The Psychometric Assessment of Empathy: Development and Validation of the Perth Empathy Scale

Jack D. Brett[1]*, Rodrigo Becerra[1], Murray T. Maybery[1], & David Preece[1, 2]

[1]University of Western Australia, School of Psychological Science, [2]Curtin University, School of Psychology

---

This is the **peer reviewed version of the following article**: Brett, J. D., Becerra, R., Maybery, M. T., & Preece, D. A. (2022). The Psychometric Assessment of Empathy: Development and Validation of the Perth Empathy Scale. *Assessment*. https://doi.org/10.1177/10731911221086987.

Author Note

Jack D. Brett https://orcid.org/0000-0002-5022-9925

Rodrigo Becerra https://orcid.org/0000-0001-8286-7254

Murray T. Maybery https://orcid.org/0000-0002-3760-1382

David Preece https://orcid.org/0000-0003-1060-2024

The paper's data are freely available in the Supplementary Material. We have no conflict of interest to disclose. Ethics approvals for Studies 1 and 2 were granted by the Edith Cowan University Ethics Committee, and the University of Western Australia Human Research Ethics Committee, respectively. All participants provided informed consent. The authors acknowledge the support provided by an Australian Government Research Training Program (RTP) Stipend to Jack Brett and by an Australian Research Council Discovery Project grant to Murray Maybery (DP190103286).

Author contributions: All authors jointly conceived the paper; R. B. and D. P. developed the scale and conducted Study 1; Data collection for Study 2 was carried out by J. D. B. with oversight from the other authors; Data analyses was carried out by J. D. B. with oversight from D. P.; J. D. B. wrote the manuscript with all other authors providing critical revisions.

***Corresponding author:** Jack D. Brett (jack.brett@research.uwa.edu.au)

Abstract

Empathy, the ability to infer and share others' affective states, plays a vital role in social interactions. However, no existing scale comprehensively assesses empathy's cognitive and affective components across positive and negative emotional valence domains. This paper explores the latent structure of the empathy construct and attempts to remedy past measurement limitations by developing and validating a new 20-item self-report measure, the Perth Empathy Scale (PES). In Study 1 (N = 316), factor analyses revealed a coherent empathy construct comprised of cognitive and valence-specific affective components. Study 2 (N = 331) replicated this factor structure, showed measurement invariance between males and females, and highlighted the importance of assessing negative and positive emotions in empathy. The PES showed convergent and discriminant validity from comparisons with alexithymia and other empathy measures. Overall, this paper empirically establishes a conceptually clear structure of the multidimensional empathy construct, which the PES reliably and validly measures.

Keywords: Empathy, Self-report, Questionnaire, Psychometric, Factor Structure, Alexithymia

The Psychometric Assessment of Empathy: Development and Validation of the Perth Empathy Scale

Empathy (literally in [en] passion or suffering [pathos]) is a complex psychological construct that plays a vital role in social interactions. Highly empathetic individuals are more cooperative and exhibit more moral thoughts than less empathetic individuals (Batson, 2010; Hoffman, 2008). Indeed, some developmental researchers view empathy as the root of compassion and morality (Baron-Cohen et al., 2002; de Waal, 2010). A consensus among cognitive neuroscientists and psychologists is that empathy involves the ability to infer and share the affective states of others (Cuff et al., 2016; Decety & Jackson, 2004; de Vignemont & Singer, 2006; de Waal, 2008; Shamay-Tsoory et al., 2009; Zaki & Ochsner, 2012).

Recent reviews (e.g., Cuff et al., 2016; Eklund & Meranius, 2021) of the empathy literature suggest that empathy comprises of cognitive and affective components. Indeed, multi-measure factor analyses on past empathy measures highlight the persistent presence of cognitive and affective empathy factors (e.g., Batchelder, et al., 2017; Reniers et al., 2011). Cognitive empathy is the ability to infer and thus recognize another's emotions, while affective empathy is the ability to share another's emotions. Additionally, empathizing with another's positively (e.g., happiness) and negatively (e.g., sadness) valenced emotions may represent closely related but distinct abilities (Andreychik & Migliaccio, 2015). This introduction will define empathy as used throughout the literature, highlighting the importance of its cognitive and affective components and the differentiation between positively and negatively valenced emotions. We will also touch on the importance of a self-other distinction and emotion congruence. A set of guidelines derived from the empathy literature will then be used to examine whether past scales of empathy reliably, and validly assess this framework of empathy. Lastly, as no scale adequately addresses the guidelines, we introduce and validate the Perth Empathy Scale (PES) as a new, comprehensive self-report measure.

Theoretical Background

There have been multiple definitions of what cognitive empathy encompasses. Here we refer to cognitive empathy as the ability to infer and recognize another's affective states (e.g., Coll et al., 2017; Ickes, 1993; Innamorati et al., 2019; Vachon & Lynman, 2016). Other definitions suggest that cognitive empathy is an ability to understand or take others' perspectives (e.g., Davis, 1983; Shamay-Tsoory et al., 2009). However, for empathy to arise, the self-to-other model of empathy (Bird & Viding, 2014) suggests that we infer others' affective states through two systems: affective cue classification and situation understanding. The affective cue classification system uses perceptual cues such as facial expressions and tone of voice to infer others' affective states. While, the situation understanding system uses situational cues, relying on perspective-taking, to infer others' affective states. As such, the definition of cognitive empathy used here encompasses the ability to integrate affective cues and perspective-taking to infer and recognize others' affective states.

Affective empathy is similar to the colloquial understanding of empathy—feeling the same emotions as others. Affective empathy occurs in the presence of affect sharing while the observer is aware that their emotional experience comes from another (de Vignemont & Singer, 2006). Affective responses to another's emotion occurs automatically (e.g., Zaki, 2014). However, mirroring another's emotions—emotion contagion—is not sufficient for affective empathy as the individual may appraise the emotion as their own and not make the distinction that the emotion is caused by another's affective state (Decety & Lamm, 2006; Decety & Moriguchi, 2007; de Vignemont & Singer, 2006; Gerdes et al., 2011). This self-other distinction is crucial for empathy since it prevents the confusion of self and otherexperienced emotions, which can create egocentric and altercentric biases (Bukowski et al., 2020; Lamm et al., 2016).

In addition to a self-other distinction, affective empathy requires a degree of emotion congruence (Cuff et al., 2016). Needless to say, if an individual observes someone upset and feels joy as a response, they are not empathizing. Emotion congruence further distinguishes empathy from closely related interpersonal constructs such as sympathy and compassion. Numerous researchers and clinicians suggest that empathy occurs when one feels *the same as* another, whereas sympathy and compassion happen when one feels *for* another (Clark, 2010; de Vignemont & Singer, 2006; Eisenberg & Fabes, 1990; Hein & Singer, 2008; Singer & Lamm, 2009). Indeed, empathy and sympathy are separate constructs (e.g., Bloom & Lambie, 2020; Vossen et al., 2015), relying on different neurological processes (Decety & Michalska, 2010).

Contemporary models of empathy suggest that affective and cognitive components work together to produce an overall empathy ability (Cuff et al., 2016; Decety & Lamm, 2006; Eklund & Meranius, 2021; Zaki & Ochsner, 2012). Moreover, the presence of cognitive and affective empathy is well documented (e.g., Batchelder et al., 2017; Bloom & Lambie, 2020; Gerdes et al., 2011; Innamorati et al., 2019; Joliffe & Farrington, 2006; Lietz et al., 2011; Olderbak et al., 2014; Vachon & Lynam, 2016; Vossen et al., 2015). As such, within this framework, a comprehensive measure of empathy should assess both components and ideally combine them in an overall assessment of the construct. However, to our knowledge, no study has tested whether cognitive and affective components empirically form a general empathy construct.

Research has suggested that empathizing with others' positive or negative emotions may represent distinct capabilities (Andreychik & Migliaccio, 2015). Indeed, when using behavioural assessments of empathy across emotional valences (e.g., the Multifaceted

Empathy Test; Dziobek et al., 2008) researchers have suggested that empathizing with positive and negative emotions arise from separable neurological processes (Morelli et al., 2014; Ziaei et al., 2021). Thus, assessing empathy across positive and negative domains is essential to comprehensibly encapsulate empathy. But unfortunately, no self-report measure exists to assess cognitive and affective empathy across positive and negative emotions.

Furthermore, there may be a clinical utility for assessing these valences in empathy. For example, people with social anxiety disorder often crave others' company but avoid social situations due to anxiety precipitated from these situations (Stein & Stein, 2008). Individuals suffering from this disorder find it more challenging to experience the positive emotions of those around them, although they still experience their negative emotions (Morrison et al., 2016). Thus it would be beneficial to understand why a reduced ability to share in others' positive emotions occurs in social anxiety disorder, including whether cognitive or affective empathy is impacted. Additionally, the empathy imbalance hypothesis of autism (Smith, 2009) posits that individuals on the spectrum have challenges in cognitive empathy but show intact affective empathy. However, some studies investigating the role of valence have suggested that individuals on the spectrum may be better at empathizing with positive rather than negative emotions (Mazza et al., 2014), thus highlighting the importance of assessing valence-specific empathy.

In summary, the current literature views empathy as comprised of cognitive and affective empathy with a self-other distinction and emotion congruence. Additionally, research has provided evidence that empathizing may be separable for positive and negative emotions. Therefore, within this theoretical framework, comprehensive empathy measures need to ideally meet the following guidelines: (1) assess cognitive and affective empathy; (2) assess negative and positive valences of empathy; and take into account the (3) self-other distinction and (4) emotion congruency; and lastly, (5) be psychometrically sound (i.e., have acceptable internal consistency and good factor structure validity—see the Statistical Analyses section below for acceptable ranges for internal consistency reliability and goodness-of-fit statistics).

Existing Empathy Scales

This section examines existing self-report empathy measures following the guidelines listed above. We identified 16 self-report psychometric tools specifically designed to measure empathy, not including measures for specific population groups, such as social workers (e.g., King & Holosko, 2012) or children (e.g., Raine & Chen, 2018). In our view, none meet all five of the abovementioned measurement guidelines (See Table 1). Concerning our first guideline, nine questionnaires assess cognitive and affective empathy (i.e., Affective and Cognitive Measure of Empathy, ACME; Adolescent Measure of Empathy and Sympathy, AMES; Basic Empathy Scale, BES; Emotion Specific Empathy, ESE; Empathy Assessment Index-17 items, EAI-17; Empathy Assessment Index-26, EAI-26; Empathy Components Questionnaire, ECQ; Empathic Experience Scale, EES; & Questionnaire of Cognitive and Affective Empathy, QCAE). Concerning our second guideline, one questionnaire showed evidence of assessing differing valenced empathy (i.e., the EAI-26 assessed happy and sad affective empathy). However, this measure was created during the development of the final EAI-17 and showed some psychometric problems (Lietz et al., 2011). Concerning our third and fourth guidelines, one questionnaire explicitly aimed to account for a self-other distinction (i.e., EES), while five explicitly ensured items had emotion congruence (i.e., AMES; BES; EAI-17; EAI-26; & Empathy Index, EI). Lastly, concerning our fifth guideline, seven questionnaires show some psychometric flaws (i.e., ACME; BES; Hogan Empathy Scale, HES; Interpersonal Reactivity Index, IRI; BES; EAI-26; & QCAE; see Chrysikou & Thompson, 2016; Cross & Sharpley, 1982; Gerdes et al., 2011; Lietz et al., 2011; Murphy et al., 2020; Reniers et al., 2011). However, we will note that many of these questionnaires are widely used and thus may likely be under more scrutiny. Notably, while many questionnaires meet some of the five guidelines, none of the empathy measures assess cognitive and affective empathy across negative and positive emotions.

[Table 1 around here]

The Perth Empathy Scale (PES)

As no existing measure adequately assesses empathy within the framework advanced here (based on the five measurement guidelines), we have developed the Perth Empathy Scale (PES). The PES is a 20-item self-report measure for clinicians and researchers who want to assess cognitive and affective empathy and their valence-specificity in adults and adolescents. By investigating the utility of the PES, this research also allowed for a detailed examination of the structure of the empathy construct. Items assess if one can accurately recognize other's emotions (cognitive empathy) and whether an emotion in someone else creates that emotion in oneself (affective empathy). It has items to do this for negative and positive emotions, creating four theoretical subscales, each with five items: negative cognitive empathy (e.g., "Just by seeing or hearing someone, I know if they are feeling sad"), positive cognitive empathy (e.g., "Just by seeing or hearing someone, I know if they are feeling happy"), negative affective empathy (e.g., "When I see or hear someone who is sad, it makes me feel sad too"), and positive affective empathy (e.g., "When I see or hear someone who is happy, it makes me happy too"). In line with the recommendations of van Sonderen et al. (2013) and others (e.g., Murphy et al., 2020; Rodebaugh et al., 2007), no items are worded to require reverse-scoring.

Items were designed to account for a self-other distinction and emotion congruence to aid content validity. For affective empathy, items ask participants whether an emotion in someone else (e.g., "When I see or hear someone who is happy…") causes them to feel the *same* emotion (e.g., "…it makes me feel happy too"). This phrasing indicates that the respondent appraises their current emotion in the context of another's, thus implicating a self-other distinction. For the cognitive empathy items, the wording "just by seeing or hearing someone, I know if they are feeling [an emotion]" directs participants to respond on their tendency to infer others' emotions; thus a self-other distinction is clear.

Furthermore, based on the conceptual model of empathy, the four subscales of the PES are designed to combine into several theoretically meaningful composite scores. First, composite scores may be ascertained for cognitive and affective empathy by combining scores across subscales for positive and negative emotions. Moreover, a general empathy score may be ascertained by combining the four subscale scores. Importantly, no study has statistically tested a general empathy factor that accounts for valence-specific cognitive and affective components. If there is support for the PES in measuring general empathy, this will support the conceptual model commonly used throughout the literature (see Cuff et al., 2016; Eklund & Meranius, 2021), thus highlighting the integration of valence-specific cognitive and affective empathy.

The present studies

We report the results of the first two psychometric studies of the PES here. Study 1 included an item selection process and assessed the measure's factor structure and internal consistency reliability. Study 2 examined the replicability of these findings and assessed criterion, convergent, and discriminant validity using the IRI, QCAE, and Perth Alexithymia Questionnaire (Preece et al., 2018). In addition, we investigated measurement invariance between males and females using the data from the two studies combined. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Sample sizes in each study were larger than 300 to obtain adequate power for the current studies' analyses (see recommendations by Lei & Shiverdecker, 2020; Mundfrom et al., 2005; Myers et al., 2011; Rhemtulla et al., 2012; Wolf et al., 2013).

Study 1

Method

Participants and Procedure

Participants were recruited via three avenues: unit websites of an undergraduate psychology course at an Australian university (49.7%), word of mouth (13.3%), or social media (37.0%). While 333 participants were initially recruited, after removing 17 for finishing the questionnaire in less than 40 seconds (i.e., at a rate of fewer than two seconds per item; likely reflecting inattentive responding), the final sample consisted of 316 (mean age 32.17 years, *SD* = 12.41; 84.8% female, 15.2% male). Participants completed the PES as part of an anonymous online survey using Qualtrics software.

Materials

Perth Empathy Scale (PES). The PES is a 20-item self-report scale aimed at assessing empathy in adults and adolescents. The last author wrote the 20 PES items, which were then examined by the other authors, with all agreeing that the items had good content validity in assessing the cognitive and affective empathy constructs across negative and positive emotions. Of the 20 items, an equal number of items (5) were written to assess cognitive empathy for negative emotions, cognitive empathy for positive emotions, affective empathy for negative emotions, and affective empathy for positive emotions. Ten different emotions were used across the items (5 negative & 5 positive) to provide a diverse spread of emotional experiences. Five of these emotions came from the six basic emotions (i.e., happiness, sadness, anger, scared, & disgust; Ekman, 1999). Surprise was not included given its vagueness in valence (e.g., Noordewier & Breugelmans, 2013; Vrticka et al., 2014). Embarrassment and pride were included to provide self-conscious emotions. Lastly, amusement, calmness, and enthusiasm were included to ensure that the number of positive emotions was equivalent to the number of negative emotions.

To help isolate the source of differences and enable direct comparisons between the subscales the wording of the items and emotions were mirrored across the domains (e.g., negative cognitive items and positive cognitive items had the same wording, with differing emotions, and negative cognitive items and negative affective items had the same emotions). Respondents answer each item on a 5-point Likert scale, ranging from *almost never* (1) to *almost always* (5), in terms of how often each item statement is true of them, with higher scores indicating higher levels of empathy. The PES is freely available for use and is provided in the Supplementary Materials.

Statistical Analyses

The analyses were conducted using Mplus 8.0, except for calculating Cronbach's alpha using SPSS 27.0.1.

Factor structure. The PES factor structure was assessed via confirmatory factor analyses (CFAs). CFA was chosen over EFA given the measure's clear theoretical basis and to allow us to test and directly compare a specific set of theoretically-derived models. However, a series of exploratory structural equation models (ESEM; using geomin rotation) was conducted to provide a data-driven approach. The pattern of ESEM results did not differ from the presented CFA results—in the manuscript we present only the CFAs, with the ESEM results provided in the Supplementary Materials (see Tables S1 & S2). The 20 PES items were treated as ordinal variables due to the 5-point Likert scale and asymmetrical thresholds, so the weighted least squares mean and variance adjusted estimator (WLSMV) was used (Rhemtulla et al., 2012). Several theoretically informed models of increasing complexity were examined (see Figure 1).

[Figure 1 around here]

Model 1 was a unidimensional model, where all 20 items were specified to load on a "general empathy" factor. Model 2 was a 2-factor correlated model, where items were specified to load on “Cognitive Empathy” or “Affective Empathy” factors. Model 3 was a 3-factor correlated model, where cognitive empathy was split into positive and negative valences. Model 4 was another 3-factor correlated model, where affective empathy was split into positive and negative valences. Model 5 was a 4-factor correlated model, where cognitive and affective empathy were valence-specific; items were specified to load on “Negative Cognitive Empathy”, “Positive Cognitive Empathy”, “Negative Affective Empathy”, or “Positive Affective Empathy” factors. Given that cognitive and affective empathy are expected to combine to form general empathy, a hierarchical model of the best fitting model was examined. Additionally, a bifactor model was investigated, with a correlation between positive and negative valences of the same empathy component. Both bifactor and hierarchical models were investigated as both have certain limitations (see Gignac, 2016; Watts et al., 2019), although both investigate the presence of a broader factor.

The goodness-of-fit of the models was judged based on four fit indices: the comparative fit index (CFI), Tucker Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI and TLI values $\geq$ .90 indicated acceptable fit, and values $\geq$ .95 indicated good fit, while RMSEA and SRMR values $\leq$ .08 indicated acceptable fit, and values $\leq$ .06 indicated good fit (Bentler & Bonnet, 1980; Browne & Cudeck, 1992; Marsh et al., 2004). Model fit indices and factor loadings were examined to determine the most appropriate model. Factor loadings $\geq$ |.40| were considered meaningful (Stevens, 1992). The bifactor model had to achieve a TLI value $\geq$ .010 over the hierarchical model to be preferred (Gignac, 2016).

Internal Consistency Reliability. Cronbach’s alpha (α) and McDonald’s omega (ω) reliability coefficients were calculated. Hierarchical omega (ω_H) was also calculated for the composite scales from bifactor models (as recommended by Gignac, 2014; see Supplementary Material, Table S5, for affective empathy bifactor model). α and ω

coefficients ≥ .70 were considered acceptable, values ≥ .80 were considered good, and values ≥ .90 were considered excellent (Groth-Marnat, 2009). $\omega_H \geq .50$ were considered minimally acceptable and values ≥ .75 were considered preferred (Reise et al., 2013).

Results

Factor structure

The unidimensional model (Model 1) produced a poor fit to the data according to all examined fit indices, suggesting that the PES measured a multidimensional construct (see Table 2). The two-factor model (Model 2) displayed a better fit than Model 1, indicating that making a distinction between cognitive and affective empathy added value to the factor solution. However, RMSEA suggested a poor fit to the data. Model fit indices suggested that splitting affective empathy into positive and negative valences (Model 4) fit the data better than splitting cognitive empathy (Model 3). A warning suggested that two latent variables were highly correlated when making valence-specific cognitive and affective empathy factors (Model 5). This model's (and Model 3's) positive and negative cognitive empathy factors correlated near 1 (r = .97), suggesting they measured the same latent construct. As such, it was deemed inappropriate to separate cognitive empathy into its valence-specific factors. Therefore, Model 4—splitting affective empathy only—was the most appropriate, as it produced acceptable model fit and produced well-defined factors with no redundant items (i.e., loadings < .40) needing to be dropped (see Table 3; Supplementary Material, Table S3, for other models' factor loadings). Model 4's factors were strongly correlated (see Table 4; Gignac & Szodorai, 2016). While Model 4 had an acceptable model fit, modification indices suggested allowing items 7 and 8 to co-vary. These items are parallel-worded (i.e., amusement), and this modification improved model fit (RMSEA = .071, CFI = .965, TLI = .960, SRMR = .053). For parsimony, this modification was not included in subsequent analyses.

[Table 2 around here]

[Table 3 around here]

[Table 4 around here]

A hierarchical variant of Model 4 (Model 4h) provided acceptable to good fit, with the second-order factor loadings suggesting a well-defined general factor (see Table 3). A bifactor variant of Model 4 that included a general empathy factor (Model 4b) provided acceptable to good fit, with well-defined general and specific factors (except for positive affective empathy). The bifactor variant provided better model fit than the hierarchical variant, with the TLI increasing by .01 (above the threshold of model fit improvement). As such, the bifactor model is preferred.[1]

Internal Consistency Reliabilities

All PES subscales showed acceptable to good internal reliabilities in terms of α and ω (see Table 3). The two composite scores provided acceptable internal reliabilities in terms of ω_H and good to excellent internal reliabilities in α and ω.

Study 2

Study 2 aimed to replicate the PES's factor structure and internal reliabilities of Study 1 using an independent sample. Model 4 and Model 4b (and Model 4h in the Supplementary Materials) were assessed for invariance between the two studies and between males and females. Additionally, Study 2 investigated the PES's criterion, convergent, and discriminant validity using other empathy scales and an alexithymia scale.

Method

Participants and Procedure

Study 2 included 403 participants, although 72 were removed for completing the questionnaires too quickly (i.e., < 430 seconds in total or 2 seconds per item), self-reporting invalid data, or failing at least one of the four attention checks. The final sample consisted of

331 undergraduate students from an Australian university, aged 16 to 56 years (*M* = 20.82, *SD* = 5.33; 44.1% male, 55.6% female, .3% gender-nonconforming people). Participants completed the study to fulfil a course component. Participants completed the PES, the QCAE, the IRI, and the Perth Alexithymia Questionnaire (PAQ), in a randomized order, as part of a larger questionnaire battery.

Materials

Interpersonal Reactivity Index (IRI). The IRI (Davis, 1980) is a 28-item self-report questionnaire assessing an individual's ability to empathize. The IRI comprises questions about perspective-taking, fantasy, empathic concern, and personal distress. The present study examined the perspective-taking and empathic concern subscales as these scales most closely relate to cognitive and affective empathy (see Murphy et al., 2020). Participants rate each item on a 5-point Likert scale ranging from *does not describe me well* (1) to *does describe me very well* (5). Higher scores indicate a greater ability to empathize. In the present study, the IRI subscale scores produced adequate reliabilities (see Table 5).

Questionnaire of Cognitive and Affective Empathy (QCAE). The QCAE (Reniers et al., 2011) is a 31-item multidimensional empathy questionnaire. The QCAE comprises of questions about perspective-taking (10 items), online simulation (9 items), emotion contagion (4 items), peripheral responsivity (4 items), and proximal responsivity (4 items). The present study examined the cognitive and affective subscales from the QCAE. Participants rate each item on a 4-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (4). Higher scores indicate greater ability to empathize. In the present study, the QCAE composite scores produced adequate reliabilities (see Table 5).

Perth Alexithymia Questionnaire (PAQ). The PAQ (Preece et al., 2018) is a 24-item questionnaire that assesses alexithymia. Alexithymia refers to difficulties in recognizing one's own emotions, irrespective of others' affective states (Preece et al., 2017). In contrast, cognitive empathy is about recognizing others' emotions, and affective empathy refers to experiencing another's emotions vicariously. On the PAQ, participants respond to statements regarding themselves on a 7-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (7). The items relate to difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and an externally oriented thinking style (G-EOT). Higher scores indicate greater alexithymic traits (i.e., more difficulties processing one's own emotions). In the present study, the PAQ showed adequate to excellent reliability (see Table 5).

Statistical Analyses

The factor structure analyses, measurement invariance, and Multiple-Indicators-Multiple-Causes (MIMIC) models were conducted in Mplus 8.0; all other statistical analyses were conducted in SPSS 27.0.1.

Factor structure and internal consistency reliability. The factor structure and internal consistency reliability of the PES were examined in the same manner as Study 1. Internal consistency reliabilities were calculated for the other measures, with factor analyses being conducted to obtain ω_H and ω values (See Supplementary Material, Tables S10-S14).

Measurement Invariance. Three models were examined for measurement invariance, Model 4, its hierarchical variant (Model 4h; found in the Supplementary Material), and its bifactor variant (Model 4b). Measurement invariance was investigated between Studies 1 and 2, and between males and females. Invariance between the two studies was tested to provide further evidence of replicability and to ensure the suitability of combining both studies' data to assess measurement invariance between males and females. Past empathy measures often reveal gender differences between males and females, with females exhibiting greater empathic ability (e.g., Baron-Cohen & Wheelwright, 2004; Jolliffe & Farrington, 2006; Reniers et al., 2011). However, it is important that empathy measures are not biased by gender. Accordingly, measurement invariance between males and females was assessed. Invariance was not tested for transgender or gender-nonconforming individuals due to the small sample of these groups (*n* = 8).

Measurement invariance of the PES was investigated by evaluating four increasingly restrictive multi-group CFAs (with theta parameterization). Initially, a *configural model* assessed whether the PES comprises the same number of factors, with the same items, across groups. Second, a *metric invariance* model constrained the unstandardized factor loadings to be equal between groups. Third, a *scalar model* constrained the items thresholds to be equal across groups. The item residual variances were constrained to be one across groups to achieve identification of the configural, metric, and scalar models (see Muthén & Muthén, 2017). A model with freely estimated residuals in one group was evaluated against the scalar model to assess *residual variance invariance* or strict invariance. If strict invariance is achieved, it suggests that group differences on the PES are due to group differences in their empathic abilities. To show invariance, each sequential model cannot significantly worsen model fit, except when freeing the residuals (to assess strict invariance), which should not significantly improve. ΔCFI or ΔTLI $\leq$ .010 and ΔRMSEA $\geq$ .015 indicated significantly worsening model fit (Chen et al., 2007; Cheung & Rensvold, 2002).

Lastly, we evaluated a model constraining factor variances and covariances to be equal across groups (Variance/Covariance Invariance). While not necessarily assessing the PES's ability to measure empathy, this invariance model will help explore whether empathy components vary and relate similarly within males and females (i.e., cognitive and valence-specific affective empathy have similar relationships). To show this form of invariance, the model fit cannot worsen compared to the scalar model.

Criterion Validity. Alexithymia has been viewed as a transdiagnostic precursor to empathy deficits (Valdespino et al., 2017). Investigating the effects of alexithymia on cognitive and affective empathy suggest a negative relationship of alexithymia with cognitive empathy, but results with affective empathy have been more inconsistent (e.g., Banzhaf et al., 2018; Khosravani et al., 2020; Mul et al., 2018; Yang et al., 2020). We conducted MIMIC models to test the PES's criterion validity. In these models, a latent alexithymia variable (obtained using the subscales as observed variables) was tested for its influence on the PES factors. Model 4 was used to obtain PES factors. We were interested in whether alexithymia would influence the PES factors, thus emphasizing the clinical relevance of assessing cognitive and affective empathy, with differentiation of valence for the latter.

Furthermore, as the PES asks participants to assess their own emotions, there might be a bias for people with pronounced alexithymia, who find such introspection difficult. As such, we investigated differential item functioning for the PES using the MIMIC models (Jöreskog & Goldberger, 1975; Muthén, 1989). This analysis checks that items function the same at varying levels of alexithymia. We evaluated three MIMIC models: a null model, a saturated model, and an invariant model (as recommended by Morin et al., 2013). For differential item functioning to be present, the invariant model must show a worse model fit than the saturated model (i.e., ΔCFA & ΔTLI $\leq$ .010, or ΔRMSEA $\geq$ .015).

Convergent validity. Pearson correlations were calculated for the PES scales with the QCAE and IRI scales to assess convergent validity. It was expected that the cognitive empathy factor from the PES would correlate more strongly with the cognitive empathy factors from the QCAE and IRI compared to their affective empathy factors, and vice versa for the affective empathy PES factors. However, as the QCAE and IRI do not assess positively valenced empathy, we expected that the positively valenced affective empathy scale of the PES might not fit neatly into this pattern.

Discriminant validity. To assess the discriminant validity of the PES, we conducted a second-order EFA (principal axis factoring using direct oblimin rotation) using the PES and PAQ subscale scores as the observed variables. Factors were extracted based on the scree-plot method (Fabrigar et al., 1999). Our purpose was to establish that the PES measured a latent construct (i.e., empathy) statistically separable from alexithymia. A central difference between alexithymia and empathy is a focus on one's own emotions in alexithymia versus a focus on others' emotions in empathy, implicating a self-other distinction. If empathy and alexithymia subscales load onto separable factors, this will provide evidence that the PES may sufficiently account for the self-other distinction.

Results

Descriptive Statistics

Table 5 provides descriptive statistics and reliabilities for the QCAE, IRI, and PAQ. All variables had normal distributions with skew and kurtosis below |2| and |9|, respectively (Byrne, 2010).

Factor Structure

The factor structure followed the same pattern of results as with Study 1. Model 4 provided the most appropriate factor structure of the first-order models. Models 3 and 5 produced valence-specific cognitive empathy factors correlating at 1.00, suggesting that they measure the same latent construct. Model 4 fit indices (see Table 2) suggested adequate fit highlighting the value in splitting affective empathy into positive and negative valences, and provided well-defined factors (see Table 3; Supplementary Material Table S4 provides factor loadings for other models). Including the modification identified in Study 1 (items 7 & 8 to co-vary) improved model fit (RMSEA = .077, CFI = .954, TLI = .947). The inter-correlations of this model's factors are found in Table 4.

The hierarchical variant of Model 4 (Model 4h) provided acceptable to good model fit and well-defined factors (see Table 3). The less restrictive bifactor model additionally showed good model fit and well-defined factors (except for positive affective empathy). Unlike Study 1, the TLI did not show a significant change (ΔTLI = .003). Thus, the bifactor variant (Model 4b) did not show a practical improvement over the hierarchical variant (Model 4h). However, to keep consistency with Study 1, we will present results with the bifactor model.[2]

[Table 5 around here]

Internal Consistency Reliabilities

All measures showed acceptable to good internal reliabilities for α and ω values, and acceptable ω_H values (see Table 3). However, caution should be taken with the QCAE's reliabilities due to poor fitting structural models (see Supplementary Material, Table S10).

Measurement Invariance

Measurement invariance was conducted on Models 4 and Model 4b (see Supplementary Materials for Model 4h measurement invariance). All models showed strict invariance between both studies (see Supplementary Material, Table S15; Study 1: n = 316, Study 2: n = 323), thus, providing evidence of the PES's replicability. Measurement invariance between males and females (Male: n = 188; Female: n = 451) revealed strict invariance, suggesting that the PES is not biased by gender. The increasingly restrictive models showed no significant reductions in model fit, nor did freely estimating residuals significantly improve fit for Model 4 and Model 4b (see Table 6). Additionally, both factor structures showed Variance/Covariance invariance, suggesting that the empathy factors vary similarly between males and females and that the factors relate similarly within these groups.

[Table 6 around here]

As strict invariance was found between males and females, independent-samples t-tests were conducted to examine the differences between males and females on the PES empathy scales. Females reported higher general empathy ability compared to males, $t(637) = 2.95$, $p = .003$, $d = .26$ (Mean difference of 2.84, 95% CI [4.74, .94]; Females: $M = 66.9$, $SD = 11.27$; Males: $M = 64.1$, $SD = 10.92$). Females rated significantly higher on general affective empathy, $t(637) = 3.53$, $p < .001$, $d = .31$ (Mean difference of 1.97, 95% CI [3.07, .87]; Females: $M = 28.5$, $SD = 6.55$; Males: $M = 26.5$, $SD = 6.14$), negative affective empathy, $t(637) = 4.07$, $p < .001$, $d = .35$ (Mean difference of 1.28, 95% CI [1.90, .66]; Females: $M = 12.4$, $SD = 3.71$; Males: $M = 11.1$, $SD = 3.39$), and positive affective empathy, $t(637) = 2.01$, $p = .045$, $d = .17$ (Mean difference of .69, 95% CI [1.37, .01]; Females: $M = 16.1$, $SD = 3.94$; Males: $M = 15.4$, $SD = 4.05$). However, there was no significant difference in cognitive empathy, $t(637) = 1.44$, $p = .150$, $d = .13$ (Mean difference of .87, 95% CI [2.08, -.34]; Females: $M = 38.4$, $SD = 7.06$; Males: $M = 37.6$, $SD = 7.13$).

Criterion Validity

The structure of Model 4 was used to examine how alexithymia influenced the specific factors of the PES. The MIMIC models suggested that the PES items do not show significant differential item functioning to levels of alexithymia as the invariant model did not show significantly worse model fit indices compared to the saturated model (ΔRMSEA = .003, ΔCFI = -.008, ΔTLI = -.005; see Supplementary Material, Table S17, for the full analysis).

From the invariant MIMIC model, more pronounced levels of alexithymia were significantly associated with reduced cognitive empathy ($\beta = -.30$, $p < .001$, 95% CI [-.40, -.19]), and reduced positive affective empathy ($\beta = -.20$, $p = .002$, 95% CI [-.33, -.08]), but increased negative affective empathy ($\beta = .16$, $p = .018$, 95% CI [.02, .28]). These results provide criterion validity, highlighting that alexithymia influences all components of empathy.

Convergent Validity

The Pearson correlations (see Table 7; see Supplementary Material, Table S18, for all subscale analyses) show that the cognitive empathy scale from the PES correlated greatest with the cognitive empathy scale from the QCAE and the empathic concern and perspective-taking scales from the IRI. The negative affective empathy scale correlated greatest with the affective empathy scale of the QCAE and empathic concern from the IRI. While the positive affective empathy subscale produced similar relationships across the other scales, potentially reflecting the QCAE's and the IRI's inability to measure positively valenced empathy. Nonetheless, the PES's general affective empathy score related greatest with other affective empathy scales. As expected, the general empathy score from the PES related strongly with all the scales.

[Table 7 around here]

Discriminant Validity

An EFA was conducted on the subscales from the PAQ and the PES. The scree plot suggested two factors were present. The first factor had only the alexithymia subscales meaningfully load onto it, while the second factor had only the empathy subscales meaningfully load onto it (see Table 8). This EFA suggests that scores on the PAQ and scores on the PES load onto separate alexithymia and empathy latent variables.

[Table 8 around here]

General Discussion

Our purpose in this paper was to document the PES's development, examine its psychometric properties across two studies, and investigate the structure of the empathy construct. Until now, no self-report empathy measure allowed for the assessment of cognitive and affective empathy across positive and negative emotions. In both studies, the PES performed well on every marker of validity and reliability that we tested. The factor structure of the PES was shown to include a general cognitive empathy factor and two valence-specific affective empathy factors. Additionally, there was evidence (via hierarchical and bifactor models) that these three factors may combine to represent a general empathy ability. This factor structure was replicable and consistent with its theoretical basis that cognitive and affective empathy are components of a coherent empathy construct. Measurement invariance testing further revealed that the PES consistently measures empathy between males and females (i.e., no evidence of gender bias) and that the variability and relationships with these components remain similar across these genders. Overall, the PES appears to assess a coherent multidimensional construct that distinguishes between cognitive and valence-specific affective empathy components.

The present results suggest that affective empathy is valence-specific while cognitive empathy may not be. Thus, according to our sample's scores on the PES, people consider their cognitive empathy to be the same regardless of valence, but they regard their affective empathy differently depending on the valence. This finding is consistent with past behavioural and neuropsychological reports suggesting that empathizing with differing emotional valences represent distinct capabilities (Andreychik & Migliaccio, 2015; Morelli et al., 2014). This study extends on previous work by providing statistical evidence that this valence split in affective empathy is an important component in the structure of empathy and that cognitive empathy levels are similar across valences.

Furthermore, the present paper provides criterion validity for the PES in assessing empathy across positive and negative valences. Alexithymia, thought of as a transdiagnostic precursor of empathy deficits (Valdespino et al., 2017), produced significant relationships to all of the PES subscales. Higher levels of alexithymia related to reduced cognitive empathy and positive affective empathy but increased negative affective empathy. The influence of alexithymia on cognitive empathy is in line with past studies (Banzhaf et al., 2018; Grynberg et al., 2010), and suggests that difficulty in identifying, describing, and focussing on one's own emotions hinders the ability to infer others' emotions. Because people commonly experience both negative and positive emotions, without assessing across valences, research will not establish a comprehensive empathetic profile of those with clinical and subclinical characteristics (such as alexithymia).

The differing impacts alexithymia has on positive and negative affective empathy may be due to impairments in emotion regulation. People may attempt to downregulate negative emotions caused by another's negative feelings, and those who are unable to downregulate may experience another's negative emotion more intensely. As alexithymia hinders the ability to downregulate such emotions (see Preece et al., 2018), arguably, individuals with pronounced alexithymia could exhibit an increased tendency to experience others' negative emotions. The same argument can be made to explain the reduced positive affective empathy associated with alexithymia. Individuals may attempt to upregulate positive emotions caused by another's positive feelings. However, if alexithymia inhibits this ability, individuals with pronounced alexithymia may exhibit a reduced tendency to experience others' positive emotions, although future research will be needed to investigate this.

A measure of empathy should distinguish between the constructs of empathy and alexithymia. While closely related, a point of distinction arises from the attention to one's own emotions (alexithymia) or another's emotions (empathy). We argue that a self-other distinction helps to tease apart these two constructs. In the present study, factor analyses suggested that alexithymia and empathy are separable latent constructs when using the PES to measure empathy, demonstrating a form of discriminant validity (i.e., the PES is not a measure of people's ability to appraise and focus on their own emotions).

The PES subscale scores related, as expected, to two previous empathy self-report measures (the QCAE and IRI). While the PES measures elements of empathy (i.e., positively valenced empathy) that the other measures do not, the cognitive and affective empathy scales from the PES were more strongly related to their corresponding cognitive and affective empathy scales from the other two measures.

The PES is likely to have considerable utility in research and other contexts. For instance, the PES will examine more comprehensive empathy profiles. For instance, people on the autism spectrum have challenges in cognitive empathy but show intact affective empathy (Smith, 2009). Further, future research could assess whether the profile of empathy in autism changes as a function of emotional valence. Beyond a research context, the PES may aid within clinical contexts by assessing individuals for suitability for therapy. For instance, within a group therapy setting, individuals with greater empathy often contribute to better therapeutic relationships (Johnson et al., 2005). Furthermore, the PES can inform the extent to which empathy deficits might underlie a client's interpersonal difficulties and be used to guide individualized therapy. Lastly, some psychotherapy approaches (e.g., mentalizing) base their intervention on increasing the ability of clients to understand and put themselves in other people's positions; the PES could be used to track the outcome of such therapies. Lastly, there may be some utility for the PES to be used in the forensic context. Studies have shown that lower levels of empathy increase the risk for recidivism (Brown et al., 2012). The PES can identify which individuals have lower empathy and track their progress throughout their rehabilitation.

Limitations and some future directions

We think the introduction of the PES makes a valuable contribution, but some limitations of our studies should be noted, and addressing these will require further research. Firstly, we designed the PES with nonclinical or clinical respondents in mind. However, our results here apply only to adults from the general community and universities. Future research should investigate whether the results are replicated within clinical samples and younger age groups. Furthermore, given that the samples were participants from Australia, generalizability across other cultures should also be investigated in future studies, particularly given that other emotional constructs (e.g., alexithymia and emotion regulation) can differ across cultures (Ryder et al., 2018; Soto et al., 2011). Secondly, we did not examine the test-retest reliability of the PES, so more research is required to determine the extent to which PES scores are consistent over time. Thirdly, the concurrent validity of the PES was explored only with other self-report measures, and we did not administer any behavioural assessments of empathy. Some self-report measures of empathy have shown poor concurrent validity with behavioural assessments of empathy (Murphy & Lilienfeld, 2019). At present, it remains to be determined whether the PES can remediate these limitations of past self-report measures. Lastly, whilst our sample sizes were adequate, future studies with larger sample sizes will be important to further test the replicability and generalizability of the findings.

Conclusion

Our study demonstrates that empathy is a coherent multidimensional construct consisting of cognitive and valence-specific affective empathy, which the PES reliably and validly measures. Of note, affective empathy for positive emotions is statistically part of this empathy construct. Past measures that do not account for positive emotions in empathy may not comprehensively assess empathy. Whilst further research is needed to confirm the reliability and validity of the PES across different populations, on present evidence, the PES appears to be a valuable new empathy assessment tool.

Endnotes

[1] Investigations comparing this CFA model and a bifactor exploratory structural equation model (with target rotation), revealed no significant differences (Supplementary Material Tables S6 & S7).

[2] As with Study 1, investigations comparing this CFA model and a bifactor exploratory structural equation model (with target rotation, revealed no significant differences (Supplementary Material Table S6 & S8).

References

Andreychik, M. R., & Migliaccio, N. (2015). Empathizing with others' pain versus empathizing with others' joy: Examining the separability of positive and negative empathy and their relation to different types of social behaviors and social emotions. *Basic and Applied Social Psychology, 37*(5), 274-291. https://doi.org/10.1080/01973533.2015.1071256.

Banzhaf, C., Hoffmann, F., Kanske, P., Fan, R., Walter, H., Spengler, S., Schreiter, S., Singer, T., & Bermpohl, F. (2018). Interacting and dissociable effects of alexithymia and depression on empathy. *Psychiatry Research, 270*, 631-638. https://doi.org/10.1016/j.psychres.2018.10.045.

Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R., & Hill, J. (2002). The exact mind: Empathizing and systemizing in autism spectrum conditions. In U. Goswami (Eds.), *Blackwell handbook of childhood cognitive development* (p. 491-508). Blackwell Publishers. https://doi.org/10.1002/9780470996652.ch22.

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*(2), 163-174. https://doi.org/10.1023/B:JADD.0000022607.19833.00.

Batchelder, L., Brosnan, M., Ashwin, C. (2017). The development and validation of the empathy components questionnaire (ECQ). *PLoS One, 12*(1), e0169185. https://doi.org/10.1371.journal.pone.0169185.

Batson, C. D. (2010). Empathy-induced altruistic motivation. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (p. 15-34). American Psychological Association. https://doi.org/10.1037/12061-001.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. https://doi.org/10.1037/0033-2909.88.3.588.

Bloom, Z. D., & Lambie, G. W. (2020). The adolescent measure of empathy and sympathy in a sample of emerging adults. *Measurement and Evaluation in Counseling and Development, 53*(2), 89-103. https://doi.org/10.1080/07481756.2019.1667243.

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*(7), e285. https://doi.org/10.1038/tp/2013.61.

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews, 47*, 520-532. https://doi.org/10.1016/j.neubiorev.2014.09.021.

Brown, S., Harkins, L., & Beech, A. R. (2012). General and victim-specific empathy: Associations with actuarial risk, treatment outcome, and sexual recidivism. *Sexual Abuse: A Journal of Research and Treatment, 24*(5), 411-430. https://doi.org/10.1177/1079063211423944.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258. https://doi.org/10.1177/0049124192021002005.

Bukowski, H., Tik, M., Silani, G., Ruff, C. C., Windischberger, C., & Lamm, C. (2020). When differences matter: rTMS/fMRI reveals how differences in dispositional empathy translate to distinct neural underpinnings of self-other distinction in empathy. *Cortex, 128*, 143-161. https://doi.org/10.1016/j.cortex.2020.03.009.

Byrne, B. M. (2010). *Structural equation modelling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge/Taylor & Francis Group.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5.

Chrysikou, E. G., & Thompson, W. J. (2016). Assessing cognitive and affective empathy through the interpersonal reactivity index: An argument against a two-factor model. *Assessment*, *23*(6), 769-777. https://doi.org/10.1177/1073191115599055.

Clark, A. J. (2010). Empathy and sympathy: Therapeutic distinctions in counselling. *Journal of Mental Health Counseling, 32*(2), 95-101.

Coll, M., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? *Proposal for a new measurement framework. Neuroscience and Biobehavioural Reviews*, *83*, 132-139. http://dx.doi.org/10.1016/j.neubiorev.2017.10.009.

Cross, D. G., & Sharpley, C. F. (1982). Measurement of empathy with the Hogan Empathy Scale. *Psychological Reports, 50*(1), 62. https://doi.org/10.2466/pr0.1982.50.1.62.

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion review*, *8*(2), 144-153. https://doi.org/10.1177/1754073914558466.

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

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. https://doi.org/10.1037/0022-3514.44.1.113.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why?. *Trends in Cognitive Sciences, 10*(10), 435-441. https://doi.org/10.1016/j.tics.2006.08.008.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279-300. https://doi.org/10.1146/annurev.psych.59.103006.093625.

de Waal, F. (2010). *The Age of Empathy*. New York: Harmony Books.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71-100. https://doi.org/10.1177/1534582304267187.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal, 6*, 1146-1163. https://doi.org/10.1100/tsw.2006.221.

Decety J., & Michalska, K. J. (2010). Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Developmental Science, 13*(6), 886-899. https://doi.org/10.1111/j.1467-7687.2009.00940.x.

Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: Implications for intervention across different clinical conditions. *BioPsychoSocial Medicine, 1*, 22. https://doi.org/10.1186/1751-0759-1-22.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger Syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders, 38*, 464-473. https://doi.org/10.1007/s10803-007-0486-x.

Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behaviour. *Motivation and Emotion, 14*(2), 131-149.

Eklund, J. H., & Meranius, M. S. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling, 104*(2), 300-3007. https://doi.org/10.1016/j.pec.2020.08.022.

Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion, 98*(45-60), 16.

Eysenck, S. B. G., & Eysenck H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports, 43*(3), 1247-1255. https://doi.org/10.2466/pr0.1978.43.3f.1247.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4*, 272-299.

Gerdes, K.E., Lietz, C. A., & Sega, E. Z. (2011). Measuring empathy in the 21st century: Development of an empathy index rooted in social cognitive neuroscience and social justice. *Social Work Research*, *35*(2), 83-93. https://doi.org/10.1093/swr/35.2.83.

Gignac, G. E. (2014). On the inappropriateness of using items to calculate total scale score reliability via coefficient for multidimensional scales. *European Journal of Psychological Assessment, 30*(2), 130-139. https://doi.org/10.1027/1015-5759/a000181.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57-68. https://doi.org/10.1016/j.intell.2016.01.006.

Groth-Marnat, G. (2009). *Handbook of psychological assessment.* John Wiley & Sons.

Grynberg, D., Luminet, O., Corneille, O., Grèzes, J., & Berthoz, S. (2010). Alexithymia in the interpersonal domain: A general deficit of empathy?. *Personality and Individual Differences*, *49*(8), 845-850. https://doi.org/10.1016/j.paid.2010.07.013.

Hein, G., & Singer, T. (2008). I feel how you feel but not always: The empathic brain and its modulation. *Current Opinion in Neurobiology, 18*(2), 153-158. https://doi.org/10.1016/j.conb.2008.07.012.

Hoffman, M. L. (2008). Empathy and prosocial behaviour. In M. Lewis, J. M. Haiviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (p. 440-455). The Guilford Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307-316. https://doi.org/10.1037/h0027580.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*(4), 587-610. https://doi.org/10.1111/j.1467-6494.1993.tb00783.x.

Innamorati, M., Ebisch, S. J.H., Gallese, V., & Saggino, A. (2019). A bidimensional measure of empathy: empathic experience scale. *PLoS ONE, 14*(4), e0216164. https://doi.org/10.1371/journal.pone.0216164.

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group therapy: Multilevel structural equation models. *Journal of Counseling Psychology*, *52*(3), 310-321. https://doi.org/10.1037/0022-0167.52.3.310.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence, 29*(4), 589-611. https://doi.org/10.1016/j.adolescence.2005.08.010.

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct?. *Emotion, 16*, 11071116. https://doi.org/10.1037/emo000028.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631-639. https://doi.org/10.2307/2285946.

Khosravani, V., Ardestani, S. M. S., Alvani, A., & Amirinezhad, A. (2020). Alexithymia, empathy, negative affect and physical symptoms in patients with asthma. *Clinical Psychology & Psychotherapy*, *27*(5), 736-748. https://doi.org/10.1002/cpp.2458.

King, S. Jr., & Holosko, M. J. (2012). The development and initial validation of the empathy scale for social workers. *Research on Social Work Practice, 22*(2), 174. https://doi.org/10.1177/1049731511417136.

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self-other representations in empathy: Evidence from neurotypical function and socio-cognitive disorders. *Phil. Trans. R. Soc. B, 371*, 20150083. https://doi.org/10.1098/rstb.2015.0083.

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*(5), 911-919. https://doi.org/10.1017/s0033291703001624.

Lei, P. & Shiverdecker, L. K. (2020). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(4), 584-601. https://doi.org/10.1080/10705511.2019.1680292.

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

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341. https://doi.org/10.1207/s15328007sem1103_2.

Mazza, M., Pino, M. C., Mariano, M., Tempesta, D., Ferrara, M., De Barardis, D., Masedu, F., & Valenti, M. (2014). Affective and cognitive empathy in adolescents with autism spectrum disorder. *Frontiers in Human Neuroscience*, *7*(8), 791. https://doi.org/10.3389/fnhum.2014.00791.

Morelli, S. A., Rameson, L. T., & Lieberman, M. D. (2014). The neural components of empathy: Predicting daily prosocial behaviour. *Social Cognitive and Affective Neuroscience*, *9*(1), 39-47. https://doi.org/10.1093/scan/nss088.

Morin, A. J.S, Marsh, H. W., Nagengast, B. (2013). Exploratory structural equation modelling. In *Structural Equation Modeling: A Second Course (2nd ed.)*, pg 395-436 [ed. Hancock, G. R., & Mueller, R. O.]. Information Age Publishing.

Morrison, A. S., Mateen, M. A., Brozovich, F. A., Zaki, J., Goldin, P. R., Heimberg, R. G., & Gross, J. J. (2016). Empathy for positive and negative emotions in social anxiety disorder. *Behaviour Research and Therapy, 87*, 232-242. https://doi.org/10.1016/j.brat.2016.10.005.

Mul, C., Stagg, S. D., Herbelin, B., & Aspell, J. E. (2018). The feeling of me feeling for you: Interoception, alexithymia and empathy in autism. *Journal of Autism and Developmental Disorders*, *48*(9), 2953-2967. https://doi.org/10.1007/s10803-018-3564-3.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159-168. https://doi.org/10.1207/s15327574ijt0502_4.

Murphy, B. A., Costello, T., Watts, A., Cheong, Y. F., Berg, J., & Lilienfeld, S. O. (2020). Strengths and weaknesses of two empathy measures: A comparison of the measurement precision, construct validity, and incremental validity of two multidimensional indices. *Assessment, 27*(2), 246-260. https://doi.org/10.1177/1073191118777636.

Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioural task performance. *Psychological Assessment, 31*(8), 1062-1072. https://doi.org/10.1037/pas0000732.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th edition). Muthén & Muthén.

Muthén, B. O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, *54*(4), 557-585.

Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A mote carlo approach. *Research Quarterly for Exercise and Sport*, *82*(3), 412-423. https://doi.org/10.1080/02701367.2011.10599773.

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, *27*(7), 1326-1334. https://doi.org/10.1080/02699931.2013.777660.

Olderbak, S., Sassenrath, C., Keller, J., & Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in Psychology*, *5*, 653. https://doi.org/10.3389/fpsyg.2014.00653.

Preece, D., Becerra, R., Allan, A., Robinson, K., Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences, 119,* 341-352. https://doi.org/10.1016/j.paid.2017.08.003.

Preece, D., Becerra, R., Robinson, K., Dandy, J., & Allan, A. (2018). The psychometric assessment of alexithymia: Development and validation of the Perth Alexithymia Questionnaire. *Personality and Individual Differences, 132*, 32-44. https://doi.org/10.1016/j.paid.2018.05.011.

Raine, A., & Chen, F. R. (2018). The cognitive, affective, and somatic empathy scales (CASES) for children. *Journal of Clinical Child and Adolescent Psychology, 47*(1), 24-37. https://doi.org/10.1080/15374416.2017.1295383.

Reise, S. P., Bonifay, W., & Haviland, M. (2013). Scoring and modelling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140. https://doi.org/10.1080/00223891.2012.725437.

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84-95. https://doi.org/10.1080/00223891.2010.528484.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods, 17*(3), 354-373. https://doi.org/10.1037/a0029315.

Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). The reverse of social anxiety is not always the opposite: The reverse-scored items of the social interaction anxiety scale do not belong. *Behavior Therapy, 38*(2), 192-206. https://doi.org/10.1016/j.beth.2006.08.001.

Ryder, A. G., Sunohara, M., Dere, J., & Chentsova-Dutton, Y. E. (2018). The Cultural Shaping of Alexithymia. In O. Luminet, R. M. Bagby, & G. J. Taylor (Eds.), *Alexithymia* (1st ed., pp. 33-48). Cambridge University Press. https://doi.org/10.1017/9781108241595.005.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617-627. https://doi.org/10.1093/brain/awn279.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*, 81-96. https://doi.org/10.1111/j.1749-6632.2009.04418.x.

Smith, A. (2009). The Empathy Imbalance Hypothesis of Autism: A theoretical approach to cognitive and emotional empathy in autistic development. *The Psychological Record, 59*(2), 489-510. https://doi.org/10.1007/BF03395663.

Soto, J. A., Perez, C. R., Kim, Y. H., Lee, E. A., & Minnick, M. R. (2011). Is expressive suppression always associated with poorer psychological functioning? A cross-cultural comparison between European Americans and Hong Kong Chinese. *Emotion, 11*(6), 1450-1455. https://doi.org/10.1037/a0023340.

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62-71. https://doi.org/10.1080/00223890802484381.

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *Lancet, 371*(9618), 1115-1125. https://doi.org/10.1016//S0140-6736(08)60488-2.

Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences*. Erlbaum.

Vachon, D. D. & Lynam, D. R. (2016). Fixing the problem with empathy: development and validation of the affective and cognitive measure of empathy. *Assessment, 23*(2), 135-149. https://doi.org/10.1177/1073191114567941.

Valdespino, A., Antezana, L., Ghane, M., & Richer, J. A. (2017). Alexithymia as a transdiagnostic precursor to empathy abnormalities: The functional role of the insula. *Frontiers in Psychology, 8*, 2234. https://doi.org/10.3389/fpsyg.2017.02234.

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*(7), e68967. https://doi.org/10.1371/journal.pone.0068967.

Vossen, H. G. M., Piotrowski, J. T., Valkenburg, P. M. (2015). Development of the adolescent measure of empathy and sympathy (AMES). *Personality and Individual Differences, 74*, 66-71. https://doi.org/10.1016/j.paid.2014.09.040.

Vrticka, P., Lordier, L., Bediou, B., & Sander, D. (2014). Human amygdala response to dynamic facial expressions of positive and negative surprise. *Emotion, 14*(1), 161-169. https://doi.org/10.1037/a0034619.

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science, 7*(6), 1285-1303. https://doi.org/10.1177/2167702619855035.

Wolf, E. R., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913-934. https://doi.org/10.1177/0013164413495237.

Yang, H., Shi, H., Ni, K., Wang, Y., Cheung, E. F. C., & Chan, R. K. C. (2020). Exploring the links between alexithymia, empathy and schizotypy in college students using network analysis. *Cognitive Neuropsychiatry, 25*(4), 245-253. https://doi.org/10.1080/13546805.2020.1749039.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience, 15*(5), 675-680. https://doi.org/10.1038/nn.3085.

Ziaei, M., Oestreich, L., Reutens, D. C. & Ebner, N. C. (2021). Age-related differences in negative cognitive empathy but similarities in positive affective empathy. *Brain Structure and Function*, *226*, 1823-1840. https://doi.org/10.1007/s00429-021-02291-y.

Figures

Figure 1. Models assessed in Studies 1 and 2. Models 1, 2, 3, 4, 5. Squares indicate item numbers, and ellipses indicate latent factors. Item error terms are not displayed. **b** = bifactor model; **h** = Hierarchical Model.

Figure 1. Models assessed in Studies 1 and 2. Models 1, 2, 3, 4, 5. Squares indicate item numbers, and ellipses indicate latent factors. Item error terms are not displayed. **b** = bifactor model; **h** = Hierarchical Model.

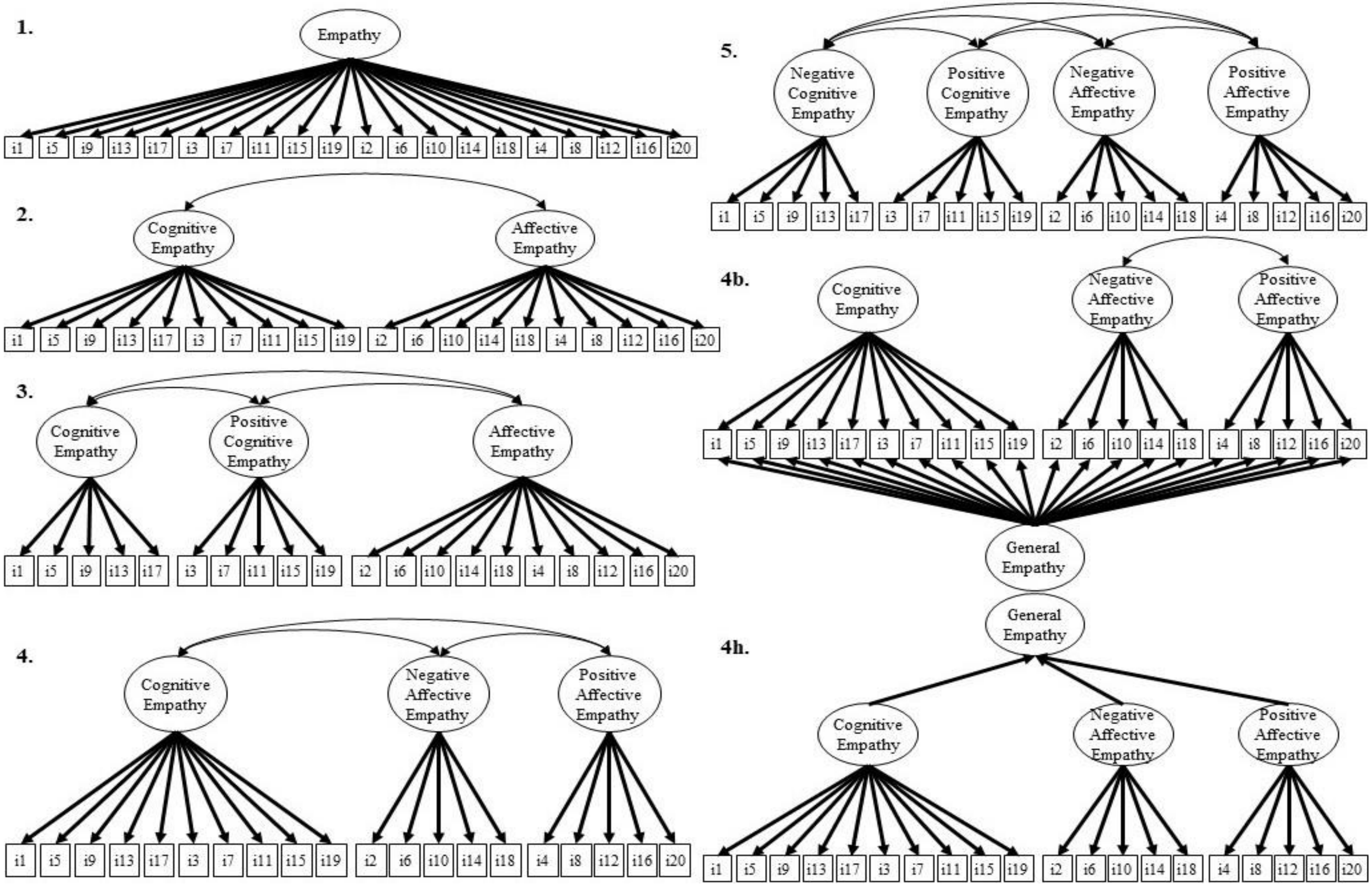

Table 1

Guidelines met by existing empathy self-report questionnaires

Name of Measure	Guidelines Met
Hogan Empathy Scale (HES; Hogan, 1969)	-
Questionnaire Measure of Emotional Empathy (QMEE; Mehrabian & Epstein, 1972)	5
Impulsiveness-Venturesomeness-Empathy Questionnaire (IVE; Eysenck & Eysenck, 1978)	5
Interpersonal Reactivity Index (IRI; Davis, 1980)	-
Empathy Quotient (EQ; Lawrence et al., 2004)	5
Basic Empathy Scale (BES; Jolliffe & Farrington, 2006)	1, 4
Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009)	5
Empathy Assessment Index-26 Items (EAI-26; Gerdes et al., 2011)	1, 2, 4
Empathy Assessment Index-17 Items (EAI-17; Lietz et al., 2011)	1, 4, 5
Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers et al., 2011)	1
Emotion Specific Empathy Questionnaire (ESE; Olderbak et al., 2014)	1, 4, 5
Adolescent Measure of Empathy and Sympathy (AMES; Vossen et al., 2015; Bloom & Lambie, 2020)	1, 4, 5
Affective and Cognitive measure of Empathy (ACME; Vachon & Lynam, 2016)	1
Empathy Index (EI; Jordan et al., 2016)	4, 5
Empathy Components Questionnaire (ECQ; Batchelder et al., 2017)	1, 5
Empathic Experience Scale (EES; Innamorati et al., 2019)	1, 3, 5

Note. This list of 16 self-report measures is based on our search of the peer-reviewed English language literature. Measures were identified if they appeared in our Google scholar searches using combinations of the term “empathy”, with either “questionnaire”, “scale”, “inventory”, “measurement”, or “assessment”. Guideline 1 = Assess cognitive and affective empathy; Guideline 2 = Positive and negative valences; Guideline 3 = Self-other distinction; Guideline 4 = Emotion Congruency; Guideline 5 = Psychometrically sound.

Table 2

Model fit of the examined PES models in Studies 1 and 2.

	Model Fit Indices				
Models	Chi-square (df)	RMSEA [90% CI]	CFI	TLI	SRMR
		Study 1			
1	1573.155 (170)*	.162 [.154/.169]*	.814	.792	.118
2	636.991 (169)*	.094 [.086/.101]*	.938	.930	.069
3	607.765 (167)*	.091 [.084/.099]*	.941	.933	.068
4	482.808 (167)*	.077 [.069/.085]*	.958	.952	.056
5†	417.060 (164)*	.070 [.062/.078]*	.966	.975	.051
4b	375.215 (149)*	.069 [.061/.078]*	.970	.962	.044
4h	482.807 (167)*	.077 [.069/.085]*	.958	.952	.056
		Study 2			
1	1499.209 (170)*	.154 [.147/.161]*	.811	.789	.114
2	630.323 (169)*	.091 [.083/.098]*	.934	.926	.069
3†	622.889 (167)*	.091 [.083/.098]*	.935	.926	.069
4	522.008 (167)*	.080 [.072/.088]*	.949	.943	.060
5†	475.258 (164)*	.076 [.068/.084]*	.956	.949	.057
4b	426.320 (149)*	.075 [.067/.083]*	.961	.950	.049
4h††	492.791 (166)*	.077 [.069/.085]*	.954	.947	.059

Note: df = degrees of freedom, RMSEA = Root Mean Square Error Approximation, CI = Confidence intervals, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual.

*Indicates $p < .05$. †Indicates models with latent variables correlating near 1. ††Indicates that the original model had presence of a negative residual variance. The modification identified in Study 1 (allowing parallel-worded items 7 & 8 to co-vary) is included in this model, which removed the presence of the negative residual variance. The original model had acceptable model fit (RMSEA = .080 [.072/.088], CFI = .949, TLI = .943, SRMR = .060).

Table 3

Standardized factor loadings, residual variances, and internal reliabilities from model 4's hierarchical and bifactor CFAs from Study 1 and 2.

Items/*Factors*	Hierarchical CFA				Bifactor CFA			
	Study 1		Study 2		Study 1		Study 2	
	λ	*ε*	λ	*ε*	λ_G (λ_s)	*ε*	λ_G (λ_s)	*ε*
Cognitive Empathy (Study 1: α = .92, ω = .95; Study 2: α = .92, ω = .94)	.51[a]	.74[a]	.47[a]	.78[a]				
1 – Just by seeing or hearing someone, I know if they are feeling sad	.81	.35	.82	.33	.32 (.76)	.33	.41 (.71)	.33
3 – Just by seeing or hearing someone, I know if they are feeling happy	.78	.39	.75	.44	.52 (.72)	.38	.41 (.63)	.44
5 – Just by seeing or hearing someone, I know if they are feeling angry	.79	.37	.73	.47	.33 (.76)	.34	.36 (.64)	.47
7 – Just by seeing or hearing someone, I know if they are feeling amused	.78	.39	.76	.42	.53 (.58)	.38	.42 (.64)	.41
9 – Just by seeing or hearing someone, I know if they are feeling scared	.84	.29	.77	.40	.41 (.74)	.29	.36 (.69)	.40
11 – Just by seeing or hearing someone, I know if they are feeling calm	.80	.37	.76	.42	.43 (.67)	.37	.38 (.66)	.42
13 – Just by seeing or hearing someone, I know if they are feeling disgusted	.80	.36	.80	.36	.36 (.72)	.35	.32 (.74)	.35
15 – Just by seeing or hearing someone, I know if they are feeling enthusiastic	.80	.36	.79	.37	.44 (.67)	.36	.43 (.66)	.37
17 – Just by seeing or hearing someone, I know if they are feeling embarrassed	.81	.35	.80	.36	.36 (.72)	.33	.30 (.76)	.33
19 – Just by seeing or hearing someone, I know if they are feeling proud	.77	.41	.80	.37	.50 (.58)	.41	.42 (.67)	.37

Negative Affective Empathy (Study 1: α = .74, ω = .85; Study 2: α = .70, ω = .74)	.63[a]	.60[a]	.62[a]	.61[a]				
2 – When I see or hear someone who is sad, it makes me feel sad too	.72	.49	.76	.43	.52 (.39)	.58	.57 (.26)	.61
6 – When I see or hear someone who is angry, it makes me feel angry too	.60	.64	.50	.76	.34 (.61)	.51	.21 (.53)	.67
10 – When I see or hear someone who is scared, it makes me feel scared too	.71	.50	.65	.58	.42 (.65)	.40	.34 (.60)	.53
14 – When I see or hear someone who is disgusted, it makes me feel disgusted too	.72	.49	.62	.62	.48 (.49)	.53	.29 (.62)	.53
18 – When I see or hear someone who is embarrassed, it makes me feel embarrassed too	.58	.67	.48	.77	.40 (.38)	.70	.14 (.64)	.57
Positive Affective Empathy (Study 1: α = .78, ω = .82; Study 2: α = .77, ω = .81)	.96[a]	.09[a]	1.00[a]	.00[a]				
4 – When I see or hear someone who is happy, it makes me feel happy too	.67	.56	.74	.45	.49 (.52)	.38	.81 (-.17)	.32
8 – When I see or hear someone who is amused, it makes me feel amused too	.73	.46	.66	.57	.77 (-.12)	.39	.64 (.24)	.54
12 – When I see or hear someone who is calm, it makes me feel calm too	.65	.58	.64	.59	.60 (.30)	.55	.64 (-.01)	.59
16 – When I see or hear someone who is enthusiastic, it makes me feel enthusiastic too	.71	.50	.70	.51	.64 (.47)	.37	.67 (.17)	.53
20 – When I see or hear someone who is proud, it makes me feel proud too	.67	.56	.67	.56	.66 (.07)	.56	.61 (.26)	.57
Composite Scores								
Affective Empathy (Study 1: α = .81, ω = .85, ω_H = .67; Study 2: α = .80, ω = .84, ω_H = .70)								
General Empathy (Study 1: α = .89, ω = .89, ω_H = .58; Study 2: α = .88, ω = .87, ω_H = .51)								

Note: First-order factor loadings from hierarchical model reflect loadings from Model 4. Study 2's hierarchical CFA includes allowing items 7 & 8 to co-vary (r = .45). [a]Factor loading of first-order factors on general empathy factor. λ = standardized factor loadings, ε = residual variance, λ_s = standardized factor loadings of specific factors, λ_G = standardized factor loadings of general empathy factor, α = Cronbach's alpha, ω = McDonald's Omega, ω_H = McDonald's Hierarchical Omega.

Table 4

Estimated factor inter-correlations of Model 4 in Studies 1 and 2.

Factor (F)	F1	F2	F3
F1 Cognitive Empathy	-	.32*	.49*
F2 Negative Affective Empathy	.29*	-	.61*
F3 Positive Affective Empathy	.48*	.62*	-

Note: Correlations above the diagonal are from Study 1, those below the diagonal are from Study 2.

*Indicates $p < .05$.

Table 5

Mean, standard deviation, and reliabilities for the PES, IRI, QCAE, and PAQ in Study 2.

	M (*SD*)	Reliabilities	
		α	ω (ω_H)
PES			
Cognitive Empathy	37.43 (7.16)	.92	.94
Negative Affective Empathy	12.38 (3.51)	.70	.74
Positive Affective Empathy	15.61 (4.00	.77	.81
Composites			
Affective Empathy	27.99 (6.38)	.80	.84 (.70)
Empathy	65.42 (11.21)	.88	.88 (.51)
IRI			
Perspective Taking	18.63 (4.54)	.78	.81
Empathic Concern	20.04 (4.48)	.78	.64
QCAE			
Cognitive Empathy	59.98 (8.35)	.91	.93† (.73†)
Affective Empathy	35.05 (5.86)	.83	.92† (.75†)
PAQ			
N-DIF	13.70 (5.80)	.87	.89
P-DIF	10.60 (4.72)	.86	.90
N-DDF	15.33 (6.28)	.90	.92
P-DDF	12.05 (5.30)	.89	.91
G-EOT	25.10 (9.71)	.89	.92
Composites			
Alexithymia	76.78 (25.00)	.94	.96 (.83)

Note: α = Cronbach's alpha, ω = McDonald's Omega, ω_H = Hierarchical Omega, PES = Perth Empathy Scale, IRI = Interpersonal Reactivity Index, QCAE = Questionnaire for Cognitive and Affective Empathy, PAQ = Perth Alexithymia Questionnaire, N-DIF = difficulty identifying negative feelings, P-DIF = difficulty identifying positive feelings, N-DDF = difficulty describing negative feelings, P-DDF = difficulty describing positive feelings, G-EOT = general externally oriented thinking.

†values obtained from poor fitting model (see supplementary material, Table S12).

Table 6

Model fit for the multi-group CFAs between males and females for model 4 and model 4b.

Level of Invariance	Chi-square (df)	RMSEA	CFI	TLI	ΔRMSEA	ΔCFI	ΔTLI
		Model 4					
Configural	948.492* (334)	.076*	.957	.952	-	-	-
Metric	905.850* (351)	.070*	.962	.958	-.006	.005	.006
Scalar	1016.746* (428)	.066*	.959	.964	-.004	-.003	.006
Freeing residuals	1026.405* (408)	.069*	.956	.960	.003	-.003	-.004
Variance/Covariance	764.432* (434)	.049	.977	.980	-.017	.018	.016
		Model 4b					
Configural	818.309* (298)	.074*	.964	.954	-	-	-
Metric	688.190* (334)	.058*	.975	.972	-.016	.011	.018
Scalar	814.774* (410)	.056	.972	.974	-.002	-.003	.002
Freeing residuals	822.202* (390)	.059*	.970	.971	.003	-.002	-.003
Variance/Covariance	659.602* (415)	.043	.983	.984	-.013	.011	.010

Note: df = degrees of freedom, RMSEA = Root mean squared error approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index. Δ = changes compared to the previous model, except for Variance/Covariance is compared to the Scalar model.

*Indicates $p < .05$.

Table 7

Pearson correlations for scales of the PES with the QCAE and the IRI

		QCAE		IRI	
		CE	AE	EC	PT
PES	CE	.62**	.19**	.32**	.36**
	NAE	.13*	.54**	.32**	.10
	PAE	.32**	.32**	.33**	.25**
	AE	.27**	.50**	.38**	.21**
	Empathy	.55**	.40**	.42**	.35**

Note: EC = empathic concern, PT = perspective-taking, CE = cognitive empathy, AE = affective empathy, NAE = negative affective empathy, PAE = positive affective empathy.

*indicates $p < .05$, **indicates $p < .001$.

Table 8

Factor loadings of the exploratory factor analysis with subscales from the PAQ and the PES.

		F1	F2
PAQ	N-DIF	.74	.15
	P-DIF	.75	-.04
	N-DDF	.73	.06
	P-DDF	.75	-.08
	EOT	.64	-.10
PES	CE	-.22	.45
	NAE	.21	.61
	PAE	-.07	.76
% variance accounted for		34.49	14.53

Note: F1 = factor one, F2 = factor two, PES = Perth Empathy Scale, PAQ = Perth Alexithymia Questionnaire, N-DIF = difficulty identifying negative feelings, P-DIF = difficulty identifying positive feelings, N-DDF = difficulty describing negative feelings, P-DDF = difficulty describing positive feelings, EOT = externally oriented thinking, CE = cognitive empathy, NAE = negative affective empathy, PAE = positive affective empathy.