

European Journal of
**Psychological
Assessment**

Editor-in-Chief
Matthias Ziegler

Official Organ of the European Association of Psychological Assessment

Associate Editors
Martin Bäckström
Gary N. Burns
Laurence Claes
Marjolein Fokkema
David Gallardo-Pujol
Samuel Greiff
Christoph Kemper
Stefan Krumm
Lena Lämmle
Carolyn MacCann
René T. Proyer
Sebastian Sauer
Marit Sijbrandij

European Journal of

Psychological Assessment

Your article has appeared in a journal published by Hogrefe Publishing. This e-offprint is provided exclusively for the personal use of the authors. It may not be posted on a personal or institutional website or to an institutional or disciplinary repository.

If you wish to post the article to your personal or institutional website or to archive it in an institutional or disciplinary repository, please use either a pre-print or a post-print of your manuscript in accordance with the publication release for your article and our “Online Rights for Journal Articles” (www.hogrefe.com/journals).



The Factorial Structure of a 15-Item Version of the Italian Empathy Quotient Scale

Vincenzo Paolo Senese,¹ Annunziata De Nicola,¹ Anna Passaro,¹ and Gennaro Ruggiero²

¹Psychometric Laboratory, Department of Psychology, Second University of Naples, Italy

²Laboratory of Cognitive Science & Immersive Virtual Reality, Department of Psychology, Second University of Naples, Italy

Abstract: The Empathy Quotient (EQ) scale is a self-report measure designed to assess empathy in adults. Although the scale is widely used and has been validated into different languages, its dimensionality is still controversial, as well as it is not clear which scale version should be considered. The aim of the present study was to investigate the psychometric properties of the EQ scale. The Italian version of the EQ scale was administered to a sample of 633 adults. Results showed that the 15-item EQ scale version proposed by Muncer and Ling (2006) had excellent reliability and validity indices, with a gender invariant three-factor structure (Cognitive Empathy, Emotional Reactivity, and Social Skills) and a higher order factor of general empathy. The overall results confirmed that the 15-item EQ scale is an eligible and stable tool for the assessment of empathy.

Keywords: Empathy, Empathy Quotient scale, Factorial Structure, Measurement invariance, Italian version

Empathy can be defined as the ability to adopt the psychological point of view of another person, to understand the thoughts and the feelings of others, and to experience in vicarious ways the emotions that the other is living (e.g., Baron-Cohen, 2011; Davis, 1983). Although there is a large body of literature on empathy (e.g., Baron-Cohen, 2011), there is still no agreement on a unique operational definition of this construct. Overall researchers agree that empathy is characterized at least by two fundamental dimensions: affective and cognitive. The former refers to the ability to experience the emotion of another person (emotion perception); the cognitive dimension represents the comprehension of the others' experience and the ability to assume the perspective of the other (perspective taking; Baron-Cohen & Wheelwright, 2004; Davis, 1983). Instead, disagreements regard whether these two dimensions are independent (multidimensional vision of empathy; Davis, 1983) or expression of a unique dimension of empathy (unidimensional vision of empathy; Baron-Cohen, 2011).

To overcome the weaknesses that emerged from previous empathy scales, Baron-Cohen and Wheelwright (2004) have proposed the Empathy Quotient (EQ) scale. According to Baron-Cohen's theory (2011), empathy should be considered a unitary dimension that supports both the emotion perception and the perspective taking abilities. The EQ is a 60-item self-report scale (20 items are filler, to detract the attention from empathy, while the remaining

40 items measure empathy; Baron-Cohen & Wheelwright, 2004). Each item presents a statement and respondents are asked to indicate the degree of agreement on a four-step scale. About half of the items are reversed. Non-empathic responses are scored 0, while empathic responses receive 1 or 2 points, depending on the degree of empathy.

The EQ scale has been validated on a sample of 197 adults (Baron Cohen & Wheelwright, 2004). In their study, Baron Cohen and Wheelwright did not directly assess the latent structure of the instrument, but showed that the total scale had a good reliability, and in terms of validity, it discriminated clinical groups from matched controls, and was sensitive to gender differences.

Since its development, the EQ scale has been translated and used in different countries (for a review see Groen, Fuermaier, Den Heijer, Tucha, & Althaus, 2015). Even if all studies have confirmed the reliability and validity of the EQ, the factorial structure of the scale and the version to be considered still represent controversial issues (see Table 1).

Lawrence, Shaw, Baker, Baron-Cohen, and David (2004) administered the original version of the EQ scale to a sample of 172 adults. Results did not confirm the unidimensional structure of the 40-item scale, therefore authors proposed a 28-item version saturated by three distinct but correlated factors: Cognitive Empathy, Emotional Reactivity, and Social Skills. Wakabayashi et al. (2006) administered an

online version of the original scale to a sample of 1,761 university students. The analyses confirmed the unidimensional structure of the scale. Moreover, authors proposed a 22-item scale, by considering the items with stronger saturation. Muncer and Ling (2006) administered the original EQ scale to 362 university students and their parents. They compared the 40-item and the 28-item versions (Lawrence et al., 2004). In both cases, results did not confirm the unidimensional factorial model. Therefore, authors proposed a shorter 15-item version. Results showed good fit indices for the 15-item scale, and that items were the expression of three correlated dimensions (Cognitive Empathy, Emotional Reactivity, and Social Skills). Allison, Baron-Cohen, Wheelwright, Stone, and Muncer (2011) administered online the original 40-item version of the scale to a large sample of 5,377 adults. Authors investigated the dimensionality of the scale with the Rating Scale model. Results confirmed the unidimensional structure of the 40-item scale. Moreover, by using the model parameters, authors identified a shorter version of the scale consisting of 26 items, and then they compared this scale with the 28-item scale (Lawrence et al., 2004) and the 15-item scale (Muncer & Ling, 2006). Results showed that the unidimensional 26-item solution was the best fitting one, but only if a second measurement factor was considered. Kim and Lee (2010) translated and administered a Korean version of the EQ scale to a sample of 478 university students. They compared the 40-item, the 28-item (Lawrence et al., 2004), and the 15-item versions (Muncer & Ling, 2006). Results showed that the 15-item scale with three correlated factor structure was the one with the best fit indices. Preti et al. (2011) administered the Italian version of the EQ scale to 256 university students. Authors analyzed the factorial structure of the 40-item scale. Results did not confirm the unidimensionality of the 40-item scale. Then, they proposed a new 28-item version, saturated by three correlated dimensions. The 28-item version was only partially equivalent to that of Lawrence et al. (2004). Dimitrijević, Hanak, Vukosavljević-Gvozden, and Opačić (2012) administered a Serbian version of the EQ scale to a sample of 367 high school adolescents. Analyses did not confirm the unifactorial structure of the 40-item scale, and showed acceptable fit indices only for the 28-item scale (Lawrence et al., 2004) with three correlated factors. Groen et al. (2015) administered an online Dutch version of the EQ scale to a sample of 685 adults. They compared the 40-item, the 28-item (Lawrence et al., 2004), and the 15-item versions (Muncer & Ling, 2006). Results showed that the 15-item scale with three correlated factor structure was the one with the best fit indices, although authors considered the 28-item because showed a better subscales' reliability.

Overall, the unidimensionality of the 40-item scale was not confirmed, with the exception of the studies carried

out by Allison et al. (2011) and Wakabayashi et al. (2006) that used an online testing. The majority of the studies showed a latent structure with three correlated factors: Cognitive Empathy, Emotional Reactivity, and Social Skills (Dimitrijević et al., 2012; Groen et al., 2015; Kim & Lee, 2010; Lawrence et al., 2004; Muncer & Ling, 2006; Preti et al., 2011). As regards the scales, five versions were proposed: two of 28-item (Lawrence et al., 2004; Preti et al., 2011), a 26-item (Allison et al., 2011), a 22-item (Wakabayashi et al., 2006), and a 15-item (Muncer & Ling, 2006).

Different possible explanations may be attributed to the variability of the results, such as the variety of linguistic versions used and, probably, the heterogeneity of the populations and administration methods used. Besides these aspects, we argued that a further explicative and more relevant factor could be the statistical method used to evaluate the dimensionality. Indeed, with only two exceptions (Allison et al., 2011; Groen et al., 2015), all remaining authors have used techniques that are valid with continuous quantitative data, but not robust for ordinal response format. Moreover, no study considered robust reliability indices. Both statistical theory and simulation studies have showed that factorial analysis assuming continuous, normally distributed variables does not perform well (e.g., fit statistics, parameters, and standard error are biased) if the variables are ordinal, especially when the number of observed categories is small (e.g., five or fewer; see Muthén & Kaplan, 1992).

Determining the specific items to be considered to measure empathy in a valid and reliable way is not trivial. Indeed, the presence of the same items to measure a given dimension is a prerequisite to evaluate the invariance of the scale across groups. In the absence of measurement invariance we cannot be certain that the same construct is being assessed across versions and whether group differences are ascribable to culture or merely to measurement artifact (Vandenberg & Lance, 2000). In this perspective, for example, we cannot be certain that the Italian version of the EQ scale proposed by Preti et al. (2011) is equivalent to the other proposed measures. Indeed, Preti et al. proposed a 28-item version composed of a different set of items in respect to that of Lawrence et al. (2004).

Therefore, the aim of the present study was to investigate the psychometric properties of the Italian EQ scale by adopting techniques for ordinal variables; and to compare the factorial structure of the 28-item (Lawrence et al., 2004) and 15-item (Muncer & Ling, 2006) versions of the EQ scale. We considered only these two versions because they have been replicated across different studies. To this end, the Italian 40-item version of the EQ scale (Preti et al., 2011) was administered to a sample of healthy adults. The dimensionality of the 40-item (Baron-Cohen &

Wheelwright, 2004), the 28-item (Lawrence et al., 2004), and the 15-item (Muncer & Ling, 2006) scales was studied by means of robust confirmatory factor analysis (CFA). Moreover, to test concurrent and convergent validity of the EQ scale, the Interpersonal Reactivity Index (IRI; Davis, 1983) and the Toronto Alexithymia Scale (TAS-20; Bagby, Taylor, & Parker, 1994) were administered; whereas, to evaluate divergent validity, the Hypomania/Mania Symptom Checklist (HCL-32; Angst et al., 2005) was considered.

Finally, given that the literature showed consistent gender differences in empathy (Baron-Cohen, 2011), here, for the first time in the literature, we tested the gender invariance of the EQ scale by adopting a multiple-group CFA approach (Vandenberg & Lance, 2000; for an item response theory approach see Thissen, Steinberg, & Gerrard, 1986), before comparing male and female scores. We expected that females were more empathic than males.

Methods

Sample

Six hundred thirty-three adults, 409 women (64.6%) and 224 men (35.4%; $M_{\text{age}} = 24.3$ years, $SD = 5.9$; range = 18–62), were independently sampled from different cities of the Campania region (Italy). All participants gave their informed consent before taking part in the study on voluntary basis. The socio-economic-status (computed by the four-factor Hollingshead Index) was $M = 26.1$ ($SD = 15.2$), and the educational levels varied from middle school to college (median = high school). Male and female groups were matched as a function of age, $F(1, 631) = 1.57$, $p = .221$, and socio-economic-status, $F(1, 625) = 0.673$, $p = .412$.

Procedure

Before administering the selected measures, participants were briefly instructed about the tests. The battery of tests was administered individually in a balanced order.

Measures

Socio-Demographics

All participants completed a socio-demographic questionnaire.

The Empathy Quotient Scale

The Italian version of the 40-item EQ scale translated by Preti et al. (2011) was administered to all participants. Similarly to the original scale, item presents a statement and respondents are asked to indicate the degree of agreement on a four-step Likert-type scale, from “strongly in disagreement” to “strongly in agreement.” About half of the items are reversed. Responses were scored into three categories according to the original scoring procedure (Baron-Cohen & Wheelwright, 2004).

The Interpersonal Reactivity Index

The Italian version of the IRI scale (Davis, 1983) was administered to a randomly selected subsample of participants ($n = 150$). The scale was considered to test the concurrent validity of the EQ scale. The IRI is a multidimensional scale designed to measure both cognitive and emotional components of empathy. It consists of 28 items, divided into four subscales, each of which includes seven items: Fantasy, Perspective Taking, Empathic Concern, and Personal Distress. The IRI has showed good convergent and discriminant validity, and test-retest reliability. In this study, all the IRI scales showed an adequate reliability (Cronbach's $\alpha > .70$).

The 20-Item Toronto Alexithymia Scale

The TAS-20 is a self-report scale commonly used to measure alexithymia (Bagby et al., 1994). The Italian version of the scale was administered to a randomly selected subsample of participants ($n = 413$). The scale is composed of 20 items divided into three subscales: Difficulty Describing Feelings (five items), measuring the difficulty in describing emotions to others; Difficulty Identifying Feelings (seven items), measuring the difficulty in identifying emotions; Externally Oriented Thinking (seven items), measuring the tendency to focus attention externally. Each item presents a statement and participants are asked to indicate to what extent they agreed on a 5-point Likert scale, from “never true” to “always true.” According to the authors, it is possible to compute either a total score or a single score for each subscale. In line with Preti et al. (2011), we used only the total score to investigate the convergent validity, because it showed a negative correlation with the EQ subscales. Here, the scale showed an adequate reliability ($\alpha = .78$).

Hypomania/Mania Symptom Checklist

The HCL-32 is a self-administered questionnaire developed by Angst et al. (2005) and subsequently validated in different countries and languages. This scale measures

Table 1. Overview of studies investigating the different version of the Empathy Quotient scale across countries

Study	Year	Version	N	Age [years] M (SD)/range	Statistical analysis	Compared versions	Unidimensionality of the 40-item EQ	Proposed version
Lawrence et al.	2004	English	172	34.6 (10.8)/n.r.	PCA	40-item 28-item	NO	28-item
Wakabayashi et al.	2006	English	1,761 (online)	21.0 (2.6)/18–26	PCA	40-item 22-item	YES	22-item
Muncer and Ling	2006	English	362	26.4 (n.r.)/n.r.	CFA	40-item 28-item 15-item	NO	15-item
Kim and Lee	2010	Korean	478	27.2 (n.r.)/n.r.	CFA	40-item 28-item 15-item	NO	15-item
Preti et al.	2011	Italian	256	n.r./18–38	CFA	40-item 28-item [^]	NO	28-item [^]
Allison et al.	2011	English	5,377 (online)	30.4 (11.4)/16–78	IRT (Rating Scale model)	40-item 28-item 26-item 15-item	YES	26-item
Dimitrijević et al.	2012	Serbian	367	16.7 (1.0)/15–19	CFA	40-item 28-item	NO	28-item
Groen et al.	2015	Dutch	685 (online)	33 (14.5)/16–84	CFA (robust)	40-item 28-item 15-item	NO	28-item

Notes. N = sample size, M = mean; SD = standard deviation; n.r. = not reported; studies are presented as a function of the year of appearance and the proposed version; [^]the scale is only partially equivalent to that of Lawrence et al. (2004).

hypomania and consists of a list of possible symptoms (32-item). The symptoms considered include: behaviors (e.g., “I spend more money/too much money”), mood (e.g., “My mood is much better”), and thoughts (e.g., “I think most fast”). For each symptom participants have to say if it occurred or not in their present life, considering the last four weeks with respect to the past. The total score is the number of affirmative answers. The Italian version of the HCL-32 scale was administered to a randomly selected subsample ($n = 150$) of participants and the total score was used to test the discriminant validity. We expected a substantial independence between hypomanic symptoms and empathy. In this study, the scale showed an adequate reliability ($\alpha = .74$).

Data Analyses

The data analyses were firstly carried out to investigate the factorial structure and the measurement invariance of the EQ scale, and then to investigate the validity of the EQ scores. Confirmatory factor analysis (CFA) and measurement invariance (MI) analysis were performed with LISREL 8.71 software. All the other analyses were performed with R 3.2.0 software.

Confirmatory Factor Analysis

To assess if the theoretical structure of the EQ scale adequately fitted the observed data, and to compare the fit of

the factorial models, CFAs were carried out for each scale configuration. Preliminarily, the fit of the 40-item unidimensional scale was evaluated, then two factorial models were compared for the 28-item and the 15-item scales, respectively: a one latent factor model (1-factor model) and a three correlated latent factor model (3-factor model). In the 3-factor models, items 1, 19, 25, 26, 36, 41, 44, 52, 54, 55, and 58 were specified to load on the first factor (Cognitive Empathy, CE), items 6, 21, 22, 27, 29, 32, 42, 43, 48, 50, and 59 on the second factor (Emotional Reactivity, ER); and items 4, 8, 12, 14, 35, and 57 on the third factor (Social Skills, SS). Asymptotic covariance matrices and robust maximum likelihood estimation methods (RML) were used to test CFA models. To evaluate and compare the models, we used the Satorra-Bentler ($SB\chi^2$) and the Maximum Likelihood ($ML\chi^2$) goodness-of-fit test statistics in combination with other practical tests of fit that are less dependent on N (Kline, 2011): the comparative fit index (CFI), the root mean square error of approximation index (RMSEA), and the Akaike information criterion (AIC). The difference in $ML\chi^2$ statistics ($ML\chi^2_{diff}$) and CFI values (CFI_{diff}) and the absolute value of the AIC were used to compare the relative fits of the nested models (Kline, 2011).

Reliability

Reliability of the 15-item EQ subscales (CE, SS, and ER) and of the total scale was examined using the ordinal Cronbach's alpha, that is an unbiased estimator of the theoretical reliability for ordinal data (Gadermann, Guhn, & Bruno,

Table 2. Confirmative factor analyses goodness-of-fit indices of the EQ scales

Scale [Model]	RMSEA	CFI	AIC	ML χ^2	SB χ^2	df	ML χ^2_{diff}	df _{diff}	CFI _{diff}
40-item [1-Factor]	0.066	0.65	2,951.8	2,371.4***	2,791.8***	740	–	–	–
28-item [1-Factor]	0.063	0.95	1,353.6	5,982.4***	1,241.6***	350	–	–	–
28-item [3-Factor]	0.035	0.98	738.0	3,279.6***	620.0***	347	2,702.9***	3 ^a	0.03
15-item [1-Factor]	0.106	0.81	793.4	1,393.2***	733.4***	90	–	–	–
15-item [3-Factor]	0.049	0.96	285.75	447.40***	219.75***	87	945.8***	3 ^b	0.15

Notes. ^aThe reference model is the 28-item 1-Factor model; ^bThe reference model is the 15-item 1-Factor model; *** $p < .001$.

2012). The reliability indices were calculated by starting from the polychoric correlation matrix and by using the *alpha* function of the *psych* package (Revelle, 2015).

Construct Validity

To evaluate the validity of the 15-item EQ scale, Pearson's correlation coefficients between the 15-item EQ subscales and the criterion measure were computed. The IRI and the TAS-20 scales were considered for concurrent and convergent validity assessment, respectively, whereas the HCL-32 was considered for the divergent validity assessment. Because of multiple testing, to control the increase of Type I error, we applied the Hommel's correction to the *p*-values of the correlation coefficients (Hommel, 1988).

Measurement Invariance (MI) Analysis and Gender Differences

The measurement invariance (MI) across gender of the 15-item EQ scale was analyzed following Vandenberg and Lance (2000). The omnibus test of the equality of covariance matrices across groups was executed to test the MI of the 15-item scale as a function of the gender. Indeed, if not rejected the equality of covariance matrices across groups can be considered as a "demonstration of overall measurement equivalence" (Vandenberg & Lance, 2000, p. 36). Covariance matrices and asymptotic covariance matrices were compared by means of LISREL and the RML method was used. Finally, the same goodness-of-fit test statistics of the CFA were considered to verify the invariance of the matrices.

To compare scores of males and females on the 15-item EQ scales a one-way MANOVA was executed. Gender was used as 2-level between-subject factor, and the total score on each of the three EQ subscales and the total 15-item scale were considered as dependent variables. The partial eta squared was computed to estimate the effect size of the mean differences. Moreover, to test if gender differences were influenced by age and socio-economic-status,

we replicated the analyses using the age and socio-economic-status as covariates (MANCOVA).

Results

Confirmatory Factor Analysis

The factor structures of the different versions of the EQ scale were also tested by means of CFAs (Table 2).

As regards the 40-item scale, fit indices did not support the single factor model. Therefore the 28-item and 15-item scales were tested.

As concerns the 28-item scale, fit indices provided a partial support for the 1-factor model, and indicated that the 3-correlated-factor model significantly improved the fit. The factor loadings table showed that the item 57 ("I don't consciously work out the rules of social situations") had a negative and not significant factor loading with the relative latent dimension (SS).

With regard to the 15-item scale, the analyses did not support the adequacy of single factor model, and indicated that the 3-correlated-factor model showed a good fit to the data and significantly improved the fit. The standardized factor loadings of the 15-item 3-factor model ranged from .50 to .79 ($M = .60$) for the CE scale, from .44 to .61 ($M = .54$) for the ER, and from .46 to .75 ($M = .59$) for the SS scale (see Table 3). Factors were rather correlated ($r = .38$, $r = .32$, and $r = .43$, respectively, for CE and the ER, the CE and the SS, and the ER and the SS), thus suggesting that it is possible to hypothesize the presence of a higher order factor of general empathy that can explain the significant correlations observed. This latter model was not tested as it is an equivalent model (Kline, 2011).

Reliability

As regards the reliability of the 15-item subscales, data showed acceptable level of internal consistency of ordinal Cronbach's α : .74, .73, and .68, respectively, for CE, ER,

Table 3. Standardized saturations of the 15-item EQ scale as a function of the factor

Stem	§Item	ªFactor		
		1 [CE]	2 [ER]	3 [SS]
I can tune into how someone else feels rapidly and intuitively.	52	.79	-	-
I am good at predicting how someone will feel.	25	.64	-	-
I can easily work out what another person might want to talk about.	54	.57	-	-
I can sense if I am intruding, even if the other person doesn't tell me.	44	.50	-	-
I am quick to spot when someone in a group is feeling awkward or uncomfortable.	26	.52	-	-
Seeing people cry doesn't really upset me.	32*	-	.61	-
I tend to get emotionally involved with a friend's problems.	59	-	.53	-
I really enjoy caring for other people.	6	-	.56	-
I usually stay emotionally detached when watching a film.	50*	-	.59	-
If I say something that someone else is offended by, I think that that's their problem, not mine.	27*	-	.44	-
I do not tend to find social situations confusing.	35	-	-	.48
I find it hard to know what to do in a social situation.	8*	-	-	.75
Friendships and relationships are just too difficult, so I tend not to bother with them.	12*	-	-	.68
I often find it difficult to judge if something is rude or polite.	14*	-	-	.58
I find it difficult to explain to others things that I understand easily, when they don't understand it first time.	4*	-	-	.46

Notes. §Item number in the original 60-item EQ scale; ªCE = Cognitive Empathy; ER = Emotional Reactivity; SS = Social Skills; *reverse item.

Table 4. Pearson's correlation coefficients between the 15-item EQ subscales and the criterion measures

Scales	EQ subscales			
	Cognitive Empathy	Emotional Reactivity	Social Skills	EQ total
ªIRI				
Fantasy scale	.064	.412***	-.156	.174*
Perspective Taking	.128	.433***	.045	.309***
Empathic Concern	.251**	.622***	.093	.489***
Personal Distress	.127	.169*	-.228**	.041
ªHCL-32	.044	-.058	.075	.027
ªTAS-20	-.131**	-.163***	-.356***	-.308***

Notes. ªn = 150; ¢n = 250; ¤n = 413; IRI: Interpersonal Reactivity Index; HCL-32: Hypomania/Mania Symptom Checklist; TAS-20: Toronto Alexithymia scale; Hommel's corrected p-values: *p < .05; **p < .01; ***p < .001.

and SS subscales. Finally, the total EQ scale showed an adequate reliability: .78.

Construct Validity

The psychometric analysis of the 15-item scale showed that the scale and subscales were reliable and had adequate concurrent, convergent, and discriminant validity (Table 4). As regards the correlation between the EQ and the IRI subscales, as expected, data revealed positive and significant correlations between the EQ-ER and all the IRI subscales. The EQ-SS subscale showed weak but negative and significant correlations with the IRI Personal Distress subscale. Finally the EQ-CE subscale showed only one weak but significant correlation with the IRI Empathic Concern subscale.

Table 5. Mean (SD) comparison as a function of Gender and EQ dimension

EQ dimensionº	Gender			F	df	ª²p#
	ªMale	ªFemales	ªTotal sample			
CE	5.2 (2.1)	5.3 (2.3)	5.3 (2.2)	0.5	1,631	0.001
ER	5.3 (2.1)	6.6 (2.3)	6.1 (2.3)	53.9***	1,631	0.079
SS	5.3 (2.3)	5.7 (2.6)	5.5 (2.5)	2.2	1,631	0.003
Total	15.8 (4.4)	17.6 (5.2)	17.0 (5.0)	18.9***	1,631	0.029

Notes. ªn = 224; ¢n = 409; ¤N = 633; ªCE = Cognitive Empathy; ER = Emotional Reactivity; SS = Social Skills; Total = 15-item scale; #Partial eta squared; ***p < .001.

With regard to the relations between the EQ subscales and the TAS-20 scale, data showed negative and significant correlations between the TAS-20 scale and all the EQ subscales.

Finally, as concerns the divergent validity analysis, data did not show significant correlation between the EQ subscales and the HCL-32 scale.

Measurement Invariance (MI) Analysis and Gender Differences

To test MI of the 15-item scale, covariance matrices were compared by contrasting males (n = 224) and females (n = 409). Results indicate a full measurement invariance of the scale across Genders, $ML\chi^2(120) = 309.5, p < .001$, $SB\chi^2(120) = 161.1, p = .007$, $RMSEA = .033$, $CFI = .99$.

The MANOVA showed a significant overall effect of the Gender on the empathy scores, Wilks' lambda = .919,

$F(3, 629) = 18.43, p < .001$, multivariate $\eta^2_p = .081$. The follow-up one-way ANOVAs showed that Gender significantly affected EQ-ER, and the total scores (Table 5), whereas no significant differences were observed between males and females on the EQ-CE and the EQ-SS subscales. Females rated themselves as more empathic than males on the emotional components of the EQ scale and on the general empathy. The MANCOVA that considered age and socio-economic-status as a covariate confirmed the same pattern of results.

Discussion and Conclusions

This research aimed at investigating the psychometric properties of the Italian version of the EQ scale by adopting techniques robust for the ordinal level, and to compare the 28-item and 15-item Italian short versions.

In line with previous studies (Dimitrijević et al., 2012; Groen et al., 2015; Kim & Lee, 2010; Lawrence et al., 2004; Muncer & Ling, 2006; Preti et al., 2011), the confirmatory factor analysis on the 40-item scale showed that the scale cannot be considered unidimensional. The confirmatory factor analysis of the 28-item scale confirmed that items can be considered as the expression of three different components of the empathy, although one item showed incongruent saturation (Item 57; "I don't consciously work out the rules of social situations"). We deem that the difference could be attributed to the statistical methodology used in the previous researches that was not robust for the ordinal scale. Indeed, the literature has shown that if the variables are ordinal, but factorial analysis assuming normally distributed variables is used, the parameters are erroneously estimated (see Muthén & Kaplan, 1992). Moreover, the only study that used robust confirmatory factor analysis showed a similar problem with this item (Groen et al., 2015). Furthermore, by the reading of the item, we may speculate that another possible explanation is that the item evaluates the cognitive component of empathy more than the assumed social ability.

The confirmatory factor analysis of the 15-item scale showed good fit indices. All items had congruent patterns with the three latent-factor structure. The factors matched the dimensions found in the previous studies (see Muncer & Ling, 2006), and suggested the presence of a higher order factor of general empathy. To verify if the scale measured the same dimension in an invariant way (Senese, Bornstein, Haynes, Rossi, & Venuti, 2012; Senese, Ruotolo, Ruggiero, & Iachini, 2012; Vandenberg & Lance, 2000), we evaluated the MI of the 15-item scale by contrasting male and female responses. In this respect, this is the first study that evaluates the MI of the EQ scale before testing the gender differences. The results showed that the 15-item

EQ is fully invariant across gender groups. Therefore, to further test the validity of the scale, we compared male and female scores on the 15-item EQ scale. Results confirmed the presence of gender self-reported differences on empathy, and, as expected, showed that, independently of age and socio-economic-status, females evaluated themselves as more empathic than males in the emotional domains of empathy; though the effect size of the difference was small ($\eta^2_p = .08$). No differences were observed with respect to cognitive and social empathy. These results are in line with previous studies that showed a greater difference between males and females on the emotional component of empathy (Baron-Cohen, 2011; Baron-Cohen & Wheelwright, 2004; Klein & Hodges, 2001), and confirm the validity of the scale. However, it is worth remarking that the gender effect observed here should be interpreted bearing in mind two aspects, that the gender difference is not related to scale bias, indeed the EQ scale was proven to be invariant across gender, and that this study is based on self-report only data; therefore we cannot distinguish if the observed gender difference reflects a women's greater willingness to declare empathic behavior consistent with sex-role stereotypes or if they have a higher level of emotional empathy. It is possible that with a different measure of empathy (e.g., an implicit measure) the gender differences would disappear (e.g., see Vellante et al., 2013). More studies are needed to better understand the gender differences on empathy.

The psychometric analysis of the 15-item scale showed that the scale and subscales were reliable and had adequate concurrent, convergent, and discriminant validity. The analysis confirmed the strong and positive correlation between the EQ Emotion Reactivity and the IRI subscales, in line with previous studies that showed a moderate association between the EQ and IRI scores, with a stronger association between the EQ Emotion Reactivity subscale and both IRI Perspective Taking and Emphatic Concern subscales (Berthoz, Wessa, Kedia, Wicker, & Grèzes, 2008; Dimitrijević et al., 2012; Kim & Lee, 2010; Lawrence et al., 2004). As regards the relation between the TAS-20 and the EQ scales, results showed significant correlations between the TAS-20 total score and 15-item EQ subscales (Preti et al., 2011). This confirms that the higher the empathy the less adults show difficulties in identifying, describing, or avoiding emotions. Finally, as expected, results showed that the EQ subscales were not associated to hypomanic symptoms.

The EQ scale has received several revisions that limit the possibility to define a unique version of the scale. Studies that investigated the dimensionality of the scale agree that the EQ scale measures three different but correlated facets of empathy, thus supporting the idea that it is possible to conceptualize empathy as a general higher order

dimension. In this study we used the Italian version of the scale to investigate its psychometric characteristics. The results confirmed a structure of three correlated factors, indicated that the 15-item scale was the one with best fit indices and excellent psychometric properties, and showed, for the first time in the literature, that the scale measures the empathy in a gender invariant way.

Further studies should replicate the analyses to investigate if the factor structure of the 15-item scale is consistent across the different language versions and, more important, if the scale shows MI across different populations. Moreover, because in this study we considered only self-report measure to test the validity of the EQ scale, further studies should investigate the relation between the 15-item EQ scores and other non-self report measures of empathy. Indeed, studies that investigated the relation between EQ scores and behavioral measures of empathy showed contrasting results (see Lawrence et al., 2004; Vellante et al., 2013).

References

- Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences, 51*, 829–835.
- Angst, J., Adolfsson, R., Benazzi, F., Gamma, A., Hantouche, E., Meyer, T. D., ... Scott, J. (2005). The HCL-32: Towards a self-assessment tool for hypomanic symptoms in outpatients. *Journal of Affective Disorders, 88*, 217–233.
- Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale-II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38*, 33–40.
- Baron-Cohen, S. (2011). *Zero degrees of empathy: A new theory of human cruelty*. London, UK: Allen Lane.
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger Syndrome or high functioning autism and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175.
- Berthoz, S., Wessa, M., Kedia, G., Wicker, B., & Grèzes, J. (2008). Cross-cultural validation of the empathy quotient in a French-speaking sample. *Canadian Journal of Psychiatry, 53*, 469–477.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.
- Dimitrijević, A., Hanak, N., Vukosavljević-Gvozden, T., & Opačić, G. (2012). Psychometric properties of the Serbian version of the Empathy Quotient (S-EQ). *Psihologija, 45*, 257–276.
- Gademmann, A. M., Guhn, M., & Bruno, D. Z. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*, 1–13.
- Groen, Y., Fuermaier, A. B. M., Den Heijer, A. E., Tucha, O., & Althaus, M. (2015). The empathy and systemizing quotient: The psychometric properties of the Dutch version and a review of the cross-cultural stability. *Journal of Autism and Developmental Disorders, 45*, 2848–2864. doi: 10.1007/s10803-015-2448-z
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika, 75*, 383–386.
- Kim, J., & Lee, S. J. (2010). Reliability and validity of the Korean version of the Empathy Quotient Scale. *Psychiatry Investigation, 7*, 24–30.
- Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*, 720–730.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- 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.
- Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the empathy quotient (EQ) scale. *Personality and Individual Differences, 40*, 1111–1119.
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19–30.
- Preti, A., Vellante, M., Baron-Cohen, S., Zucca, G., Petretto, D. R., & Masala, C. (2011). The empathy quotient: A cross-cultural comparison of the Italian version. *Cognitive Neuropsychiatry, 16*, 50–70.
- Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*, Evanston, IL: Northwestern University. Retrieved from <http://CRAN.R-project.org/package=psychVersion=1.5.8>
- Senese, V. P., Bornstein, M. H., Haynes, O. M., Rossi, G., & Venuti, P. (2012). A cross-cultural comparison of mothers' beliefs about their parenting very young children. *Infant Behavior and Development, 35*, 479–488.
- Senese, V. P., Ruotolo, F., Ruggiero, G., & Iachini, T. (2012). The Italian version of the Weinstein Noise Sensitivity Scale: Measurement invariance across age, gender, and context. *European Journal of Psychological Assessment, 28*, 118–124.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin, 99*, 118–128.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–69.
- Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The “Reading the Mind in the Eyes” test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18*, 326–354.
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., ... Weil, L. (2006). Development of short forms of the Empathy Quotient (EQShort) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences, 41*, 929–940.

Received May 25, 2015

Revision received September 15, 2015

Accepted November 3, 2015

Published online August 3, 2016

Vincenzo Paolo Senese

Psychometric Laboratory

Department of Psychology

Second University of Naples

31, Viale Ellittico

81100 Caserta

Italy

Fax +39 0823 274759

E-mail vincenzopaolo.senese@unina2.it