to be emotionally appropriate in that situation, without necessarily experiencing the emotional or "gut-level" reaction affective empathy is meant to capture. Although such behavior would reflect cognitive empathy, it would not provide valid assessments of affective empathy. Therefore, the nature of the IRI introduces a response bias that may make it difficult or impossible to accurately measure affective empathy. Even if some items on the IRI refer to situations that would involve affective empathy if experienced in real life, it is more likely that each subscale of the questionnaire may, in

fact, only capture different facets of cognitive empathy, some of which pertain to affective circumstances more than others. As such, although the IRI may, thus, be used to assess—and operationally define—behaviors associated with cognitive empathy, it may not provide a valid measure of affective empathy.

A number of recent studies have attempted to address this limitation of the IRI by evaluating affective empathy using behavioral measures (Derntl et al., 2010; Dziobek et al., 2008; Krause, Enticott, Zangen, & Fitzgerald, 2012; Masten, Eisenberger, Pfeifer, & Dapretto, 2010; Rameson, Morelli, & Lieberman, 2011; Thoma et al., 2011). For instance, using fMRI, Derntl et al. (2010) assessed cognitive empathy by asking participants to identify the emotion portrayed in images depicting different facial expressions. Similarly, they measured affective empathy by having participants read sentences that provoke emotion and then asking them to choose which of two faces best reflected the emotion they were experiencing (for a similar measure, see Thoma et al., 2011). Thus, although the original, validated four-factor model of the IRI can be an invaluable tool in collecting data on empathy overall, it may better perform this role in future studies if used in conjunction with behavioral measures that can better capture cognitive and affective empathy.

Beyond developing reliable behavioral assessments of the different empathy subcomponents as discussed above, we note that individual IRI subscales have also been used to capture cognitive and affective empathy (e.g., Cox et al., 2012; Hengartner et al., 2013). For example, Cox et al. (2012) excluded from their empathy assessments the Fantasy and Personal Distress subscales of the IRI due to reliability concerns associated with them. Instead, they used only the Perspective Taking subscale as a measure of cognitive empathy and only the Empathic Concern subscale as a measure of affective empathy, an approach they found to be more reliable relative to the two-factor model. Nevertheless, issues stemming from the response bias toward cognitive empathy, as discussed above, still persist with this approach. It is further currently unknown whether the Perspective Taking and Empathic Concern subscales offer a comprehensive measure of cognitive and affective empathy or rather reflect only facets of much larger constructs. Similarly, it is unclear exactly which, if any, aspects of cognitive and affective empathy are captured by the Fantasy and Personal Distress subscales, respectively. For example, in a study

comparing the Hogan Empathy Scale (Hogan, 1969) with the IRI, Johnson, Cheek, and Smither (1983) found the Personal Distress subscale in particular measured a very different type of emotional response, as this subscale did not significantly correlate with any of the aspects measured by the Hogan Empathy Scale. Thus, there would be no reason to believe that Personal Distress and Empathic Concern define a single factor. Future research should explore in more depth the construct validity of the subscales of the IRI with regard to their potential to measure reliably cognitive and affective empathy.

We note that our study may potentially be limited by sample characteristics. Data collected through MTurk might not be as representative of the general population in terms of different psychological characteristics that may influence participants responses on the IRI. For instance, Goodman, Cryder, and Cheema (2013) found that MTurk participants tend to be less extraverted and also tend to show lower self-esteem than other research samples. However, the majority of research findings also suggest that although samples from MTurk are more diverse in terms of ethnic and socio-economic characteristics, the data collected are very similar to data attainted through more conventional means (Casler, Bickel, & Hackett, 2013; Paolacci, Chandler, & Ipeirotis, 2010). Although our analyses replicated individual differences in performance with regard to gender that have been reported in past studies, comparing participants' IRI scores with their performance on other empathy measures would offer further support for the validity of the latent variables and could inform hypothesis on the best model for the IRI. Future studies using more traditional samples (e.g., psychology participant pools) and additional empathy measures (e.g., Hogan's Empathy Scale) should attempt to replicate these results to confirm the underlying structure of the IRI using the methods presented in this study.

Overall, researchers have employed a two-factor approach to the IRI in an effort to define empirically cognitive and affective empathy, the two key empathy components that lack precise definitions in the literature. However, our findings indicate that this practice may be misguided, as our analysis failed to support the presence of these constructs within the IRI. Instead, our results would strongly support the use of the four-factor model of the IRI, as it was originally validated. The model fit indices show much better fit for the original structure of the IRI than the broadly used two-factor model that was tested in this study (see Table 2). In light of these findings, the persistent use of the two-factor approach to the IRI as a measure of cognitive and affective empathy, despite the lack of psychometric validation as revealed in this study, will unavoidably continue to litter the literature with incorrect operational definitions of these constructs. A potentially more productive approach to empathy research would be to focus on the development of reliable and valid measures of cognitive and affective empathy, either through implicit behavioral measures

(e.g., Derntl et al., 2010; Thoma et al., 2011) or by means of self-report measures designed to specifically capture each of these constructs. That the IRI does not support a two-factor structure does not preclude the possibility of a new self-report measure of cognitive and affective empathy. For example, Vachon and Lynam (2016) have recently developed the Affective and Cognitive Measure of Empathy that includes one scale for cognitive empathy and two scales for affective empathy, reflective of a three-factor model of the construct. Nevertheless, for either of these options to succeed, it is imperative to first establish clear definitions of what these constructs entail so that valid assessment approaches can be created (for a more in-depth description of how this process can proceed, see Huff, Steinberg, & Matts, 2010; Mislevy & Haertel, 2006; Mislevy, Steinberg, & Almond, 2003).

The findings of the present work have significant implications for the study of empathy. As the IRI is the most widely used measure of empathy, using the correct structure of this scale is critical to obtaining results that represent empathy in a valid way and can, thus, promote our understanding of this important psychological construct. Our results strongly support the conclusion that the combination of the Perspective Taking and the Fantasy subscales of the IRI into a single "Cognitive Empathy" factor, and the Empathic Concern and the Personal Distress subscales into a single "Affective Empathy" factor is a misguided practice that compromises the IRI as a valid measure of empathy. Our analyses demonstrate that this two-factor approach is not actually reflected in the structure of the scale; hence, cognitive and affective empathy may not be accurately measured by this version of the IRI. This finding raises concerns for studies that have used this two-factor approach to understand the implication of empathy deficits in various forms of psychopathology. For example, using the two-factor model, Maurage et al. (2011) reported that affective, but not cognitive, empathy is impaired in alcoholism and, hence, should be a critical component of alcoholism rehabilitation programs. However, if these factors are not present in the IRI, the measurement of cognitive and affective empathy in this study is likely not valid and any rehabilitation efforts inefficient and difficult to evaluate. The same holds true for other psychiatric illnesses that have been tied to empathy such as depression, sociopathy, and autism spectrum disorders (Bock & Hosser, 2014; Cusi et al., 2011; Samson, Huber, & Gross, 2012; Thoma et al., 2011), as well as studies attempting to link these empathic subprocesses to distinct brain systems (Calabria et al., 2009; Dziobek et al., 2011; Hooker et al., 2010; Shamay-Tsoory et al., 2004; Shamay-Tsoory et al., 2009). Because of the importance of empathy for several aspects of social life, reevaluating the use of the two-factor approach to the IRI as a valid measure of cognitive and affective empathy is critical to advance our understanding of this valuable psychological construct, as well as support the use of the IRI as it was originally

intended and the development of behavioral and self-report measures that can accurately capture cognitive and affective empathy.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We note that investigations using less popular empathy measures further support this conclusion. For example, when developing the Empathy Quotient, Lawrence, Shaw, Baker, Baron-Cohen, and David (2004) examined the relationship of this new scale to the IRI and reported that the Empathy Quotient only correlated with the Perspective Taking and the Empathic Concern subscales, but not with the other subscales within each of the two factors (cognitive and affective empathy). These results further suggest that the four subscales of the IRI may not be truly reflective of these two latent factors, otherwise we would expect other measures of cognitive and affective empathy to correlate with both subscales in each factor, not just the one.

2. We also conducted a CFA on the two-factor model using item parceling (Little, 2013; Little, Cunningham, Shahar, & Widaman, 2002; Little, Rhemtulla, Gibson, & Schoemann, 2013). Both balanced and facet-representative parceling methods failed to yield acceptable model fit. Thus, all of the methods we used for estimating the two-factor model of the IRI, as widely used in the empathy literature, provided unacceptable model fit.

References

Bock, E. M., & Hosser, D. (2014). Empathy as a predictor of recidivism among young adult offenders. *Psychology, Crime & Law*, *20*, 101-115. doi:10.1080/1068316X.2012.749472

Calabria, M., Cotelli, M., Adenzato, M., Zanetti, O., & Miniussi, C. (2009). Empathy and emotion recognition in semantic dementia: A case report. *Brain and Cognition*, *70*, 247-252. doi:10.1016/j.bandc.2009.02.009

Carey, J. C., Fox, E. A., & Spraggins, E. F. (1988). Replication of structure findings regarding the Interpersonal Reactivity Index. *Measurement and Evaluation in Counseling and Development*, *21*, 102-105.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156-2160. doi:10.1016/j.chb.2013.05.009

Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy. *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*, 49-59.

Cox, C. L., Uddin, L. Q., Di Martino, A., Castellanos, F. X., Milham, M. P., & Kelly, C. (2012). The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. *Social, Cognitive, and Affective Neuroscience*, *7*, 727-737. doi:10.1093/scan/nsr051

Cusi, A. M., Macqueen, G. M., Spreng, R. N., & McKinnon, M. C. (2011). Altered empathic responding in major depressive disorder: Relation to symptom severity, illness burden, and psychosocial outcome. *Psychiatry Research*, *188*, 231-236. doi:10.1016/j.psychres.2011.04.013

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

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.

Davis, M. H., & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality*, *25*, 70-87.

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

Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage*, *57*, 539-548. doi:10.1016/j.neuroimage.2011.05.005

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*, *38*, 464-473.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, *26*, 213-224. doi:10.1002/bdm.1753

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research*, *175*, 277-279. doi:10.1016/j.psychres.2009.03.002

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*, 96-106.

Hengartner, M. P., Ajdacic-Gross, V., Rodgers, S., Müller, M., Haker, H., & Rossler, W. (2013). Fluid intelligence and empathy in association with personality disorder trait-scores: Exploring the link. *European Archives of Psychiatry and Clinical Neuroscience*, *264*, 441-448. doi:10.1007/s00406-013-0441-0

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307-316.

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2010). Neural activity during social

signal perception correlates with self-reported empathy. *Brain Research*, *1308*, 100-113. doi:10.1016/j.brainres.2009.10.006

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53-60.

Horton, J. J., & Chilton, L. B. (2010, June). *The labor economics of paid crowdsourcing*. Paper presented at the Proceedings from EC '10: The 11th ACM Conference on Electronic Commerce, New York, NY.

Huff, K., Steinberg, L., & Matts, T. (2010). The promises and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education*, *23*, 310-324.

Johnson, J. A., Cheek, J. M., & Smither, R. (1983). The structure of empathy. *Journal of Personality and Social Psychology*, *45*, 1299-1312.

Krause, L., Enticott, P. G., Zangen, A., & Fitzgerald, P. B. (2012). The role of medial prefrontal cortex in theory of mind: A deep rTMS study. *Behavioral Brain Research*, *228*, 87-90. doi:10.1016/j.bbr.2011.11.037

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*, 911-924. doi:10.1017/s0033291703001624

Lietz, C. A., Gerdes, K. E., Sun, F., Geiger, J. M., Wagaman, M. A., & Segal, E. A. (2011). The Empathy Assessment Index (EAI): A confirmatory factor analysis of a multidimensional model of empathy. *Journal of the Society for Social Work and Research*, *2*, 104-124. doi:10.5243/jsswr.2011.6

Little, T. D. (2013). Overview and SEM foundations. In *Longitudinal structural equation modeling* (p. 2). New York, NY: Guilford Press.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 151-173. doi:10.1207/s15328007sem0902_1

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*, 285-300. doi:10.1037/a0033266

Masten, C. L., Eisenberger, N. I., Pfeifer, J. H., & Dapretto, M. (2010). Witnessing peer rejection during early adolescence: Neural correlates of empathy for experiences of social exclusion. *Social Neuroscience*, *5*, 496-507. doi:10.1080/17470919.2010.490673

Maurage, P., Grynberg, D., Noël, X., Joassin, F., Philippot, P., Hanak, C., . . . Campanella, S. (2011). Dissociation between affective and cognitive empathy in alcoholism: A specific deficit for the emotional dimension. *Alcoholism: Clinical and Experimental Research*, *35*, 1662-1668. doi:10.1111/j.1530-0277.2011.01512.x

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*, 6-20.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective*, *1*, 3-62.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411-419.

Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality*, *32*, 355-360.

Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2011). The neural correlates of empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, *24*, 235-245.

Rankin, K. P., Kramer, J. H., & Miller, B. L. (2005). Patterns of cognitive and emotional empathy in frontotemporal lobar degeneration. *Cognitive and Behavioral Neurology*, *18*, 28-36.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*, 84-95. doi:10.1080/00223891.2010.528484

Samson, A. C., Huber, O., & Gross, J. J. (2012). Emotion regulation in Asperger's syndrome and high functioning autism. *Emotion*, *12*, 659-665.

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., . . . Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *Journal of Social Psychology*, *141*, 523-536.

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

Shamay-Tsoory, S. G., Shur, S., Harari, H., & Levkovitz, Y. (2007). Neurocognitive basis of impaired empathy in schizophrenia. *Neuropsychology*, *21*, 431-438. doi:10.1037/0894-4105.21.4.431

Shamay-Tsoory, S. G., Tomer, R., Goldsher, D., Berger, B. D., & Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: Anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology*, *26*, 1113-1127. doi:10.1080/13803390490515531

Smith, A. (2006). Cognitive empathy and emotional empathy in human behavior and evolution. *Psychological Record*, *56*, 3-21.

Tangney, J. P. (1991). Moral affect: The good, the bad, and the ugly. *Journal of Personality and Social Psychology*, *61*, 598-607.

Thoma, P., Zalewski, I., von Reventlow, H. G., Norra, C., Juckel, G., & Daum, I. (2011). Cognitive and affective empathy in depression linked to executive control. *Psychiatry Research*, *189*, 373-378. doi:10.1016/j.psychres.2011.07.030

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*, 135-149

Zoll, C., & Enz, S. (2005). A questionnaire to assess affective and cognitive empathy in children. *Journal of Child Psychology*, *15*, 165-174.

Corrigendum

Chrysikou, 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. doi:1073191115599055.

In the paper "Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model" we, the authors, Chrysikou and Thompson, state that Hengartner et al. (2013) used a modified two-factor model of the IRI according to which the Empathic Concern and the Personal Distress subscales were combined as a measure of affective empathy. We clarify that, although discussed as a possible single affective empathy construct, these subscales were, in fact, analyzed separately in this study.

[1]University of Kansas, Lawrence, KS, USA

Corresponding Author:
Evangelia G. Chrysikou, Department of Psychology, University of Kansas, 1415 Jayhawk Blvd., 426 Fraser Hall, Lawrence, KS, 66045, USA.
Email: lilachrysikou@ku.edu