

MEASURING EMPATHIC TENDENCIES: RELIABILITY AND VALIDITY OF THE DUTCH VERSION OF THE INTERPERSONAL REACTIVITY INDEX

Kim DE CORTE ⁽¹⁾, Ann BUYSSE ⁽²⁾, Lesley L. VERHOFSTADT ⁽²⁾,
Herbert ROEYERS ⁽²⁾,
Koen PONNET ⁽³⁾, & Mark H. DAVIS ⁽⁴⁾
Ghent University Hospital (1), Ghent University (2), University of Antwerp (3), & Eckerd College, St. Petersburg, United States (4)

The Interpersonal Reactivity Index (IRI; Davis, 1980) is a commonly used self-report instrument designed to assess empathic tendencies. The IRI consists of four separate subscales: Perspective Taking (PT), Fantasy (FS), Empathic Concern (EC), and Personal Distress (PD). The objective of this study was to examine the psychometric properties of a Dutch version of the IRI. The IRI was administered to a Dutch sample of 651 normal functioning adults. The factor structure of the IRI was examined by using confirmatory factor analysis (CFA). The results of the CFA revealed that there is room for improvement and modification of the original theoretical model. The validity of the IRI was tested using internal criteria (i.e., scale intercorrelations) and external criteria (i.e., correlations with subscales of the EQ-i (Bar-On, 1997), the NEO-FFI (Hoekstra, Ormel, & De Fruyt, 1996), Mach-IV (Van Kenhove, Vermeir, & Verniers, 2001), Rosenberg Self-esteem Scale (Rosenberg, 1965), and the WAIS-III (Wechsler, 2000)). Overall, the internal consistency, construct validity, and factor structure of scores from the Dutch version of the IRI suggest that it is a useful instrument to measure people's self-reported empathic tendencies.

Empathy is a central component of normal social functioning, providing a foundation for pro-social behaviour (Charbonneau & Nicol, 2002), maintaining social relationships (Noller & Ruzzene, 1991), and enhancing psychological well-being (Musick & Wilson, 2003). In view of this, the value of being able to conceptualise and assess empathy reliably and validly seems clear-cut (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). In recent

Kim De Corte is Doctor, now attached at the Ghent University Hospital, Department of Child and Adolescent Psychiatry; Ann Buysse is Doctor at the Department of Experimental Clinical and Health Psychology, Ghent University; Lesley L. Verhofstadt is Doctor at the Department of Experimental Clinical and Health Psychology, Ghent University; Herbert Roeyers is Doctor at the Department of Experimental Clinical and Health Psychology, Ghent University; Koen Ponnet is Doctor, now attached at the University of Antwerp, Research Centre for Longitudinal and Life Course Studies; Mark H. Davis is Doctor at the Department of Behavioural Science, Eckerd College, St. Petersburg, USA.

Correspondence concerning this article should be addressed to Kim De Corte, Ghent University Hospital, Department of Child and Adolescent Psychiatry, De Pintelaan 185, 9000 Ghent, Belgium. E-mail: kim.decorte@uzgent.be

years, the theoretical consensus about a *multidimensional* conception of empathy that comprises both cognitive and affective components has substantially grown (Kerem, Fishman, & Josselson, 2001; Thornton & Thornton, 1995).

Over the years, various self-report measures of empathy have been developed. Currently, Davis' Interpersonal Reactivity Index (IRI; Davis, 1980) is the most widely and frequently used scale to measure individual differences in empathic tendencies (Pulos, Elison, & Lennon, 2004). Its popularity is attributable to several desirable qualities. First, this scale is the only one that is based on a multidimensional conceptualisation of empathy. Second, the IRI is regarded as the most comprehensive measure of self-reported empathic dispositions. Finally, this scale is relatively short and thus simple to administer.

Based on a multidimensional approach to empathy, the IRI was designed to assess a set of empathic tendencies, related in that they all have to do with the dispositional tendencies to be responsive to others, but also clearly discriminable from each other. To assess these different empathic dispositions, four seven-item scales were created: (a) Perspective taking (PT), the tendency to adopt another's psychological perspective, (b) Fantasy (FS), the tendency to identify strongly with fictitious characters, (c) Empathic concern (EC), the tendency to experience feelings of warmth, sympathy, and concern toward others, and (d) Personal distress (PD), the tendency to have feelings of discomfort and concern when witnessing others' negative experiences (Davis, 1994).

The PD and EC scales assess *affective* components, whereas the PT scale represents the *cognitive* component. Although the FS scale, with its focus on identifying with fictional characters, is frequently included in the "affective" components of the IRI, we find it harder to characterise it along the affective-cognitive dimension (see also Baron-Cohen & Wheelwright, 2004). The four IRI scales exist in the same instrument because they represent separate facets of what is termed "empathy". Davis and colleagues' data (see Davis, 1980, 1983; Davis & Franzoi, 1991) validated the IRI's multidimensional conceptualisation of empathy by demonstrating that the four dimensions constituted unique but related aspects of empathy and provided further evidence for this theory through predicted significant relationships of the IRI scale scores with interpersonal functioning, social competence and other empathy-related measures

A number of studies have shown that the IRI provides a reliable and valid way of measuring people's empathic tendencies via self-report (see Davis, 1994, for a review). However, although the IRI shows much promise, there is still some need to further investigate certain validity issues. First, some uncertainty about the IRI's factor structure exists as research on the structure has revealed different results. Consequently, further examination of the fac-

tor structure of the IRI is desirable – in particular, theory-based validation of its factor structure. Second, it would also be advisable to further investigate convergent and discriminant validity. Third, although the IRI has been translated into several languages (e.g., Swedish, Spanish, French, Chinese, and German), a reliable and validated Dutch version of this instrument does not yet exist.

Thus, the purpose of the present paper is to describe the development of a Dutch version of the IRI and to evaluate the psychometric properties of its obtained scores. The first goal is to examine the hypothesised four-factor structure of the IRI scores, and to assess the internal reliability of the subscale scores. The second goal is to examine the construct validity evidence for scores of the new translation using a large Dutch sample. Third, we will examine evidence for the convergent and discriminant validity of its scores by examining associations with scores of other relevant measures.

Factor structure and scale reliability of the IRI

Evidence regarding the underlying structure of the IRI has been mixed. Some studies have found the presence of a stable four-factor structure consistent with the four IRI subscales (e.g., Litvack-Miller, McDougall, & Romney, 1997), while other studies have found alternative (and mutually inconsistent) factor solutions (e.g., Alterman, McDermott, Cacciola, & Rutherford, 2003; Cliffordson, 2002; Pulos et al., 2004; Siu & Shek, 2005). For example, some studies (e.g., Cliffordson, 2002) found a higher-order model with two global factors; others found that a unidimensional structure best represented the IRI data (e.g., Alterman et al., 2003). Two possible explanations may account for this pattern. First, the majority of these studies applied an empirical model testing procedure to evaluate the best model fit, rather than testing specified a-priori model structures based on theoretical considerations. Although empirical model testing can be useful, the problem here is that the rationale for applying certain modifications to the model is determined *post hoc*, and such models cross-validate very badly (MacCallum, Roznowski, & Necowitz, 1992). A second potential explanation is that the use of different translations of the instrument may account for this pattern (Brislin, 1988).

Bearing in mind that establishing the factor structure of a measure is essential to the credibility of empirical findings and theory development (Byrne, 1994), our first goal was to evaluate the factor validity of the Dutch IRI using Confirmatory Factor Analysis (CFA). Confirming Davis' four-factor structure in a different culture would help validate the score structure of the Dutch IRI.

Construct validity of the scores of the Dutch translation

An additional goal of this investigation was to examine the validity of the Dutch IRI by examining whether the subscale scores display relationships consistent with prior work using the original IRI.

Scale intercorrelations

One method of establishing construct validity is to examine the pattern of correlations among the four IRI scales. This intercorrelation pattern has shown to be fairly consistent across prior studies (e.g., Carey, Fox, & Spraggins, 1988; Cliffordson, 2001; Pulos et al., 2004) and thus provides additional evidence for the scale scores' construct validity. Accordingly, the expected pattern of correlations between the scores of the Dutch IRI scales is as follows: a) EC scores will be significantly and positively associated with FS and PT scores, b) PD will be either negatively correlated with, or independent of PT and EC scores, and c) FS scale scores will be independent of PT and PD scores.

Gender differences

In the literature, empathy is considered a gendered belief (Shields, 1995) that entails the assumption that women are more emotional and more caring than males (Zahn-Waxler, Cole, & Barrett, 1991). Consistent with this view, women frequently score significantly higher than men do on self-reported empathy (Eisenberg & Miller, 1987; Hojat et al., 2002). More specifically, women generally score higher on all four IRI scales (e.g., Davis, 1980). Accordingly, one method for evaluating construct validity is to examine gender differences for the obtained scale scores of the Dutch IRI.

Convergent and discriminant validity of the Dutch IRI

Another approach to establishing the validity of the IRI scale scores is to examine their relationships with scores of other, related scales or instruments. The relationships between the four IRI scale scores and the scores of seven potentially associated constructs are considered in this paper. These constructs are emotional intelligence (EQ), the Big Five personality traits, Machiavellianism, self-esteem, and three intellectual ability indices. Each of these constructs, with exception of the intellectual ability indices, is expected – on theoretical, logical, or empirical grounds – to be related to one or more of the IRI scale scores.

EQ

Mayer and colleagues (e.g., Mayer, Caruso, & Salovey, 1999) have shown that EQ, defined as the ability to be aware of and express, assimilate, understand and manage one's emotions, is positively correlated with empathy indicators. However, to date the correlations between the IRI scale scores and EQ have not been investigated. To assess EQ, we used the Emotional Quotient Inventory (EQ-i; Bar-On, 1997), a self-report measure that taps the Intrapersonal, Interpersonal, Adaptability, Stress Management, and General Mood components of EQ. The Interpersonal component – the ability to be aware of and understand another's feelings (Bar-On, 1997) – is expected to be positively related to the PT and EC scales as these scales deal with one's tendency to imagine others' perspectives and to experience other-oriented emotions. The other components of the EQ-i focus more on emotional processes occurring *within* the individual (see Bar-On, 1997). Therefore, we expect the Intrapersonal, Stress Management, Adaptability, and General Mood components – all of which reflect the successful regulation of emotion – to be negatively correlated with scores on the PD scale and independent of PT and EC scores. How FS scores will relate to the scores on the different EQ-i scales is unclear.

Personality traits

One way to conceptualise and operationalize personality is in terms of five basic factors, labelled the 'Big Five': Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness (Costa & McCrae, 1992b). The NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992a) is one measure developed to operationalize the Big Five model. Empathy is expected to correlate with various traits of the 'Big Five' (Del Barrio, Aluja, & García, 2004), but again the associations between the IRI scale scores and the scores of the personality traits have not received much attention. It is expected that higher scores on the Neuroticism factor will be associated with higher scores on the PD scale (Shiner & Caspi, 2003). Higher scores on Agreeableness and Extraversion – the primary dimensions of interpersonal behaviour (Costa, McCrae, & Dye, 1991) – are expected to be positively correlated with EC and PT scores. Scores on Openness are expected to be associated with the FS, PT, and EC scale scores, since Openness shows significant positive correlations with pro-social behaviours (Kosek, 1995). No relationships are expected between Conscientiousness scores and the IRI scale scores, because it is not apparent that a tendency to be well organised, self-disciplined and dutiful will be related to one's empathic tendencies.

Machiavellianism

Machiavellianism (Mach) is regarded as a cluster of traits characterised by

distrust, cynicism, selfishness, and a tendency for interpersonal manipulation (McHoskey, Worzel, & Szyarto, 1998). Relative to people with low scores on Mach, high Machs lack interpersonal warmth as well as the ability to identify emotions; as such, they may consequently have a diminished capacity to be empathic (Wastell & Booth, 2003). Several studies provide evidence for this assumption (e.g., Valentine, Fleischman, & Godkin, 2003). In line with previous research, we expect scores on PT and EC scales to be negatively related to scores on Mach, given that both scales address the tendency to concern oneself with another's state of mind. Neither PD nor FS scale scores are expected to display a significant relation with Mach scores.

Self-esteem

Self-esteem is defined as a global orientation characterised by self-oriented positive emotionality (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). In view of this, it seems likely that self-esteem will be most strongly related (negatively) to the scores on the PD scale, as personal distress is a negative emotional reaction in response to another's distress (Batson, 1991). Because engaging in pro-social behaviour may be important in the development of feelings of self-worth (Laible, Carlo, & Roesch, 2004), we might presume that having the tendency to adopt another's perspective may be positively related to self-esteem. These theoretical assumptions are in accordance with Davis' (1983) empirical findings. Consequently, the following predictions can be made: 1) PT scores will be positively associated with self-esteem, 2) PD scores will be negatively related with self-esteem, and 3) no relations are expected between self-esteem and either FS or EC scores.

Intellectual ability indices

Previous research examining the association between the original IRI scales and measures of intelligence has found little consistent association (e.g., Davis, 1983; Mayer & Geher, 1996). In line with this, we expect to find no consistent pattern of relationships between intellectual ability – measured by means of an IQ test – and the IRI scale scores.

Method

Participants and procedure

Data were drawn from eight studies conducted with 651 Belgian participants. The participants were solicited using two methods. Advertisements were placed in magazines recruiting individuals who were willing to participate in a research project on empathy (13% of participants). In addition, a

snowball sampling procedure was used to obtain the remaining participants (87%). First, a team of research assistants recruited individuals in their personal social network. In a second step, additional participants were obtained from this initial sample. The persons who responded positively to either recruitment method were given a standard description of the research (e.g., aims and procedure). The sample consisted of 299 men (46%) and 352 women (54%). The mean age of the men was 24.48 years ($SD = 4.79$) and of the women 27.37 years ($SD = 5.42$). Seventy-three percent of the participants were unmarried, 20% cohabiting, and 7% married. After providing their informed consent, all participants completed a package of questionnaires in a quiet room as part of a wider testing session.

Materials

The composition of the questionnaire package varied across the eight studies. All the participants completed a questionnaire inquiring into self-reported empathy ($N = 651$), while the other measures were completed by only some of the participants: EQ ($n = 310$), personality traits ($n = 235$), Machiavellianism ($n = 182$), and self-esteem ($n = 221$). In two small studies ($n_1 = 37$, $n_2 = 36$), an intelligence test was administered and subsequently participants' Total IQ, Verbal IQ, and Performance IQ were calculated.

Empathy

Empathic tendencies were assessed using the Dutch version of the IRI. The English version of the IRI was previously translated into Dutch/Flemish by the fourth author (see Roeyers, Buysse, Ponnet, & De Corte, under revision). To pursue semantic equivalence to the original IRI measure, the Dutch translation was conducted in accordance with the standardised back-translation procedure (Bontempo, 1993). (The items of the final Dutch version of the IRI appear in Appendix A.) The IRI consists of 28 items. Participants are asked to indicate the extent to which each item describes them on a 5-point Likert scale ranging from 0 (does not describe me well) to 4 (describes me very well). PT, FS, EC, and PD scale scores were computed by summing the scores on the seven items, so that the minimum (0) and maximum (28) score of each subscale is the same. (We refer the reader to the results section for the internal consistency reliability of the scores on the IRI scales.)

EQ

Bar-On's Emotional Quotient Inventory (EQ-i; Bar-On, 1997) comprises 133 items, scored on a 5-point scale anchored by 1 (very seldom or not true of me) to 5 (very often true of me or true of me). This self-report measure assesses the trait indicators of EQ and provides a Total EQ score and five

composite scale scores. The five composite scales represent the Intrapersonal, Interpersonal, Adaptability, Stress Management, and General Mood components of EQ. Raw scores on scales are transformed into standard scores. The Dutch version of the EQ-i was used (Derksen, 1998). Participants were excluded if any of the four validity indices suggested that the results were invalid (see Bar-On, 1997). In the present study, 17 participants were excluded from further analysis involving emotional intelligence based on these criteria. The EQ-i produced an overall Cronbach's Alpha coefficient of .93. The reliability coefficient values for the composite scales were .91 for the Intrapersonal, .77 for the Interpersonal, .79 for the Adaptability, .81 for the Stress Management, and .84 for the General Mood EQ component.

Personality traits

The Dutch version (Hoekstra et al., 1996) of the NEO-Five Factor Inventory (NEO-FFI) is a short version of the NEO-PI-R Personality Inventory (Costa & McCrae, 1992a), and was administered to assess the Big Five personality traits: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Participants were presented 60 items, 12 for each domain, and were asked to indicate the extent to which they agreed or disagreed with each statement on a 5-point Likert scale. In this study, the Cronbach's alpha coefficients for the five personality domains were .78 for Extraversion, .69 for Agreeableness, .82 for Conscientiousness, .84 for Neuroticism, and .71 for Openness.

Machiavellianism

The Dutch version of the Mach-IV (Christie & Geis, 1970; Dutch version by Van Kenhove et al., 2001) is a 20-item inventory that measures the use of interpersonal manipulation strategies and agreement with Machiavellian statements. Items are scored on a 7-point Likert scale from 7 (strongly agree) through 4 (no opinion) to 1 (strongly disagree). Cronbach's alpha in this study was .66.

Self-esteem

The Dutch version of the Rosenberg Self-esteem Scale (Rosenberg, 1965) assesses a person's feelings of self-acceptance and self-worth. The statements of this 10-item scale are rated on a 4-point Likert scale ranging from 0 (strongly agree) to 3 (strongly disagree). Cronbach's alpha in this study was .76.

Intellectual ability indices

General intelligence was measured by means of the Dutch version of the

Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 2000). For each participant a Total IQ score, a Verbal IQ score, and a Performance IQ score was calculated.

Results

Statistical analyses

CFA was conducted using LISREL window version 8.50 (Jöreskog & Sörbom, 2001) to examine the factor structure of the IRI scores. According to the literature on CFA, the goodness-of-fit was evaluated based on several fit indices (Hu & Bentler, 1999). With a large sample size, as is the case in the present study, the χ^2 test statistic will almost certainly be significant, even when there are good-fitting models (Gerbing & Anderson, 1993)¹; therefore, χ^2/df ratio is also reported. A χ^2/df ratio of 2:1 to 5:1 is required (Marsh & Hocevar, 1985) and indicates an acceptable fit, but values of less than 3 are considered favourable in large sample analyses (Kline, 1998). In addition, we examined several indices that are less sensitive to sample size (Marsh & Balla, 1994): (1) the comparative fit index (CFI), (2) the goodness-of-fit index (GFI), (3) the adjusted goodness-of-fit index (AGFI), (4) the root mean square error of approximation (RMSEA), and (5) the standardised root mean square residual (SRMR). The GFI index is an absolute fit index, and CFI and AGFI are incremental fit indices (Jöreskog & Sörbom, 2001). For these three fit indices, values greater than 0.90 indicate an acceptable fit. RMSEA, which is a non-centrality based index, is a highly recommended tool in the evaluation of model fit. A value of about 0.05 (or less) for RMSEA would indicate a close fit of the model and a value of about 0.08 would indicate a reasonable fit. The 90% confidence interval (CI) around the RMSEA point estimate should contain 0.05 to indicate the possibility of close model-data fit (Browne & Cudeck, 1993). The fifth indicator is SRMR, a standardised summary of the average covariance residuals (Kline, 1998). A relatively good fit of the model is indicated when the SRMR is smaller than 0.08 (Hu & Bentler, 1999).

Distributional properties of the IRI items

Prior to further analyses, skewness and kurtosis statistics were used to inspect the distribution of the responses to the IRI items. All items and factors displayed skewness and kurtosis statistics within an acceptable range

¹Following a reviewer's suggestion, we refrained from reporting the results of the χ^2 tests.

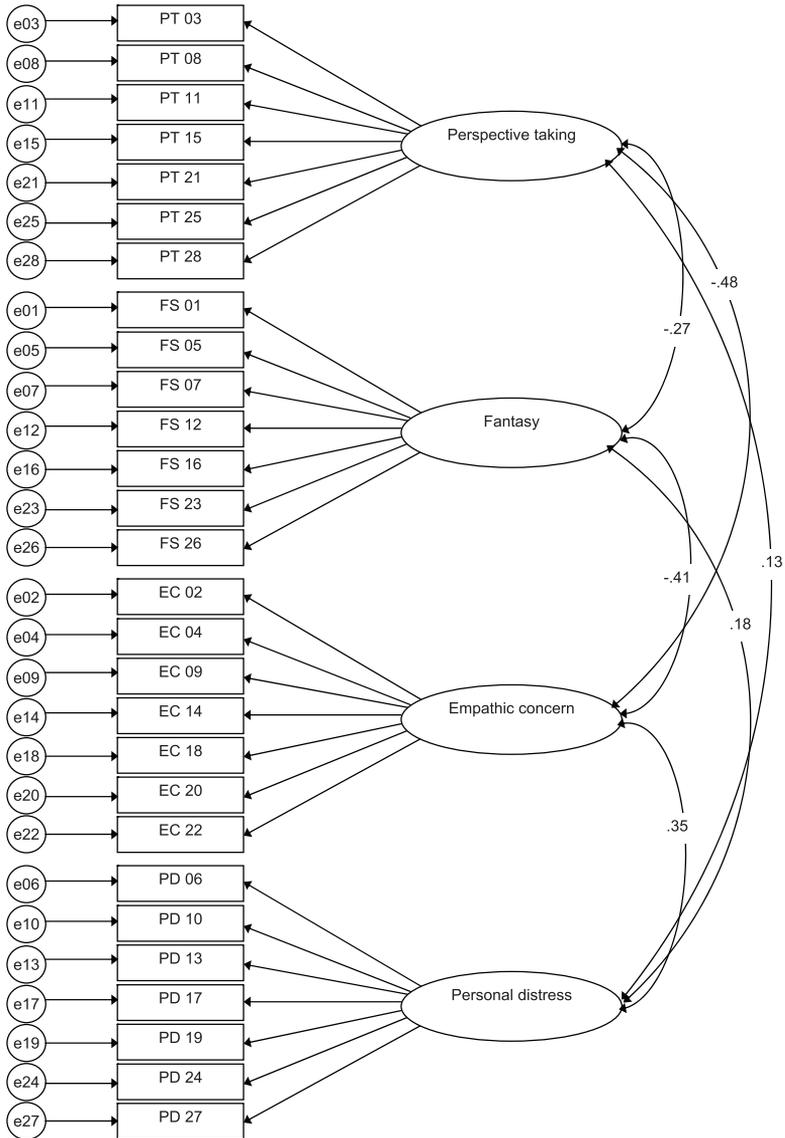
(Byrne, 1998). The percentage of missing values was negligible (0.31%), and distributed across the items. Therefore, these values were substituted with the mean value of the relevant variable (Gold & Bentler, 2000). All variables were included in the analyses, because the descriptive statistics showed that all items approximated the normal distribution (Muthén & Kaplan, 1992).

Factor structure of the IRI scores

We first attempted to replicate the four-factor structure identified by Davis (1980) by using CFA and utilising the iterated maximum likelihood procedure to estimate the four-factor model². The observed variance/covariance matrix was used for input on all analyses. Factors were allowed to correlate (analogous to an oblique rotation). Each item of the IRI was allowed to load freely on its hypothesised factor, but was not allowed to load on other factors. However, error covariances between observed variables were not allowed to correlate. The fit indices are: $\chi^2/df = 2.93$, CFI = 0.86, GFI = 0.90, AGFI = 0.87, RMSEA = 0.06 (90% CI = 0.05-0.06), SRMR = .06, AIC = 1219.06. Although some fit indices indicated an acceptable-to-reasonably-good model fit to the data (χ^2/df ratio < 3 and RMSEA = 0.06), the values of the other fit indices were acceptable but not excellent. Even though Davis' four-factor model provided a reasonable fit to the data, some improvement in model fit is possible.

An investigation of the modification indices suggested that substantial improvement in this model could be gained if error covariances of the items making up the FS scale were allowed to be estimated freely. Thus, these modification indices suggest that there is an unusually high level of semantic overlap among the FS items. Why might this be? One possibility has to do with the process by which the FS scale was initially created. The starting point was a set of three items from Stotland's Fantasy-Empathy scale (FES; Stotland, 1969); new items were then created to match their content (see Davis, 1980, 1983). The three original FES items all focus on transposing oneself into fictitious works (e.g., books, movies), and the additional four items largely reflect the same content. Given this strong semantic overlap between these seven Fantasy items, we can be more relaxed about freeing up the covariance between them (Y. Rosseel, personal communication, April 2, 2005).

²To investigate whether the IRI items measure general empathy, a hierarchical model could be tested in which the four latent factors, PT, FS, EC, and PD, load freely on one second-order latent construct, here, general empathy. However, such a model is mathematically equivalent to a four-factor model in which the four latent factors are allowed to correlate freely, and both models will provide identical fit to the data (Bollen, 1989). An examination of the correlations between the four latent factors should sufficiently inform us about their shared variance.



Note. PT = Perspective taking; FS = Fantasy; EC = Empathic concern; PD = Personal distress.

Figure 1.

Four-factor model of the IRI with factor structure identified by IRI item numbers and factor intercorrelations.

Thus, based on both theoretical arguments and modification indices, seven modifications were made to the original model (see Figure 1); specifically seven error covariances between the FS items were freed up. The error covariances to be freely estimated are between IRI item 7 and 12, between IRI item 16 and 23, between IRI item 5 and 12, between IRI item 7 and 26, between IRI item 12 and 16, between IRI item 1 and 26, and between IRI item 12 and 26. Importantly, this strategy did not require adding or deleting any paths between the observed and latent variables. This modified model was tested and the fit indices were: $\chi^2/df = 2.47$, CFI = 0.90, GFI = 0.91, AGFI = 0.90, RMSEA = 0.05 (90% CI = 0.04-0.05), SRMR = 0.06, AIC = 1014.74. The values of the fit indices for this revised CFA were noticeably improved, with relatively minor modifications.

We used Akaike's (1973) Information Criterion (AIC) to evaluate the competing models: the model with the lowest AIC is preferred (Bozdogan, 2000). As the results show (see above), the AIC criterion favours the modified four-factor model (i.e., AIC = 1014.74) rather than Davis' four-factor model (i.e., AIC = 1219.06). Since the fit indices indicated that the modified four-factor model³ offered a statistically more adequate account of the data than Davis' four-factor model, the standardised factor loadings of each item of this modified four-factor model were examined. Table 1 displays the loadings for the modified four-factor solution. As can be seen, all factor loadings are significant and above .32.

Internal reliability of the IRI scale scores

Cronbach's alpha coefficients were calculated for the scores of each of the four IRI scales. As presented in Table 2, the results indicate that the four scales of the IRI have satisfactory internal consistency in this Dutch sample.

³In addition, we examined the measurement invariance of the obtained scores on the Dutch version of the IRI across two independent groups by means of the *full measurement invariance test* (Kline, 1998). Therefore, the sample ($N = 651$) was randomly split into two independent sub-samples ($n = 325$ in each sub-sample). Both the modified four-factor model without equality constraints (i.e., unconstrained model) and the very restrictive four-factor model (i.e., constrained model: equating the factor loadings, factor correlations, and error variances) fitted the data adequately across both sub-samples. Moreover, the change in overall χ^2 between the unconstrained and the constrained model was statistically not significant, $\Delta\chi^2(69) = 4.11$, $p > .05$, indicating that the factor loadings, factor correlations, and error variances were invariant across both independent sub-samples (Vandenberg & Lance, 2000). In sum, the modified four-factor model was invariant across the two independent sub-samples.

Table 1.
Standardised factor loading (SFL) for each item of the IRI in the oblique modified four-factor solution obtained using CFA.

Item number and item content	IRI scale			
	PT	FS	EC	PD
1 I daydream and fantasise, with some regularity, about things that might happen to me.	0	.40	0	0
2 I often have tender, concerned feelings for people less fortunate than me.	0	0	.62	0
3 I sometimes find it difficult to see things from the "other guy's" point of view.	.47	0	0	0
4 Sometimes I don't feel very sorry for other people when they are having problems.	0	0	-.60	0
5 I really get involved with the feelings of the characters in a novel.	0	.88	0	0
6 In emergency situations, I feel apprehensive and ill-at-ease.	0	0	0	.81
7 I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.	0	-.74	0	0
8 I try to look at everybody's side of a disagreement before I make a decision.	-.55	0	0	0
9 When I see someone being taken advantage of, I feel kind of protective towards them.	0	0	.32	0
10 I sometimes feel helpless when I am in the middle of a very emotional situation.	0	0	0	.58
11 I sometimes try to understand my friends better by imagining how things look from their perspective.	-.49	0	0	0
12 Becoming extremely involved in a good book or movie is somewhat rare for me.	0	-.88	0	0
13 When I see someone get hurt, I tend to remain calm.	0	0	0	-.57
14 Other people's misfortunes do not usually disturb me a great deal.	0	0	-.61	0
15 If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.	.41	0	0	0
16 After seeing a play or movie, I have felt as though I were one of the characters.	0	.79	0	0
17 Being in a tense emotional situation scares me.	0	0	0	.56
18 When I see someone being treated unfairly, I sometimes don't feel very much pity for them.	0	0	-.39	0
19 I am usually pretty effective in dealing with emergencies.	0	0	0	-.60
20 I am often quite touched by things that I see happen.	0	0	.62	0
21 I believe that there are two sides to every question and try to look at them both.	-.62	0	0	0
22 I would describe myself as a pretty soft-hearted person.	0	0	.44	0
23 When I watch a good movie, I can very easily put myself in the place of a leading character.	0	.77	0	0
24 I tend to lose control during emergencies.	0	0	0	.73
25 When I'm upset at someone, I usually try to "put myself in his shoes" for a while.	0	0	0	0
26 When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.	.54	0	0	0
27 When I see someone who badly needs help in an emergency, I go to pieces.	0	.92	0	0
28 Before criticizing somebody, I try to imagine how I would feel if I were in their place.	-.59	0	0	.44

Note. (1) Factors were allowed to correlate and that (2) each item was allowed to load freely on its hypothesised factor but not allowed to load on other factors. All standardized factor loadings are significant at $p < .001$. IRI = Interpersonal Reactivity Index; PT = Perspective taking factor; FS = Fantasy factor; EC = Empathic concern factor; PD = Personal distress factor.

Table 2.
Means, standard deviations, ranges, internal reliability estimates
for the PT, FS, EC, and PD scales scores.

	IRI scale	<i>M</i>	<i>SD</i>	Range	Internal reliability	2	3	4
1	PT	17.29	4.30	3 – 28	.73	.24 **	.36 **	-.09 *
2	FS	16.48	5.91	1 – 28	.83	–	.37 **	.21 **
3	EC	18.05	4.23	3 – 28	.73		–	.27 **
4	PD	11.92	4.87	0 – 24	.77			–

Note. IRI = Interpersonal Reactivity Index; PT = Perspective taking; FS = Fantasy; EC = Empathic concern; PD = Personal distress. Each scale consists of 7 items, rated on a 5-point Likert scale (0 = does not describe me well, 4 = describes me very well).

* $p < .05$; ** $p < .001$.

Construct validity of the IRI scale scores

Scale intercorrelations

In Table 2, the relationships among the scores of the IRI scales are presented with the magnitude of the correlations ranging from $-.09$ to $.37$. As expected, EC scores were significantly and positively related to PT and FS scores. In addition, the correlation between PT and PD scores was weak; given the size of the sample, this correlation is significant, although small in size (see Cohen, 1992). Other substantial and significant positive correlations were between PD and EC scale scores, on the one hand, and between FS and both PD and PT scale scores, on the other hand.

Gender differences

To assess gender differences in scores on the four IRI scales while controlling for the multiple comparisons, we used a multivariate analysis of variance (MANOVA) with gender as the independent variable and the four IRI scales as the dependent variables. The analysis revealed a significant main effect for gender, Wilks's lambda = 0.77, $F(6, 646) = 47.44$, $p < .001$, $\eta^2 = .23$. Furthermore, the results revealed that the effect of gender was significant for all four scales, with women scoring higher than men on each one (see Table 3). The effect sizes of the FS, EC, and PD scale were in the range that Cohen (1988) describes as "large". A medium effect size, which is approximately 0.50 standard deviation units, was found for the PT scale.

Table 3.
Gender differences and effect sizes for the four IRI scales.

IRI scale	Men (<i>n</i> = 299)		Women (<i>n</i> = 352)		Effect size (<i>d</i>) ^a	<i>F</i> value ^b
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
1 PT	16.43	4.33	18.01	4.13	.37	22.69*
2 FS	14.45	6.10	18.21	5.14	.67	72.79*
3 EC	16.55	4.01	19.55	3.82	.77	112.46*
4 PD	10.26	4.55	13.32	4.68	.66	70.73*

Note. IRI = Interpersonal Reactivity Index; PT = Perspective taking; FS = Fantasy; EC = Empathic concern; PD = Personal distress. ^a The effect size measure used is Cohen's *d* (Cohen, 1988). ^b *Df* = (1, 649).

* *p* < .001.

Convergent and discriminant validity of the IRI scale scores

Table 4 displays the Pearson correlation coefficients between the four scale scores of the Dutch IRI and the scores of a variety of other instruments. Inasmuch as statistical significance is largely dependent on sample size, the effect size provides a more informative index of relations between study variables. The estimations of effect size are based upon Cohen's (1992) criteria from the magnitude of correlation coefficients: Values less than 0.1 are regarded as insubstantial, values from 0.1 to 0.3 as small, values of 0.3 to 0.5 as moderate; and values greater than 0.5 as large. In every case, the effect size could be described as small or moderate. Because samples differed greatly in size across the different instruments, differentiation between small and moderate is not considered justified.

Table 4.

Pearson correlation coefficients between the IRI scale scores and the scores of other psychological measures.

External measures	<i>n</i>	PT	FS	EC	PD
Total EQ (EQ-i) ^a	310	.31**	-.07	.02	-.29**
Intrapersonal		.11	-.07	-.05	-.29**
Interpersonal		.32**	.16*	.27**	.06
Stress Management		.21		-.12	-.24**
Adaptability		.16	-.12	.01	-.12
General Mood		.02	-.10	.01	-.23**
Personality traits (NEO-FFI) ^b	235				
Neuroticism		-.06	.14	.17	.42**
Extraversion		.03	-.05	.08	-.13
Openness	.29**	.36**	.13	-.09	
Agreeableness		.21*	.16	.31**	.08
Conscientiousness		.12	.03	.15	-.05
Machiavellism (MACH-IV) ^c	182	-.15	-.11	-.28**	.08
Self-esteem ^d	221	.16	.05	-.01	-.25**
Total IQ (WAIS-III) ^e	73	.08	.15	-.12	-.22
Verbal IQ		.13	.16	-.07	-.21
Performance IQ		-.02	.11	-.15	-.18

Note. IRI = Interpersonal Reactivity Index; PT = Perspective taking; FS = Fantasy; EC = Empathic concern; PD = Personal distress. ^a Derksen (1998); ^b Hoekstra et al. (1996); ^c Van Kenhove et al. (2001); ^d Rosenberg Self-esteem Scale (1965); ^e Wechsler (2000).

* $p < .01$, ** $p < .001$.

EQ

Correlations between IRI scores and scores on the EQ measure were largely as expected. Scores on the PT scale of the IRI were moderately and positively associated with Total EQ scores, and as expected, this was the result of PT scores being most positively related to the Interpersonal dimension of EQ. Higher PT scores were also associated with being able to cope with stress and being flexible in social settings. Scores on the EC scale were also positively associated with better “interpersonal” EQ scores, again consistent with expectations. They were not, however, associated with Total EQ. Scores on the PD scale were negatively associated with Total EQ, but this pattern resulted not from poorer interpersonal abilities, but from the predicted lower scores on the Intrapersonal dimension; higher PD scores were also associated with lower tolerance for stress and lower levels of optimism. Finally, scores on the FS scale were not related to most of the EQ domains, although higher FS scores were positively associated with greater scores on the Interpersonal measures of EQ.

Personality traits

With regard to the Big Five traits, results were again largely in accord with predictions. Scores on the PT scale were associated with being open-minded and agreeable. Scores on the EC scale were moderately and positively associated with Agreeableness scores. Scores on the PD scale were also moderately and positively associated with neuroticism. Scores on the FS scale were moderately and positively associated with greater open-mindedness. Conscientiousness and Extraversion did not display consistent relationships with any of the IRI scale scores.

Machiavellianism

Regarding Mach, results were only partially in accord with predictions. Consistent with expectations, scores on the EC scale were negatively associated with Mach, and scores on the FS and PD scales were unrelated with Mach. However, the PT scale score was non-significantly related to Mach scores, inconsistent with expectations.

Self-esteem

Scores on the PD scale were, consistent with predictions, negatively associated with self-esteem. Both the EC and FS scale score displayed no relation with self-esteem, again consistent with expectations. However, unexpectedly, higher scores on the PT scale were not significantly associated with higher self-esteem.

Intellectual ability indices

With regard to the intellectual ability indices, results were largely in accord with predictions. None of the IRI scales, except for PD, were related to any intellectual ability measure. The PD scale score was negatively associated with the Total IQ and Verbal IQ, inconsistent with expectations.

Discussion

The current study sought to examine the psychometric properties of the scores of the Dutch version of the IRI. Almost without exception, the results supported the psychometric adequacy of the scores of this version in terms of factor structure and scale reliability, construct validity as reflected in scale intercorrelations and gender differences, and the discriminant and convergent validity as evidenced by correlations with other related measures. Thus, the Dutch version of the IRI appears to be a useful complement to the original instrument.

Factor structure

By employing CFA, we examined the factor structure of the IRI. The first aim was to determine whether Davis' four-factor model – based on both empirical and theoretical considerations – represented the score structure of the Dutch IRI. Goodness of fit indices suggested that the fit between the four-factor model and the data was acceptable but not excellent. To improve model fit, we made some *post hoc* adjustments in Davis' four-factor structure by allowing within-factor correlated measurement errors for some of the items making up the FS scale. CFA revealed that this modified four-factor model provided a better fit to the data. Thus, the key question is: What can explain this need to relax certain constraints in order to achieve an adequate fit for the four-factor model?

Three possible reasons can be advanced to explain the presence of the within-factor correlated measurement errors in the IRI scores (Netemeyer, 2001). First, there may be some semantic overlap that gives rise to covariation between the FS items, above and beyond any covariation that may exist between the concepts that the FS items tap. This result might thus indicate that the unidimensional measurement of the FS factor is threatened, as an extra source of correlation in variance exists in this factor. It should be noted in this regard that unlike the items on the other three IRI scales – all of which were written expressly for the IRI – the items making up the FS scale came from two separate sources. As mentioned previously, four FS items were written for the IRI, but three others were taken from Stotland's (1969) FES. Perhaps the different origins of these two sets of items helps account for this pattern. It is also possible that something about the idiosyncrasies of translating the FS scale into Dutch created additional overlap in semantic content for this scale. It would be informative to conduct similar analyses on scores of other translations of the IRI to see if the need for within-factor correlated error variances appears for the FS scale score in other languages as well. A second possibility is that there are unwanted or unexplained sources of correlation in the variance beyond the four factors specified *a priori* in the measurement model of the IRI. In other words, it could be that the covariation between the IRI items of the FS scale has not been adequately accounted for by the four factors of Davis' factor structure. Finally, it is possible that these correlated errors are sample idiosyncratic and may not replicate to other samples. However, given that these correlated error variances were obtained from a relatively large sample (Cote, 2001), and that these correlated error variances were homogenous across two random sub-samples (see Footnote 3), it seems less likely that this pattern is a simple anomaly.

The FS scale is also one for which a reasonably strong case can be made for eliminating a scale item. IRI item 1 emerged as a relatively weak indica-

tor of the FS factor in terms of factor loading, a finding consistent with results reported by Davis (1980). Content analysis of the FS items further revealed that item 1 does not reflect the tendency to empathise with another person; in contrast, the remaining six FS items all assess the tendency to imagine oneself in another person's position. This might explain why this item appears to be a relatively weak contributor to the FS factor in the IRI. This theoretical rationale allows us to consider eliminating this item – as long as it does not appreciably reduce reliability – since doing so might generate higher overall semantic coherence within the FS scale (Frary, 2000)⁴. Whether this is also true of the English IRI and other translations remains to be seen.

Furthermore, we support other authors' assumption to question the relevance of including the FS scale for the measurement of pure empathy (Baron-Cohen & Wheelwright, 2004; Lawrence et al., 2004).

Construct validity

The internal consistency coefficients of the four Dutch IRI scales range from acceptable to high. We found relationships of relatively low strength between the IRI scale scores, which seems to be logical and theoretically meaningful. The intercorrelations among these scale scores suggest that PT, FS, EC, and PD are four statistically related but also (relatively) discriminable constructs. Moreover, the gender differences found for each of the four IRI scales are consistent with traditional gender stereotypes that women are more emotional and more caring than men (Zahn-Waxler et al., 1991) and thus perceive themselves as being more empathic than men. This pattern is also consistent with the sex differences typically found with the original IRI (Davis, 1980). Thus, these results provide additional support for the construct validity of the IRI scale scores.

⁴An examination of the internal consistency (Cronbach's Alpha) of the 6-item FS scale score ($\alpha = .85$), formed by eliminating IRI item 1, and of the original 7-item FS scale score ($\alpha = .83$) revealed an increase rather than a decrease in reliability. This indicates that the FS scale without IRI item 1 forms a more content-homogeneous subset of items than the original 7-item FS scale. Furthermore, the fit indices of another CFA based on a 27-item IRI scale (after elimination of item 1) are: $\chi^2/df = 2.67$, CFI = .90, GFI = .91, AGFI = .89, RMSEA = .05 (90% CI = 0.05-0.06), SRMR = .06, AIC = 966.10. As the results show (see above), the AIC criterion favours the modified 27-item model (i.e., AIC = 966.10) rather than the modified four-factor model (i.e., AIC = 1014.74).

Convergent and discriminant validity

Evidence for convergent and discriminant validity of the Dutch IRI scale scores came from their pattern of relationships with the scores on measures of EQ, the Big Five, Machiavellianism, self-esteem, and three intellectual ability indices. Overall, the data indicated high levels of convergent and discriminant validity for each of the four IRI scale scores.

As expected, no IRI scale score, except for the PD scale score, was associated with any of the intellectual ability. Nonetheless, this finding gives additional evidence for the discriminant validity of the scores on three IRI scales. The fact that the PD score was negatively related to Total and Verbal IQ was a surprising finding. One hypothetical explanation is that during episodes of intense distress, stable traits, like Total and Verbal IQ (Furnham, Forde, & Cotter, 1998), may be somewhat contaminated by current distress levels and might significantly lower self-esteem (e.g., Ormel & Schaufeli, 1991). In other words, people who tend to experience high levels of personal distress might report lower levels of Total and Verbal IQ due to less self-confidence.

Moreover, there were clearly different patterns of relations between the scores of each IRI scale and those of other psychological measures. The PT scale score was related to overall EQ more strongly than was any other IRI scale. This relationship was primarily due to the Interpersonal component, but the PT scale score was also associated with better Stress Management and Adaptability. Scores on the PT scale were also related to scores on the Openness and Agreeableness dimensions of the NEO-FFI. These findings seem to indicate that persons with high PT scores are able to regulate emotions and thus function smoothly in social environments. With regard to the EC scale, scores on this IRI scale were also associated with better Interpersonal EQ but not with better Stress Management or Adaptability. EC scores were also related to high scores on the Agreeableness personality trait and low scores on the Mach-IV scale. These results seem to indicate that persons with high EC scores are somewhat good-natured, warm-hearted, and non-manipulative; these are all qualities that can enhance social success.

The PD scale was negatively related to total EQ, and through a much different path than the PT scale score. PD scale scores were associated with lower ability to control one's own stress and mood, and the lack of *intrapersonal* EQ skills. Completing this pattern was the fact that higher PD scores were associated with low self-esteem and higher neuroticism. Those high in personal distress thus seem to be at the mercy of their emotions and cannot regulate them in an effective way; this may contribute to their more negative self-evaluation. The FS scale score was the only scale that was not very much related to the other psychological variables. These findings strongly suggest that the FS scale is the least 'social' of the four IRI scales.

Conclusion

In sum, the findings presented in this study give evidence for the reliability and validity of the Dutch version of the IRI and indicate that this scale is useful in measuring the perception of empathic tendencies in a Dutch sample. The findings, however, should be interpreted in the context of certain limitations. The replication of this investigation with other samples within the Dutch population could strengthen conclusions regarding the validity and reliability of the scores on the Dutch version of the IRI. Especially, since the present investigation is the first empirical analysis of the measurement invariance of the IRI, our results await replication in other samples. Furthermore, it should be taken into account that our data are based on self-report measures only as we refrained from including behavioural (non self-reported) indicators of empathic responding. Consequently, it should be noted that associations between the variables under study might be spuriously inflated or otherwise distorted due to shared method variance (Lorenz, Conger, Simons, Whitbeck, & Elder, 1991). Additionally, we did not explicitly test the IRI by comparing it with other existing empathy measures. Thus, incorporation of additional self-report empathy measures and alternative assessment methodologies into future research may provide additional evidence needed for conclusions related to convergent and discriminant validity.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest: Akademiai Kiado.
- Alterman, A.I., McDermott, P.A., Cacciola, J.S., & Rutherford, M.J. (2003). Latent structure of the Davis Interpersonal Reactivity Index in methadone maintenance patients. *Journal of Psychopathology and Behavioral Assessment*, *25*, 257-265.
- Bar-On, R. (1997). *BarOn Emotional Quotient Inventory: A measure of emotional intelligence. Facilitator's resource manual*. Toronto: Multi Health Systems.
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient (EQ). An investigation of adults with Asperger Syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163-175.
- Batson, C.D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bontempo, R. (1993). Translation fidelity of psychological scales: An item response theory analysis of individualism-collectivism scale. *Journal of Cross-Cultural Psychology*, *24*, 149-166.
- Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, *44*, 62-91.

- Brislin, R.W. (1988). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Beverly Hills, CA: Sage.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Byrne, B.M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications and programming*. Thousand Oaks, CA: Sage.
- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 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.
- Charbonneau, D., & Nicol, A.A.M. (2002). Emotional intelligence and prosocial behaviors in adolescents. *Psychological Reports*, 90, 361-370.
- Christie, R., & Geis, F.L. (1970). *Studies in Machiavellianism*. New York: Academic Press.
- Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the Interpersonal Reactivity Index (IRI). *European Journal of Psychological Assessment*, 17, 36-47.
- Cliffordson, C. (2002). The hierarchical structure of empathy: Dimensional organization and relations to social functioning. *Scandinavian Journal of Psychology*, 43, 49-59.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Costa, P.T., & McCrae, R.R. (1992a). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P.T., & McCrae, R.R. (1992b). The 5-factor model of personality and its relevance to personality-disorders. *Journal of Personality Disorders*, 6, 343-359.
- Costa, P.T., McCrae, R.R., & Dye, D.A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the Neo Personality Inventory. *Personality and Individual Differences*, 12, 887-898.
- Cote, J. (2001). Structural equation modelling. In D. Iacobucci (Ed.), *Methodological and statistical concerns of the experimental behavioural researcher*. *Journal of Consumer Psychology*, 10, 1-120.
- Davis, M.H. (1980). A multidimensional approach to individual differences in empathy. *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. (1994). *Empathy: A social psychological approach*. Colorado: Westview Press.
- Davis, M.H., & Franzoi, S.L. (1991). Stability and change in adolescent self-con-

- sciousness and empathy. *Journal of Research in Personality*, *25*, 70-87.
- Del Barrio, V., Aluja, A., & García, L.F. (2004). Relationship between empathy and the Big Five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality*, *32*, 677-681.
- Derksen, J.J. (1998). *EQ en IQ in Nederland. [EQ and IQ in the Netherlands]*. Nijmegen: Pen Test Publisher.
- Eisenberg, N., & Miller, P.A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*, 91-119.
- Frary, R.B. (2000). Higher validity in the face of lower reliability: Another look. *Applied Measurement in Education*, *13*, 249-253.
- Furnham, A.F., Forde, L.D., & Cotter, T. (1998). Personality and intelligence. *Personality and Individual Differences*, *24*, 187-192.
- Gerbing, D.W., & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.
- Gold, M.S., & Bentler, P.M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, *7*, 319-355.
- Hoekstra, H.A., Ormel, H., & De Fruyt, F. (1996). *NEO persoonlijkheidsvragenlijsten: NEO-PI-R & NEO-FFI. [NEO personality inventories: NEO-PI-R & NEO-FFI]*. Lisse: Swets & Zeitlinger.
- Hojat, M., Gonnella, J.S., Mangione, S., Nasca, T.J., Veloski, J.J., Erdmann, J.B., et al (2002). Empathy in medical students as related to academic performance, clinical competence and gender. *Medical Education*, *36*, 522-527.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- Jöreskog, K.G., & Sörbom, D. (2001). *LISREL 8.50*. Chicago, IL: Scientific Software International.
- Kerem, E., Fishman, N., & Josselson, R. (2001). The experience of empathy in everyday relationships: Cognitive and affective elements. *Journal of Social & Personal Relationships*, *18*, 709-729.
- Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kosek, R.B. (1995). Measuring prosocial behavior of college-students. *Psychological Reports*, *77*, 739-742.
- Laible, D.J., Carlo, G., & Roesch, S.C. (2004). Pathways to self-esteem in late adolescence: The role of parent and peer attachment, empathy, and social behaviors. *Journal of Adolescents*, *27*, 703-716.
- 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*, 911-919.
- Litvack-Miller, W., McDougall, D., & Romney, D.M. (1997). The structure of empathy during middle childhood and its relationship to prosocial behavior. *Genetic Social and General Psychology Monographs*, *123*, 303-324.
- Lorenz, F.O., Conger, R.D., Simons, R.L., Whitbeck, L.B., & Elder, G.H., Jr. (1991). Economic pressure and marital quality: An illustration of the method variance

- problem in the causal modeling of family processes. *Journal of Marriage and the Family*, *53*, 375-388.
- MacCallum, R.C., Roznowski, M., & Necowitz, L.B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.
- Marsh, H.W., & Balla, J.R. (1994). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality and Quantity: International Journal of Methodology*, *28*, 185-217.
- Marsh, H.W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*, 562-582.
- Mayer, J.D., Caruso, D.R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, *27*, 267-298.
- Mayer, J.D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence*, *22*, 89-113.
- McHoskey, J.W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology*, *74*, 192-210.
- Musick, M.A., & Wilson, J. (2003). Volunteering and depression: The role of psychological and social resources in different age groups. *Social Science & Medicine*, *56*, 259-269.
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of nonnormal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.
- Netemeyer, R. (2001). Structural equations modeling and statements regarding causality. *Journal of Consumer Psychology*, *10*, 83-84.
- Noller, P., & Ruzzene, M. (1991). Communication in marriage: The influence of affect and cognition. In G.J.O. Fletcher & F.D. Fincham (Eds.), *Cognitions in close relationships* (pp. 203-233). Hillsdale, NJ: Erlbaum.
- Ormel, J., & Schaufeli, W.B. (1991). Stability and change in psychological distress and their relationship with self-esteem and locus of control: A dynamic equilibrium model. *Journal of Personality and Social Psychology*, *60*, 288-299.
- Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality*, *32*, 355-359.
- Robins, R.W., Tracy, J.L., Trzesniewski, K.H., Potter, J., & Gosling, S.D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality*, *35*, 463-482.
- Roeyers, H., Buysse, A., Ponnet, K., & De Corte, K. (under revision). Perceived and performed mind-reading in adults with a pervasive developmental disorder. *Autism*.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Shields, S.A. (1995). The role of emotion beliefs and values in gender development. In N. Eisenberg (Ed.), *Social development: Review of personality and social psychology* (pp. 212-232). Thousand Oaks, CA: Sage.
- Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry*, *44*, 2-32.

- Siu, A.M.H., & Shek, D.T.L. (2005). Validation of the Interpersonal Reactivity Index in a Chinese context. *Research in Social Work Practice, 15*, 118-126.
- Stotland, E. (1969). Exploratory studies of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 271-313). New York: Academic Press.
- Thornton, S., & Thornton, D. (1995). Facets of empathy. *Personality and Individual Differences, 19*, 765-767.
- Valentine, S., Fleischman, G., & Godkin, L. (2003). Work social agency as a function of self-esteem and Machiavellianism. *Psychological Reports, 93*, 855-858.
- 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.
- Van Kenhove, P., Vermeir, I., & Verniers, S. (2001). An empirical investigation of the relationships between ethical beliefs, ethical ideology, political preference and need for closure of Dutch-speaking consumers in Belgium. *Journal of Business Ethics, 32*, 347-361.
- Wastell, C., & Booth, A. (2003). Machiavellianism: An alexithymic perspective. *Journal of Social and Clinical Psychology, 22*, 730-744.
- Wechsler, D. (2000). *WAIS-III Nederlandstalige bewerking [Dutch Wechsler adult intelligence scale]*. Lisse: Swets Test Publishers.
- Zahn-Waxler, C., Cole, P.M., & Barrett, K.C. (1991). Guilt and empathy: Sex differences and implications for the development of depression. In J. Garber & K.A. Doge (Eds.), *The development of emotion regulation and dysregulation* (pp. 243-272). New York: Cambridge University Press.

Appendix A

Items of the Dutch Version of the Interpersonal Reactivity Index

Items	
1 Ik dagdroom en fantaseer, met enige regelmaat, over dingen die zouden kunnen gebeuren met mij	FS
2 Ik heb vaak tedere, bezorgde gevoelens voor mensen die minder gelukkig zijn dan ik	EC
3 Ik vind het soms moeilijk om dingen te zien vanuit andermans gezichtspunt	PT*
4 Soms heb ik niet veel medelijden met andere mensen wanneer ze problemen hebben	EC*
5 Ik raak echt betrokken bij de gevoelens van de personages uit een roman	FS
6 In noodsituaties voel ik me ongerust en niet op mijn gemak	PD
7 Ik ben meestal objectief wanneer ik naar een film of toneelstuk kijk, en ik ga er niet vaak volledig in op	FS*
8 Ik probeer naar ieders kant van een meningsverschil te kijken alvorens ik een beslissing neem	PT
9 Wanneer ik iemand zie waarvan wordt geprofiteerd, voel ik me nogal beschermend tegenover hen	EC
10 Ik voel me soms hulpeloos wanneer ik in het midden van een zeer emotionele situatie ben	PD
11 Ik probeer mijn vrienden soms beter te begrijpen door me in te beelden hoe de dingen eruit zien vanuit hun perspectief	PT
12 Uitermate betrokken geraken in een goed boek of film is eerder zeldzaam voor mij	FS*
13 Wanneer ik zie dat iemand zich bezeert, ben ik geneigd kalm te blijven	PD*
14 Andermans ongelukken verstoren me meestal niet veel	EC*
15 Als ik zeker ben dat ik over iets gelijk heb, verspil ik niet veel tijd aan het luisteren naar andermans argumenten	PT*
16 Na het zien van een toneelstuk of film, heb ik mij gevoeld alsof ik een van de karakters was	FS
17 In een gespannen emotionele situatie zijn, schrikt me af	PD
18 Wanneer ik zie dat iemand unfair wordt behandeld, voel ik soms weinig medelijden met hen	EC*
19 Ik ben meestal behoorlijk effectief in het omgaan met noodsituaties	PD*
20 Ik ben vaak nogal geraakt door dingen die ik zie gebeuren EC 21 Ik geloof dat er twee zijden zijn aan elke vraag en probeer te kijken naar hun beide	PT
22 Ik zou mijzelf beschrijven als een vrij teerhartig persoon	EC
23 Wanneer ik naar een goede film kijk, kan ik mezelf zeer gemakkelijk in de plaats stellen van het hoofdpersonage	FS
24 Ik neig ertoe controle te verliezen tijdens noodsituaties	PD
25 Wanneer ik overstuur ben door iemand, probeer ik mijzelf meestal voor een tijdje "in zijn schoenen" te verplaatsen	PT
26 Wanneer ik een interessant verhaal of roman aan het lezen ben, beeld ik me in hoe ik me zou voelen indien de gebeurtenissen in het verhaal mij zouden overkomen	FS
27 Wanneer ik iemand zie die zeer hard hulp nodig heeft in een noodsituatie, ga ik kapot	PD
28 Alvorens iemand te bekritisieren, probeer ik mij voor te stellen hoe ik mij zou voelen mocht ik in hun plaats zijn	PT

Note. The item order of the Dutch version of the IRI is in accordance with this of the original IRI. The asterisk sign (*) indicates reversed items. PT = Perspective Taking; FS = Fantasy; EC = Empathic concern; PD = Personal distress.

Received August 20, 2006
Revision received February 8, 2008
Accepted February 8, 2008