Article

Development of an Empathy Rating Scale for Young Children

Assessment
1–14

DOI: 10.1177/10731911211038629
journals.sagepub.com/home/asm
SAGE

Eva R. Kimonis[1], Natasha Jain[1], Bryan Neo[1], Georgette E. Fleming[1], and Nancy Briggs[2]

Abstract

Empathy is critical to young children's socioemotional development and deficient levels characterize a severe and pervasive type of Conduct Disorder (i.e., with limited prosocial emotions). With the emergence of novel, targeted early interventions to treat this psychopathology, the critical limitations of existing parent-report empathy measures reveal their unsuitability for assessing empathy levels and outcomes in young children. The present study aimed to develop a reliable and comprehensive parent-rated empathy scale for young children. This was accomplished by first generating a large list of empathy items sourced from both preexisting empathy measures and from statements made by parents during a clinical interview about their young child's empathy. Second, this item set was refined using exploratory factor analysis of item scores from parents of children aged 2 to 8 years (56.6% male), recruited online using Amazon's Mechanical Turk. A five-factor solution provided the best fit to the data: Attention to Others' Emotions, Personal Distress (i.e., Emotional Contagion/Affective Empathy), Personal Distress–Fictional Characters, Prosocial Behavior, and Sympathy. Total and subscale scores on the new "Measure of Empathy in Early Childhood" (MEEC) were internally consistent. Finally, this five-factor structure was tested using confirmatory factor analysis and model fit was adequate. With further research into the validity of MEEC scores, this new rater-based empathy measure for young children may hold promise for assessing empathy in early childhood and advancing research into the origins of empathy and empathy-related disorders.

Keywords

empathy, assessment, prosocial behavior, children, scale development, parent-report

Empathy is a key feature of human responding, enabling the understanding and vicarious experience of others' feelings as if they were one's own (Clark, 1980). Individuals who fail to properly develop the ability to empathize with others during early childhood are at risk of various negative outcomes in later life. Empathy deficits are a core feature of a severe and persistent form of antisocial behavior called Conduct Disorder with limited prosocial emotions (American Psychiatric Association, 2013; Miller & Eisenberg, 1988). Novel early interventions are emerging to target the putative familial and dispositional mechanisms involved in this treatment resistant form of psychopathology (Dadds et al., 2019; Kimonis et al., 2019), but existing child empathy measures are proving unsuitable for measuring empathy levels and outcomes among young children. The most commonly used parent rating scales, which represent the dominant and arguably most clinically useful method for assessing empathy in young children, fail to reliably and comprehensively assess empathy subcomponents that are developmentally important indicators of empathy in young children (Dadds et al., 2008; Miller & Eisenberg, 1988; Rieffe et al., 2010). Consequently, the present study aimed to develop a new, comprehensive and reliable rating scale for measuring empathy in young children.

The Measurement of Child Empathy

To date, only two parent-report tools for measuring empathy in young children exist: The Griffith Empathy Measure (GEM; Dadds et al., 2008) and the Empathy Questionnaire (EmQue; Rieffe et al., 2010).

GEM. The 23-item GEM was adapted from the widely used self-report Bryant Empathy Index (BEI; Bryant, 1982) by rewording its items into third person. However, unlike the

[1]Parent-Child Research Clinic, University of New South Wales, Sydney, New South Wales, Australia
[2]Mark Wainwright Analytical Centre, University of New South Wales, Sydney, New South Wales, Australia

Corresponding Author:
Eva R. Kimonis, School of Psychology, The University of New South Wales, Mathews Building, Room 1102, Sydney, New South Wales 2052, Australia.
Email: e.kimonis@unsw.edu.au

unidimensional BEI, which was designed to assess only affective empathy (*experiencing* and *emotionally responding* to another's emotional state), the GEM is used to measure both affective and cognitive (*understanding* another's emotional state) empathy (Dadds et al., 2008). For example, a cognitive empathic response may involve recognizing why another individual is sad, whereas an affective empathic response would entail sharing in that sadness. This divergence highlights three key issues with the GEM.

First, the six-item GEM Cognitive Empathy subscale has been criticized for not reliably and comprehensively capturing "cognitive empathy" (Dadds, 2019; Murphy, 2019). In several studies including young children, the GEM Cognitive Empathy subscale demonstrated borderline acceptable internal consistency (Cronbach's $\alpha = .49$-$.62$; Dadds et al., 2008; Deschamps et al., 2014; Kimonis et al., 2016). This may reflect the general difficulty in assessing cognitive empathy in young children because of its delayed development relative to affective empathy (Hoffman, 1979). Alternatively, it may be due to the scale comprising several items that assess low frequency events (e.g., "My child doesn't understand why other people cry out of happiness") or lack face validity as indicators of cognitive empathy (e.g., "My child reacts badly when she or he sees people kiss and hug in public"). Other studies report poor construct validity of GEM Cognitive Empathy subscale scores (e.g., Dadds et al., 2009; Georgiou et al., 2019; Lui et al., 2016), and it has been suggested that this subscale instead taps into domains other than cognitive empathy, namely callousness (e.g., "When I get sad my child doesn't seem to notice"; Murphy, 2019; cf. Dadds, 2019). In support of this claim, Dadds et al. (2009) reported substantial GEM-rated cognitive empathy deficits in 3- to 13-year-old children with psychopathic traits, despite the general consensus that psychopathic traits are associated with impaired affective empathy, but intact cognitive empathy (e.g., Jones et al., 2010). The poor psychometric properties of the GEM Cognitive Empathy subscale possibly reflect that its item set was never intended to measure cognitive empathy (Murphy, 2019).

Second, the GEM does not have a "clean" factor structure in that a large number of its items substantially cross-load on both Cognitive and Affective Empathy factors. Of these, four items center on the child feeling upset in response to various situations, and thus appear to better tap into affective empathy. Eight GEM items used to calculate the total empathy score are not included in either Cognitive or Affective empathy subscales, of which several appear to measure sympathy (e.g., "It makes my child sad to see another child who can't find anyone to play with"; "My child feels sorry for another child who is upset"; "My child feels sad for other people who are physically disabled [e.g., in a wheelchair]"). In contrast to empathy, which involves experiencing the same emotional state as another person in a given situation, sympathy involves feeling sorrow or concern for another person. While both involve other-oriented vicarious emotional responses that are central to moral development, each is thought to play a unique role in motivating prosocial behavior and inhibiting aggressive behavior towards others (Eisenberg & Eggum, 2009; Eisenberg et al., 2013; Zahn-Waxler & Radke-Yarrow, 1990). Thus, any comprehensive assessment of childhood empathy should assess sympathy as distinct from other subcomponents.

Third, the GEM has insufficient item content coverage to adequately measure empathy in young children. Because the GEM comprises entirely reworded items from the BEI, which was developed as a self-report tool for children aged 6 years and older and so is not designed to capture earlier forms of empathy, it does not include any items tapping into prosocial action that is considered to be a developmentally significant empathic behavior in young children (Hoffman, 2001). Thus, using the GEM to measure the multidimensional empathy construct in young children (e.g., Pasalich et al., 2014) could potentially underestimate their empathic abilities.

EmQue. The 20-item EmQue is a parent-report tool specifically designed to measure three developmentally important dimensions of empathy in infants and toddlers: Personal Distress, Attention to Others' Feelings, and Prosocial Action. It was constructed to capture three of Hoffman's (1987) stages of empathic development: Global Empathy, Egocentric Empathy, and Empathy for Another's Feelings, and thus may be better suited than the GEM to measure empathy in young children. However, some methodological issues with this scale's construction may limit its ability to do so. The EmQue item list was generated from a factor analysis of $n = 83$ parent ratings about their young child ($M_{age} = 35$ months, $SD = 13$) on 60 empathy items generated by a team of developmental psychologists, teachers, and a child psychiatrist (Rieffe et al., 2010). While there is no agreed on minimum sample size for factor analysis, larger sample sizes are favorable (Yong & Pearce, 2013). When using small samples, the factor analytic solution often does not generalize to the overall population, compounded by a high rate of missing values in Rieffe et al.'s (2010) study that led to half of EmQue items being deleted—items potentially important to operationalizing empathy in young children. For example, EmQue items reflect developmentally advanced prosocial responses (e.g., "When two children are quarrelling, my child tries to stop them"), overlooking key behavioral indicators of empathy in very young children and again risking underestimation of empathic abilities.

In terms of its specific subscales, the EmQue is unique in its inclusion of prosocial behavior—a construct that may capture early childhood manifestations of empathy

(Hoffman, 1987). It is therefore unfortunate that items within this scale lack face validity as indicators of empathy and may instead tap into a fear of punishment or behavioral regulation (e.g., "When I make clear that I want some peace and quiet, my child tries not to bother me"). Furthermore, the Emotion Contagion (i.e., Personal Distress) subscale demonstrated poor psychometric properties, with internal consistency estimates below the acceptable minimum value (Cronbach's $\alpha = .58$; Rieffe et al., 2010). Emotion Contagion scores also correlated negatively with a measure of consoling (prosocial) behaviors (Rieffe et al., 2010), such that the vicarious experience of another's distress decreased, rather than increased, helping behaviors toward them. Although this finding was based on only a small subsample of 12- to 30-month-old children ($n = 31$) and could possibly indicate that they had not yet developed the ability to act prosocially, this is difficult to establish due to the limited use and psychometric testing of the EmQue to date.

The Development of Empathy

Understanding how empathy develops can provide valuable insights into how to best measure this construct at different developmental stages, and in the present case, early childhood. According to Hoffman (2000), empathy develops through five distinct stages: reactive newborn cry, egocentric empathic distress, quasi-egocentric empathic distress, veridical empathic distress, and empathy for another's experience beyond just their current situation.

The first stage of empathy development captures the innate aspects of empathy, whereby newborns and infants experience automatic contagious crying when exposed to others' negative emotional states, as they are unable to distinguish between their own and others' distress cues (Dondi et al., 1999; McDonald & Messinger, 2011). By 12 months, when children can better distinguish between their own and others' emotional responses and show a rudimentary understanding of others' emotions, they enter the egocentric empathic distress stage (Hoffman, 2000). This new awareness of emotions in others as separate from self leads children to allocate attentional resources toward others in distress, which Rieffe et al. (2010) refer to as "Attention to Others" in their EmQue model. In the egocentric empathic distress stage, toddlers are increasingly able to respond to others' displays of distress with appropriate empathic affect, but tend to behave egocentrically and engage in self-comforting behaviors because it is overwhelming and personally distressing to them (Eisenberg et al., 2010; McDonald & Messinger, 2011). In the following few months, children advance to the quasi-egocentric empathic distress stage, during which they understand that they need to comfort others in distress but do this using prosocial behaviors that are more appropriate for comforting themselves (e.g., giving a doll to a distressed parent; Hoffman, 1979).

The 24-month age marks the most significant gains in empathic abilities. It is between 24 and 36 months old that the child enters the refined stage of empathic responding, termed veridical empathy for another's distress (Hoffman, 2000). By this stage, the child understands that others have their own inner states, and is more receptive to others' needs and feelings, as distinct from their own (Eisenberg et al., 2015). The child also better understands what others are experiencing (i.e., has some rudimentary cognitive empathy; Zahn-Waxler & Radke-Yarrow, 1990). Thus, the empathic responses observed after 24 months are other-oriented and involve prosocial behaviors that are specific to another's situation, state, and needs (e.g., giving a distressed child their preferred toy; Eisenberg et al., 2015; Zahn-Waxler & Radke-Yarrow, 1990). Other more sophisticated empathic responses also emerge at the 24-month mark, including empathizing and sympathizing with a wider range of emotions, advanced verbal responses, and the expression of facial concern (McDonald & Messinger, 2011).

The final stage of empathy involves more advanced cognitive skills, including perspective-taking (Eisenberg et al., 2015) and an understanding that both the self and others have identities, histories and feelings (Hoffman, 1979). For example, children understand that despite appearing happy in the present, an individual may be experiencing hardships elsewhere (e.g., illness, homelessness, poverty; Hoffman, 2000). During this stage children develop the ability to experience empathic responses even when the other person is not physically present. Research indicates that this final stage begins developing before age 7 years but is not solidified until later childhood (Gnepp & Gould, 1985).

In sum, Hoffman's developmental theory explains that major empathic gains occur in children between 2 and 8 years old. While the empathic responses observed during this developmental period are mostly affective, children also show rudimentary aspects of cognitive empathy. Children's empathic and sympathetic experiences begin to motivate helping and comforting prosocial behaviors toward individuals in distress, which are important components of the construct of empathy in early childhood (Zahn-Waxler et al., 1992; Zahn-Waxler & Radke-Yarrow, 1990).

An understanding of this normative developmental process has guided theoretical accounts of the etiology of Conduct Disorder with limited prosocial emotions, with deficits in empathic concern for others implicated as a defining characteristic of this developmental pathway (Frick & Kemp, 2021; Waller et al., 2020). Specifically, emotional processing deficits present during early development—for example, poor accuracy in recognizing facial, vocal, and postural distress cues; deficient attentional orienting to distress cues—are thought to give rise to low aversive arousal in response to others' distress (i.e., affective empathy). In turn, these empathy deficits result in reduced motivation to engage in prosocial behavior and reduced inhibition against

(proactive) aggression, and are thus a critical causal mechanism explaining the severe, aggressive, and stable conduct problems associated with Conduct Disorder with limited prosocial emotions (see Frick et al., 2014).

Numerous empirical studies have provided support for the role of empathy deficits in the development of limited prosocial emotions (e.g., Bedford et al., 2017; Wagner et al, 2020) and several novel interventions have been developed to target empathy deficits as a putative risk and maintaining factor for Conduct Disorder with limited prosocial emotions. For example, Dadds, Cauchi, Wimalaweera, Hawes, and Brennan (2012) found that emotion recognition training combined with parent management training improved both mother-rated affective empathy and conduct problems among children (M_{age} = 10.38 years, SD = 2.41) with conduct problems and limited prosocial emotions, although improvement in affective empathy did not explain improvement in conduct problems. Given the young age at which empathy development begins, Kimonis et al. (2019) developed an adapted version of parent–child interaction therapy to target empathy deficits and other causal mechanisms unique to Conduct Disorder with limited prosocial emotions in young children. This targeted intervention improved parent-rated affective and cognitive empathy among 3- to 6-year-old children (M_{age} = 4.5 years, SD = .92) with conduct problems and limited prosocial emotions, although internal consistency of GEM cognitive empathy scale scores was poor and reliable change indices from pretreatment to posttreatment were lowest for GEM scores relative to all other treatment outcomes. While this finding may reflect that the intervention did not change empathy to the same extent as other emotional and behavioral outcomes measured, it is also possible that the GEM was unable to assess changes in developmentally important indicators of empathy in young children, such as prosocial behavior and sympathy.

The Present Study

With the emergence of these novel interventions targeting mechanisms thought to partly contribute to the profound empathic deficits of children with limited prosocial emotions-type conduct disorders, there has emerged a need to reliably and comprehensively assess empathy in young children. Current rating tools for measuring empathy during this developmental stage are limited and inconsistent with leading theoretical models of empathic development. Thus, the present study aimed to develop a new tool, the Measure of Empathy in Early Childhood (MEEC), to measure empathy in young children aged between 2 and 8 years. This age range was selected because the construct of empathy only properly develops in children at 2 years old, after which significant empathic gains occur, until approximately 8 years old (Hoffman, 2000). The empathic behaviors observed in children during this developmental period are basic and thus manifest differently to those seen in older children. Therefore, the items on the new scale need to reflect the specific empathic behaviors observed in younger children.

The MEEC was developed by first sourcing items from both preexisting empathy measures and from statements made by parents about their young child's empathy during a semistructured clinical interview (Clinical Assessment of Prosocial Emotions [CAPE]; Frick, 2013). Consistent with the empathy literature and prior factor analytic studies, it was hypothesized that factor analysis of the empathy item set would identify four factors indicative of empathy-related behaviors evident in the early childhood years: personal distress (i.e., emotion contagion/affective empathy), attention to others' feelings, prosocial behavior, and sympathy. It was expected that scores on the MEEC and its subscales would be internally consistent. It was hypothesized that total MEEC empathy scores would be higher for girls than boys given consistent findings across the literature that females are more empathic than males when using rater-based measures (e.g., Hastings et al., 2000; Eisenberg & Fabes, 1998). Consistent with developmental theory (Hoffman, 2000), it was hypothesized that older child age would be associated with higher prosocial behavior, sympathy, and total empathy scores and younger age associated with greater personal distress scores.

Method

Participants

Participants were 1,226 parents of children (694 male, 532 female) aged 2 to 8 years (M = 4.94 years, SD =1.87), recruited online using Amazon's Mechanical Turk (MTurk; http://www.mturk.com). MTurk provides access to high-quality data at a low cost (Paolacci & Chandler, 2014) from workers that are reliable (Goodman et al., 2013) and attentive (Hauser & Schwarz, 2016). The sample was predominantly Caucasian (70.2%), but also included African American (9.6%), Asian (10.6%), Hispanic/Latino (4.9%), Native American (2.7%), Pacific Islander (0.3%, n = 4), Middle Eastern (0.1%, n = 1), and mixed/other ethnicity (1.6%) families. All participants reported English as their first language as this was an inclusion criterion. The majority of parents were between 25 to 34 years old (56.4%) and 35 to 44 years old (32.2%). Most reported being married or in a domestic partnership (62.6%) with an average of 2.1 (SD = 1.1; Range: 1-8) children.

To participate in the study, participants were required to provide informed consent and meet strict eligibility criteria. These criteria included (a) being a parent of a child aged between 2 and 8 years, (b) speaking English as their primary language, and (c) having an MTurk approval rating of 95% or above (Peer et al., 2014). Participants were

excluded from the study if they produced low-quality data, as indicated by (a) attempting the survey multiple times to try to meet the eligibility criteria, (b) failing the attention check, (c) completing the survey in an insufficient amount of time (<7 minutes), or (d) providing false demographic information. To ensure recruitment of parents of 2- to 8-year-old children and not individuals only seeking remuneration, participants were not told that they needed to be parents to participate. Participants that had multiple children in the target age range were advised to choose one child and respond only about that child. Participants were reimbursed $2.00, consistent with Buhrmester et al. (2011). Previous findings suggest that higher reimbursement rates increase the rate of recruitment (Berinsky et al., 2012), but have little influence on the quality of the data (Buhrmester et al., 2011).

Measures

Empathy. The MEEC was developed by obtaining items from two sources. First, a content analysis of 73 video recorded CAPE semistructured interviews was conducted. These interviews were previously conducted with parents of 2- to 8-year-old children as part of a comprehensive assessment, to determine their eligibility for a clinical trial of a parent training intervention for children with conduct problems and limited prosocial emotions (Kimonis et al., 2019). The CAPE is a clinical tool for measuring callous–unemotional (CU) traits in individuals aged 3 to 21 years. The CAPE semistructured interview permits an in-depth understanding of the individual's empathic responding, being a core component of CU traits. The present study focused on only three CAPE semi structured interview questions (item numbers 3, 5, and 9), which tap into concern for others, prosocial behaviors, and affective empathy, respectively. All statements made by parents during the interviews that were relevant to empathic responding were paraphrased and adapted into third-person format for inclusion in the new empathy scale item pool.

Additional empathy items (n = 188) were sourced from nine preexisting empathy measures: the GEM, EmQue, Feeling and Thinking Scale (Garton & Gringart, 2005), Adolescent Measure of Empathy and Sympathy (Vossen et al., 2015), Basic Empathy Scale (Jolliffe & Farrington, 2006), Empathy Quotient (Baron-Cohen & Wheelwright, 2004), interpersonal reactivity index (Davis, 1980), Questionnaire of Cognitive and Affective Empathy (Reniers et al., 2011), and the Affective and Cognitive Measure of Empathy (Vachon & Lynam, 2016). These tools were chosen as they conceptualize empathy as a multidimensional construct, including at least an affective and cognitive empathy component. Thus, the resulting scale would allow for a broad measure of overall empathy, and also specific measures of empathy subcomponents via subscale scores. Apart from the GEM and EmQue, all items from the remaining scales were reworded into third person. Any adverbs of frequency (e.g., "often," "usually") were removed to avoid confusion when the response scale is added. Duplicate items between scales were removed (n = 21) and the remaining items were retained in the draft empathy item pool (n = 287).

The draft item pool was then refined based on consultation with a developmental psychologist with expertise in empathy research, to ensure that the items were appropriate for measuring empathy in young children. Items were excluded if they (a) were redundant with other items (n = 35), (b) were not developmentally appropriate (n = 66, e.g., "My child can usually predict how someone will feel"), (c) represented low frequency events (n = 16, e.g., "My child tends to lose control during emergencies"), (d) measured a construct other than empathy (n = 52), (e) comprised two distinct clauses (these n = 3 items were split into two separate items), or (f) obviously lacked face validity (n = 1). One other item was formed by combining two separate CAPE statements, as they considerably overlapped. The remaining items (n = 119) formed the first iteration of the MEEC. The refined item list was randomly reordered and given a 4-point response scale from 1 (*strongly disagree*) to 4 (*strongly agree*). This 4-point scale adopted from the Empathy Quotient and Questionnaire of Cognitive and Affective Empathy was used, since it demonstrated the best fit across all empathy items. Other agreement-based scales also fit (e.g., 5-point Basic Empathy Scale), but were avoided due to the mixed findings around the efficacy of rating scales with middle points—research shows that participants may select the middle rating when they are unsure of which response option to choose (Armstrong, 1987; Nadler et al., 2015). Participants rated the extent to which they agreed with each statement about their child (e.g., "My child comforts me when I am upset"), such that higher scores indicated higher empathy levels.

Attention Check. An attention check question ("Select *slightly agree* if you are paying attention.") was embedded within the empathy questionnaire to ensure that participants were reading the items. This question was scored as either correct or incorrect. Individuals who failed to select *slightly agree* were excluded (n = 147).

Social Desirability. The Balanced Inventory of Desirable Responding, Short Form (BIDR-16; Hart et al., 2015) is a 16-item scale for measuring socially desirable responding. Participants indicated how true various statements (e.g., "I never regret my decisions.") were on a 7-point scale from 1 (*not true*) to 7 (*very true*). Scores on the BIDR-16 correlated highly with the full-length BIDR (r = .84-.87) and showed acceptable reliability (Cronbach's α = .63-.82; Hart et al., 2015). A measure of social desirability was included in the

present study since empathy is a socially desirable construct and parents may thus misrepresent their child's empathy levels (Jolliffe & Farrington, 2006).

Procedure

Ethical approval for the current study was obtained from the University of New South Wales Human Research Ethics committee. The 119 empathy items were administered to participants on MTurk via a Qualtrics (https://www.qualtrics.com) survey, where they provided informed consent and completed demographic questions (n= 6) to assess their eligibility. Those participants who met eligibility criteria completed the empathy items about their 2- to 8-year-old child. Participants then completed the BIDR-16 and were told to respond about themselves. Data were checked for validity and valid responders were reimbursed. To check whether responses reflected socially desirable responding, polychoric correlations were calculated between the 119 empathy items and the BIDR-16 score.

Next, an exploratory factor analysis was conducted in SAS Version 9.4 software with a randomly selected half of the sample (n = 596) to examine the factor structure of empathy items. Factor analysis was conducted on polychoric correlations rather than Pearson's correlations, since the latter assumes a continuous (interval/ratio) level of measurement and the former is recommended when item-level data are scored on less than a 5-point scale (Horn, 1965). Parallel analysis (with the 50th percentile used for the random data eigenvalues, since a general factor is expected in these data) was used to determine the number of factors to extract (O'Connor, 2000). This method is considered superior to the Eigenvalue above 1 method (i.e., the Kaiser-Guttman rule), which results in retention of excessive factors[1] (Velicer & Jackson, 1990), and the relatively unsophisticated "scree test" method (Cattell, 1966). The data were then forced to load on the number of factors suggested using the parallel analysis approach and an oblique rotation was used, since some correlation was expected between the scale factors (Costello & Osborne, 2005). Items with a communality $<$0.30 were removed from the analysis (n = 5 for five-factor, n = 8 for four-factor solution), as such items are generally unrelated to other items (Costello & Osborne, 2005).

Once a "clean" factor analytic output was obtained (i.e., at least 3 high-loading items per factor and no cross-loading items; Costello & Osborne, 2005), each factor was inspected to identify the underlying dimension and labelled accordingly. Internal consistency reliability was estimated with Ordinal alpha for both the subdimension (subscale) scores and the overall (total) scores (Bonanomi et al., 2015). In order to condense the number of items to be comparable to other parent-rating empathy measures for young children, alphas with item removed were examined. A confirmatory factor analysis was run in *M*plus Version 8.2 (Muthén & Muthén, 2001) on the remaining confirmatory subsample (n = 596), and measurement invariance across sex was tested. Raw data were analyzed using weighted least square mean and variance adjusted (WLSMV) estimation, and all variables were specified as ordinal categorical. Latent factor variances were constrained to 1.0, and all factors were allowed to correlate with all other factors.

Sex differences in empathy levels were tested using independent samples t tests. Associations between empathy total and subscale scores and child age were tested using Pearson zero-order correlations.

Results

Preliminary Analyses

No empathy items correlated with the BIDR-16 at a value greater than .30 and thus, all items were retained (Lawrence et al., 2004). The Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) value was greater than the minimum recommended value of .65 (KMO = .977), thus verifying the sampling adequacy of the data. Additionally, the Barlett's Test of Sphericity (Bartlett, 1954) was significant, $\chi^2(7021) = 85269.587$, $p < .001$, indicating that the data were suitable for factor analysis. Furthermore, the sample size met the minimum requirements (>300 participants, 5-10 cases per item) for factor analysis (Yong & Pearce, 2013).

Exploratory Factor Analysis

Parallel analysis indicated that the number of factors to retain was four or five, whereas the "scree test" method suggested four factors. Rotated factor solutions for four- and five-factor models were examined, and the five-factor solution was determined to be most meaningful. The five-factor solution explained 95% of the total variance, with the five factors explaining 50.7%, 23.6%, 9.4%, 6.1%, and 5.1%, respectively. Factor 1 involved compassionate, comforting, and caring behaviors, and was labelled "Prosocial Behavior." Factor 2 captured sensitivity and responsiveness to others' feelings and was labelled "Attention to Others' Emotions" to align with the EmQue. Factor 3 reflected the vicarious experience of others' emotions and was labelled "Personal Distress (Emotional Contagion/Affective Empathy)." Factor 4 reflected personal distress to the experiences of fictitious television/movie and book characters and was labelled "Personal Distress–Fictional Characters." Factor 5 involved sorrow for others' misfortune and was labelled "Sympathy."

Internal Consistency

The internal consistencies of scores on all factors were high (Ordinal $\alpha > .88$). Since examination of alphas with items

removed was unhelpful for informing which items to eliminate, in order to reduce the large number of items loading onto factors 1 to 3, groups of items with factor loadings <.60, <.70, and <.75 (Factor 1 only) were iteratively removed. One redundant item was also removed from Factor 3. This method resulted in a nine-item Prosocial Behavior, eight-item Attention to Others' Emotions, and eight-item Personal Distress subscale. Due to their brevity, no items were removed from the six-item Personal Distress–Fictional Characters and four-item Sympathy subscales. As shown in Table 1, internal consistency was high for all subscales (Ordinal $\alpha > .89$). All items correlated with their corresponding scale at a value greater than .30. The internal consistency of scores on the 35-item MEEC empathy total score was good (Ordinal $\alpha = .89$). Table 1 presents factor loadings and correlations between factors for the final 35-item MEEC.[2]

Confirmatory Factor Analysis

Figure 1 shows the path diagram for the CFA model and the resulting coefficient estimates. Model fit statistics showed adequate model fit ($\chi^2 = 2202.54$, $df = 550$). Root mean square error of approximation (RMSEA) showed acceptable fit (.071) with 90% CI [.068, .074]. The comparative fit index (CFI) was .918, indicating adequate fit. The standardized root mean square residual (SRMR) was 0.072, indicating good fit. The Tucker–Lewis index (TLI = 0.912) was lower than the recommended value for good fit of 0.95. However, most fit indices showed adequate fit for this model (Hu & Bentler, 1998).[3]

Sex Differences and Age Analyses

Girls (G) had higher total empathy, $t(1224) = -4.49$, $p < .001$, G: $M = 105.69$, $SD = 15.24$ versus boys (B): $M = 101.68$, $SD = 15.71$, and higher Prosocial Behavior, $t(1224) = -5.49$, $p < .001$, G: $M = 31.44$, $SD = 4.73$ versus B: $M = 29.86$, $SD = 5.21$, Attention to Others' Emotions, $t(1205) = -5.05$, $p < .001$, G: $M = 26.22$, $SD = 4.57$ versus B: $M = 24.80$, $SD = 5.25$, and Personal Distress–Fictional Characters, $t(1224) = -2.11$, $p = .035$, G: $M = 16.71$, $SD = 4.21$ versus B: $M = 16.19$, $SD = 4.39$, scores than boys. Older child age was associated with greater total empathy ($r = .09$, $p = .001$), Prosocial Behavior ($r = .10$, $p < .001$), and Sympathy ($r = .32$, $p < .001$) scores, but not other empathy dimensions.

Discussion

The goal of the present study was to develop a parent-report tool for measuring empathy in young children between the ages of 2 and 8 years, when important empathic gains occur and motivate prosocial behaviors (Hoffman, 2000). Exploratory and confirmatory factor analyses were used to refine and determine the factor structure of the new MEEC.

Factor Structure of the MEEC

Exploratory factor analysis of empathy items identified five factors: Prosocial Behavior (i.e., engagement in compassionate, comforting, and caring behaviors); Attention to Others' Emotions (i.e., sensitivity and responsiveness to others' emotional displays); Personal Distress (i.e., emotional contagion/affective empathy involving vicarious experience of others' emotions); Personal Distress–Fictional Characters; and Sympathy (i.e., sorrow for others' misfortune). These findings inform our conceptualization and measurement of empathy in young children in five key ways.

First, given that no adult measure of empathy incorporates a prosocial subscale, the present findings support that prosocial behaviors are an important indicator of empathy for young children. Indeed, Zahn-Waxler and Radke-Yarrow (1990) suggest that definitions of empathy for young children should incorporate prosociality. Consequently, the MEEC may be a more suitable measure of empathy in young children than the GEM that does not include prosocial items. MEEC prosocial items are all face-valid and developmentally appropriate for measuring prosociality in young children, improving upon the EmQue Prosocial Actions subscale that includes more developmentally advanced prosocial behaviors and items lacking face validity.

Importantly, cognitive empathy did not emerge as a factor in analyses. This finding aligns with research indicating that this final stage of empathic development that involves more advanced cognitive skills does not develop until later childhood (Gnepp & Gould, 1985). Alternatively, young children may be capable of experiencing rudimentary cognitive empathy but lack the metacognitive and/or verbal abilities to report on these internal experiences to parents. Consequently, using parent-report of young children's understanding of others' emotional states, as in the case of the GEM, is likely to be problematic and possibly explains the poor psychometric properties of its cognitive empathy scale when used with samples including young children (Deschamps et al., 2014; Kimonis et al., 2016). On the other hand, parent ratings of young children's displays of prosocial behavior might better tap into nascent cognitive empathic abilities since understanding that another is experiencing distress is a prerequisite to engaging in prosocial caring and comforting behaviors towards them.

Second, the identification of a Personal Distress (Emotion Contagion/Affective Empathy) factor reiterates that the vicarious experience of others' emotive states and specifically, personal distress to another's negative emotional state, is a key indicator of empathy in young children.

Table 1. Final Factor Loadings, Alphas, and Correlations From Exploratory Factor Analysis of the Measure of Empathy in Early Childhood (MEEC).

Variable	Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Source
emp1_17	My child comforts me when I am upset	**0.875**	−0.063	0.017	−0.036	−0.145	CAPE
emp3_21	When I am hurt, my child comes over and checks on me	**0.848**	0.077	0.055	−0.035	−0.033	CAPE
emp2_5	My child tries to make people feel better when he or she realizes they are upset	**0.818**	0.020	0.106	0.046	−0.059	CAPE
emp2_21	My child does things to cheer me up, such as picking me flowers, drawing a picture or getting something for me	**0.801**	0.044	−0.001	−0.042	−0.016	CAPE
emp3_24	When I am upset or tired, my child does nice things for me (e.g. getting tissues)	**0.792**	0.103	−0.023	−0.060	0.058	CAPE
emp3_19	When I am hurt, my child reassures me that everything will be okay	**0.765**	0.024	0.026	0.000	0.098	CAPE
emp1_24	My child comforts children who are distressed	**0.749**	0.066	−0.036	0.110	0.011	CAPE
emp2_11	When someone is upset, my child shares their toys with them	**0.747**	−0.127	−0.116	0.078	0.125	CAPE
emp2_14	My child tries to comfort a friend when they are upset	**0.723**	0.119	−0.018	0.040	0.101	CAPE
emp5_20R	My child notices but does not show concern when someone is hurt/injured[a]	0.108	**0.818**	−0.003	−0.028	0.037	CAPE
emp5_2R	My child does not notice when someone else is crying[a]	0.054	**0.807**	−0.066	0.048	−0.135	CAPE
emp5_14R	My child ignores situations where someone is badly injured or upset[a]	0.123	**0.784**	−0.076	−0.055	−0.031	CAPE
emp4_10R	If someone gets hurt, my child is not emotionally affected[a]	−0.065	**0.783**	0.044	0.047	0.076	CAPE
emp3_15R	When another child cries, my child looks away[a]	−0.068	**0.781**	−0.103	0.059	−0.251	CAPE
emp3_17R	When someone gets hurt, my child may look up, but will not be concerned[a]	0.084	**0.777**	−0.043	0.001	0.091	CAPE
emp3_14R	Other people's misfortunes do not disturb my child a great deal[a]	−0.057	**0.720**	0.071	0.039	0.177	IRI
emp2_18R	My child does not change his/her behavior or emotions in response to someone being upset[a]	0.020	**0.709**	0.147	−0.065	0.007	CAPE
emp5_11	When others around my child are nervous or worried, he or she gets a bit scared and worried too	0.060	0.033	**0.836**	−0.039	−0.036	F&T
emp4_6	My child becomes nervous when other children around him or her are nervous	−0.028	0.007	**0.826**	−0.018	−0.048	GEM
emp2_15	My child becomes scared when he or she sees someone who is scared	0.102	−0.010	**0.786**	0.049	−0.097	AMES
emp3_9	When another child gets frightened, my child freezes or starts to cry	−0.066	−0.213	**0.738**	0.077	−0.057	EmQue
emp4_12	It worries my child when others are worrying and panicky	0.095	0.165	**0.654**	−0.011	0.050	QCAE
emp5_13	My child becomes sad when other children around him/her are sad	0.016	0.100	**0.631**	0.065	0.209	GEM
emp2_2	My child also needs to be comforted when another child is in pain	−0.058	−0.052	**0.622**	0.065	0.141	EmQue
emp1_9	When another child is upset, my child needs to be comforted too	−0.081	−0.022	**0.576**	0.144	0.044	EmQue
emp4_20	My child is upset when something bad happens to a character in a movie	0.016	0.040	−0.051	**0.908**	−0.026	CAPE
emp5_19	My child gets deeply involved with the feelings of a character in a film, play, or novel	−0.011	−0.085	0.085	**0.814**	0.052	QCAE
emp4_14	My child is emotionally affected by book characters that are in trouble	−0.031	0.066	0.087	**0.741**	0.011	CAPE
emp4_18	Sad movies or TV shows make my child sad	−0.057	0.096	0.104	**0.735**	0.019	GEM
emp2_1	My child gets upset/worried when characters in a book are lost, in danger or hurt	0.189	−0.098	0.022	**0.654**	0.046	CAPE
emp1_5	My child gets upset if movie/TV show characters are upset	0.005	0.024	0.149	**0.636**	−0.039	CAPE
emp5_21	My child gets emotional when he/she sees someone less fortunate than them, such as a homeless person	−0.059	−0.014	−0.014	0.013	**0.972**	CAPE

(continued)

Table 1. (continued)

Variable	Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Source
emp5_1	Seeing posters of homeless or sick children upsets my child	−0.020	−0.047	0.039	0.039	**0.766**	CAPE
emp4_15	My child feels worried about people that are not as lucky as them	0.130	0.022	0.054	0.022	**0.700**	F&T
emp4_1	My child feels sad for other people who are physically disabled (e.g. in a wheelchair)	0.120	−0.011	0.008	0.009	**0.691**	GEM
Factor correlations							
Factor 1		—					
Factor 2		.543	—				
Factor 3		.317	.135	—			
Factor 4		.316	.190	.575	—		
Factor 5		.432	.176	.528	.526	—	
Alpha values							
Cronbach's alpha		.948	.947	.910	.914	.891	
Ordinal alpha		.948	.947	.910	.915	.892	

Note. Bolded factor loadings are test items best explained by each factor. Factor 1 = Prosocial Behavior; Factor 2 = Attention to Others' Emotions; Factor 3 = Personal Distress (Emotional Contagion/Affective Empathy); Factor 4 = Personal Distress–Fictional Characters; Factor 5 = Sympathy; CAPE = Clinical Assessment of Prosocial Emotions; GEM = Griffith Empathy Measure; IRI = Interpersonal Reactivity Index; EmQue = Empathy Questionnaire; F&T = Feeling and Thinking Scale; AMES = Adolescent Measure of Empathy and Sympathy; OCAE = Questionnaire of Cognitive and Affective Empathy.
[a]Reverse-scored items.

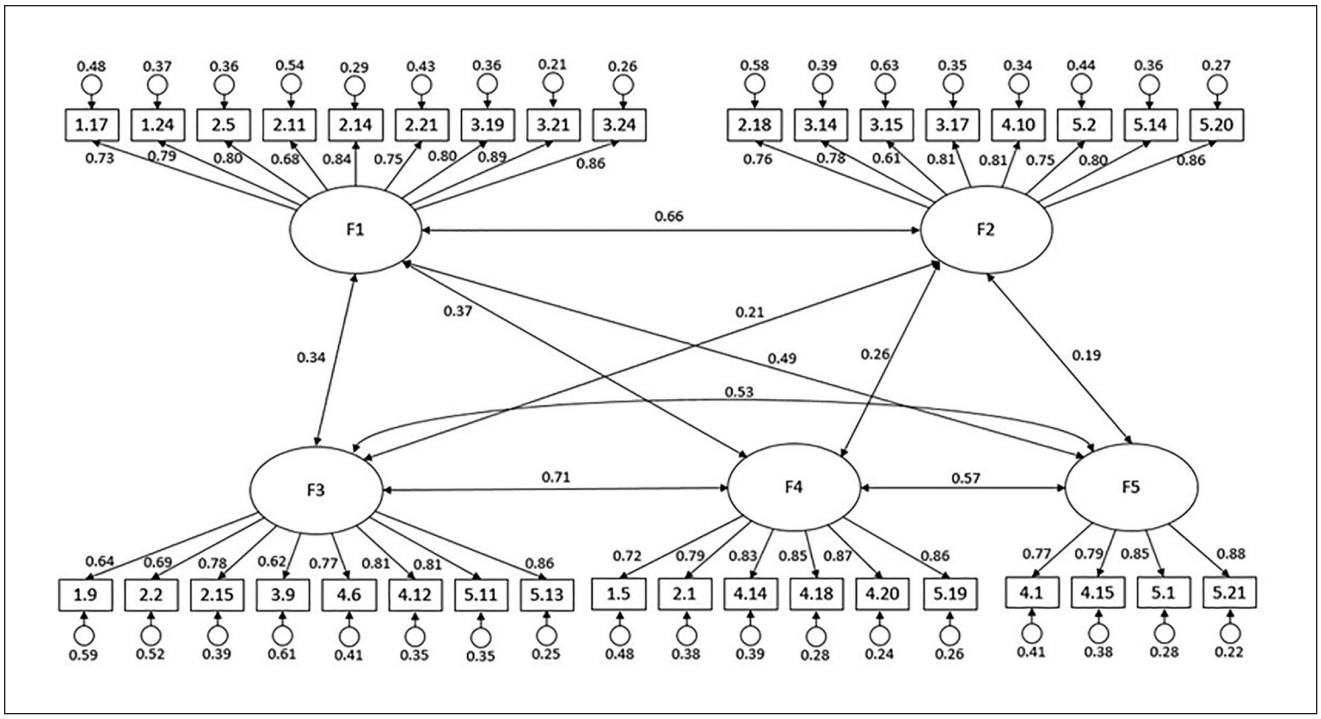

Figure 1. Path diagram of the confirmatory factor analysis model and resulting coefficient estimates.
Note. F1 = Prosocial Behavior; F2 = Attention to Others' Emotions; F3 = Personal Distress (Emotional Contagion/Affective Empathy); F4 = Personal Distress–Fictional Characters; F5 = Sympathy.

This finding is consistent with Hoffman's (2000) theory of empathy development, and the idea that affective empathy develops early-on in childhood. Improving on the EmQue's Emotion Contagion subscale that demonstrated poor psychometric properties, internal consistency of the MEEC Personal Distress scale scores was good.

A novel contribution of the MEEC beyond existing rater-based measures of childhood empathy is its subscale

focusing on personal distress to the emotional experiences of fictional characters. The identification of this factor suggests that exposure to television/movie and book characters plays a key role in young children's empathic development. According to Feshbach and Feshbach (1997), developmentally appropriate television fosters empathic development by invoking emotion states in children and teaching them empathic skills, such as prosocial responding and conflict resolution, via social learning. Research also shows that exposure to visual media improves children's ability to identify different emotional states and teaches social-emotional skills, such as emotional understanding and helping behaviors (Calvert & Kotler, 2003; Peebles et al., 2018). Exposure to media may be especially important for young children who are not yet of school-age, as they may infrequently encounter real-life empathy-eliciting emotional displays.

Third, the Attention to Others' Emotions subscale captures the shift towards attending and responding to others' displays of emotion that is seen in children aged 12 months and older (Hoffman, 2000), conceptually overlapping but not sharing items with the EmQue's Attention to Others' Feelings subscale (Rieffe et al., 2010). All but one item loading on this factor were sourced from parents' responses to questions about their child with conduct problems on the empathy section of CAPE interviews—a clinical tool designed explicitly to assess limited prosocial emotions—and appeared to partly capture indifference to others' distress before reverse coding (i.e., ignoring and/or failing to respond to someone who is hurt or upset). Our approach of using parental statements to create empathy items is novel and the emergence of this factor highlights how attention to others' emotions is a subcomponent of empathy that is relevant to both adaptive and maladaptive ends of the empathy spectrum.

Scores on the Attention to Others' Emotions factor showed the weakest correlations with other MEEC factors, suggesting it taps into a separable but related construct to empathy. Consistent with this possibility, an EEG/ERP study found that children showed an early N200 component indexing attentional orienting to emotionally salient stimuli as well as a late-positive potential reflecting more complex cognitive processing, while passively viewing stimuli depicting others being physically injured (Cheng et al., 2014). Future research can investigate whether the Attention to Others' Emotions subscale is more predictive of disorders involving empathic deficits than other empathy subcomponents, having implications for developmental theory. Indeed, individuals with psychopathic traits show deficient emotional attention to others' distress that is evident from early childhood (Kimonis et al., 2016; 2020). Alternatively, this subscale may capture the tendency for younger children to become overwhelmed by others' displays of distress and to avoid them (Hoffman, 2000), although it was uncorrelated with age. While distress avoidant behaviors are typically reported in 12–24-month-olds, some residual aspects of this behavior remain in older children.

Fourth, a Sympathy dimension emerged from factor analytic results, consistent with findings that children as young as 2 years old are capable of demonstrating sympathy towards another in distress (Vaish, Carpenter, & Tomasello, 2009; Zahn-Waxler, Radke-Yarrow, & King 1979). Sympathy scores increased with older child age and cognitive sophistication in the current study, as did Prosocial Behavior and total MEEC scores, consistent with prior research (Kienbaum, 2014; Zahn-Waxler et al., 1992). Together, these findings indicate that sympathy is important to the conceptualization of empathy in early childhood. Although sympathy is often conflated with empathy, its lack of differentiation from other empathy subcomponents such as personal distress is thought to have contributed to inconsistent findings across the empathy literature (Eisenberg et al., 2013). For example, the GEM includes several items tapping into sympathy, which cross-load fairly equally on both affective and cognitive empathy factors in factor analytic studies and are used to calculate total empathy scores (Dadds et al., 2008). The BEI (Bryant, 1982), from which the GEM was developed, as well as its precursor the Mehrabian and Epstein (1972) measure of emotional empathy, similarly tap into sympathetic feelings toward others. Thus, an advantage of the MEEC is its yielding of both a "cleaner" and more comprehensive factor structure relative to other parent-report empathy measures. Future longitudinal research might use the MEEC to examine whether these distinct but related empathy subcomponents differentially predict socioemotional outcomes in later development, which can inform theoretical models. For example, individual differences in the tendency to experience sympathy over personal distress in the presence of others' distress is associated with greater emotional regulation (Eisenberg et al., 1994; Spinrad & Stifter, 2006).

Finally, with regard to sex differences, girls showed significantly higher overall scores on the MEEC than boys, as well as greater Prosocial Behavior, Attention to Others' Emotions, and Personal Distress (Fictional Characters). Boys and girls did not differ in their levels of Personal Distress (not involving fictional characters) or Sympathy for others. According to Jolliffe and Farrington (2006), such sex differences are important for validating empathy measures, as females consistently show greater empathy than males across studies using rater-based measures (Eisenberg & Fabes, 1998). Notably, prior studies of the EmQue failed to detect sex differences, suggesting this tool may require refinement.

Together, these preliminary results suggest that the MEEC should be further investigated as a measure of empathy in young children, notwithstanding some study

limitations. First, we did not examine the construct validity of MEEC scores by examining their convergence with other measures of empathy and key criterion measures (e.g., social competence). It is critical for future research to test whether MEEC scores correlate with alternate observational, rater, laboratory, and physiological measures of empathy. Ultimately, using a multimethod approach is likely to best capture young children's comprehensive empathic abilities (Eisenberg & Strayer, 1987). Second, our partial sourcing of items from existing parent-rating empathy scales leaves open the possibility that this new tool shares similar limitations, although the majority of MEEC items (22 out of 35) were generated from parental statements made during a clinical interview about their child's empathy. Finally, while our sample size was sufficient to address our study aims, it was insufficient to test measurement invariance across sex. Thus, it will be important to replicate and extend findings using a larger sample in future research.

Children who fail to feel appropriate levels of empathy for others are at risk of engaging in antisocial patterns of behavior, endangering both society and themselves (Miller & Eisenberg, 1988). Thus, it is vital that empathy deficits be detected early when they are most amenable to intervention. Although parent-report tools for measuring empathy in young children exist, these tools have some serious flaws, necessitating their refinement and/or development of new tools as endeavored in this study. The present study developed a new parent-report tool for measuring empathy in young children. When administered to the parents of 2- to 8-year-old children, the MEEC demonstrated a unique five-factor structure that captured empathy-related behaviors common to the early childhood period (Personal Distress, Personal Distress–Fictional Characters, Attention to Others' Emotions, Prosocial Behavior, and Sympathy), which has important implications for future conceptualizations and developmental models of empathy in young children. Findings provided preliminary support for the tool's psychometric properties, but further validation studies are needed including the sensitivity of MEEC scores to assess changes in response to intervention. Collectively, the results suggest that with further research the "MEEC" has potential as a new rater-based tool for detecting multidimensional empathic deficits in young children.

Declaration of Conflicting Interests

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

Funding

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

ORCID iDs

Eva R. Kimonis https://orcid.org/0000-0002-4325-6108

Natasha Jain https://orcid.org/0000-0002-9673-8805

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In the present study, 10 factors yielded an eigenvalue above 1.
2. The final 35-item MEEC is available in the online supplemental material.
3. Measurement invariance across sex was not possible in the confirmatory sample. When separated into boys and girls, sample sizes were 330 (sex = male) and 266 (sex = female). These sample sizes combined with the skewed ordinal item-level data resulted in nonadmissible solutions in a multiple group analysis (factor correlations greater than 1, negative item error variances, etc.). Bandalos (2014) suggests that a sample size of 200 to 500 should be sufficient for estimation of latent variable models with ordinal outcomes; this was not the case for these data.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Armstrong, R. L. (1987). Midpoint 5 on a five-point Likert-type scale. *Perceptual and Motor Skills*, *64*(2), 359-362. https://doi.org/10.2466/pms.1987.64.2.359

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(1), 102-116, https://doi.org/10.1080/10705511.2014.859510

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), 163-175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society*, *16*(2), 296-298.

Bedford, R., Wagner, N. J., Rehder, P. D., Propper, C., Willoughby, M. T., & Mills-Koonce, R. W. (2017). The role of infants' mother-directed gaze, maternal sensitivity, and emotion recognition in childhood callous unemotional behaviours. *European Child & Adolescent Psychiatry*, *26*(8), 947-956. https://doi.org/10.1007/s00787-017-0967-1

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351-368. https://doi.org/10.1093/pan/mpr057

Bonanomi, A., Cantaluppi, G., Ruscone, M. N., & Osmetti, S. A. (2015). A new estimator of Zumbo's ordinal alpha: A copula approach. *Quality & Quantity*, *49*(3), 941-953. https://doi.org/10.1007/s11135-014-0114-8

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413-425. https://doi.org/10.1111/j.1467-8624.1982.tb01331.x

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5. https://doi.org/10.1177/1745691610393980

Calvert, S. L., & Kotler, J. A. (2003). Lessons from children's television: The impact of the Children's Television Act on children's learning. *Journal of Applied Developmental Psychology*, *24*(3), 275-335. https://doi.org/10.1016/S0193-3973(03)00060-1

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Cheng, Y., Chen, C., & Decety, J. (2014). An EEG/ERP investigation of the development of empathy in early and middle childhood. *Developmental Cognitive Neuroscience*, *10*(October), 160-169. https://doi.org/10.1016/j.dcn.2014.08.012

Clark, K. B. (1980). Empathy: A neglected topic in psychological research. *American Psychologist*, *35*(2), 187-190. https://doi.org/10.1037/0003-066X.35.2.187

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1-9. https://doi.org/10.7275/jyj1-4868

Dadds, M. R., Cauchi, A. J., Wimalaweera, S., Hawes, D. J., & Brennan, J. (2012). Outcomes, moderators, and mediators of empathic-emotion recognition training for complex conduct problems in childhood. *Psychiatry Research*, *199*(3), 201-207.

Dadds, M. R. (2019). In search of better measurement of empathy in children: A reply to Murphy (2018) regarding the Griffith Empathy Measure. *Australian Psychologist*, *54*(3), 165-166. https://doi.org/10.1111/ap.12357

Dadds, M. R., English, T., Wimalaweera, S., Schollar-Root, O., & Hawes, D. J. (2019). Can reciprocated parent–child eye gaze and emotional engagement enhance treatment for children with conduct problems and callous-unemotional traits: A proof-of-concept trial. *Journal of Child Psychology and Psychiatry*, *60*(6), 676-685. https://doi.org/10.1111/jcpp.13023

Dadds, M. R., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Hunter, K., & Merz, S. (2009). Learning to "talk the talk": The relationship of psychopathic traits to deficits in empathy across childhood. *Journal of Child Psychology and Psychiatry*, *50*(5), 599-606. https://doi.org/10.1111/j.1469-7610.2008.02058.x

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Merz, S., & El Masry, Y. (2008). A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry Human Development*, *39*(2), 111-122. https://doi.org/10.1007/s10578-007-0075-4

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85. https://www.uv.es/~friasnav/Davis_1980.pdf

Deschamps, P. K. H., Been, M., & Matthys, W. (2014). Empathy and empathy induced prosocial behavior in 6- and 7-year-olds with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(7), 1749-1758. https://doi.org/10.1007/s10803-014-2048-3

Dondi, M., Simion, F., & Caltran, G. (1999). Can newborns discriminate between their own cry and the cry of another newborn infant? *Developmental Psychology*, *35*(2), 418-426. https://doi.org/10.1037/0012-1649.35.2.418

Eisenberg, N., & Strayer, J. (1987). Critical issues in the study of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 3-13). Cambridge University Press.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, C, & Suh, K. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of Personality and Social Psychology*, *66*(4), 776-797.

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71-83). MIT Press.

Eisenberg, N., Eggum-Wilkens, N. D., & Spinrad, T. L. (2015). The development of prosocial behavior. In D. A. Schroeder & W. G. Graziano (Eds.), *The oxford handbook of prosocial behavior* (pp. 114-136). Oxford University Press.

Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 701-778). John Wiley.

Eisenberg, N., Spinrad, T. L., & Morris, A. (2013). Empathy-related responding in children. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 184-207). Psychology Press.

Feshbach, N. D., & Feshbach, S. (1997) Children's empathy and the media: Realizing the potential of television. In S. Kirschner & D. A. Kirschner (Eds.), *Perspectives on psychology and the media* (pp. 3-27). American Psychological Association. https://doi.org/10.1037/10509-001

Frick, P. J. (2013). *Clinical Assessment of Prosocial Emotions (CAPE)*. https://sites01.lsu.edu/faculty/pfricklab/wp-content/uploads/sites/100/2015/11/CAPE-Manual.pdf

Frick, P. J., & Kemp, E. C. (2021). Conduct Disorders and Empathy Development. *Annual Review of Clinical Psychology*, *17*, 391-416. https://doi.org/10.1146/annurev-clinpsy-081219-105809

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Annual research review: A developmental psychopathology approach to understanding callous-unemotional traits in children and adolescents with serious conduct problems. *Journal of Child Psychology & Psychiatry*, *55*(6), 532-548. https://doi.org/10.1111/jcpp.12152

Garton, A. F., & Gringart, E. (2005). The development of a scale to measure empathy in 8- and 9-year old children. *Australian Journal of Education and Developmental Psychology*, *5*, 17-25. https://www.newcastle.edu.au/__data/assets/pdf_file/0017/100394/v5-garton-gringart.pdf

Georgiou, G., Kimonis, E. R., & Fanti, K. A. (2019). What do others feel? Cognitive empathy deficits explain the association between callous-unemotional traits and conduct problems among preschool children. *European Journal of Developmental Psychology*, *16*(6), 633-653. https://doi.org/10.1080/17405629.2018.1478810

Gnepp, J., & Gould, M. E. (1985). The development of personalized inferences: Understanding other people's emotional reactions in light of their prior experiences. *Child Development*, *56*(6), 1455-1464. https://doi.org/10.2307/1130465

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213-224. https://doi.org/10.1002/bdm.1753

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, *5*(4), 1-9. https://doi.org/10.1177/2158244015621113

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental Psychology*, *36*(5), 531-546. https://doi.org/10.1037/0012-1649.36.5.531

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavioural Research Methods*, *48*(1), 400-407. https://doi.org/10.3758/s13428-015-0578-z

Hoffman, M. L. (1979). Development of moral thought, feeling, and behavior. *American Psychologist*, *34*(10), 958-966. https://doi.org/10.1037/0003-066X.34.10.958

Hoffman, M. L. (1987). The contribution of empathy to justice and moral judgment. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 47-80). Cambridge University Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Hoffman, M. L. (2001). Toward a comprehensive theory of prosocial moral development. In D. Stipek & A. Bohart (Eds.), *Constructive and destructive behavior* (pp. 61-86). American Psychological Association. https://doi.org/10.1037/10433-003

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *32*(2), 179-185. https://doi.org/10.1007/BF02289447

Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453. https://doi.org/10.1037/1082-989X.3.4.424

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

Jones, A. P., Happé, F. G. E., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: Different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *51*(11), 1188-1197. https://doi.org/10.1111/j.1469-7610.2010.02280.x

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36. https://doi.org/10.1007/BF02291575

Kienbaum, J. (2014). The development of sympathy from 5 to 7 years: Increase, decline, or stability? A longitudinal study. *Frontiers in Psychology*, *5*(468), 1-10. https://doi.org/10.3389/fpsyg.2014.00468

Kimonis, E. R., Fanti, K. A., Anastassiou-Hadjicharalambous, X., Mertan, B., Goulter, N., & Katsimicha, E. (2016). Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*, *44*(4), 625-638. https://doi.org/10.1007/s10802-015-0075-y

Kimonis, E. R., Fleming, G., Briggs, N., Brouwer-French, L., Frick, P. J., Hawes, D. J., Bagner, D. M., Thomas, R., & Dadds, M. R. (2019). Parent-child interaction therapy adapted for preschoolers with callous-unemotional traits: An open trial pilot study. *Journal of Clinical Child & Adolescent Psychology*, *48*(1), S347-S361. https://doi.org/10.1080/15374416.2018.1479966

Kimonis, E. R., Kidd, J., Most, S. B., Krynen, A., & Liu, C. (2020). An elusive deficit: Psychopathic personality traits and aberrant attention to emotional stimuli. *Emotion*, *20*(6), 951-964. https://doi.org/10.1037/emo0000601

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), 911-920. https://doi.org/10.1017/S0033291703001624

Lui, J. H. L., Barry, C. T., & Sacco, D. F. (2016). Callous-unemotional traits and empathy deficits: Mediating effects of affective perspective-taking and facial emotion recognition. *Cognition and Emotion*, *30*(6), 1049-1062. https://doi.org/10.1080/02699931.2015.1047327

McDonald, N. M., & Messinger, D. S. (2011). The development of empathy: How, when, and why. In J. J. Sanguineti, A. Acerbi, & J. A. Lombo (Eds.), *Moral behavior and free will: A neurobiological and philosophical approach* (pp. 333-359). If Press.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, *40*(4), 525-543. https://doi.org/10.1111/j.1467-6494.1972.tb00078.x

Miller, P. A., & Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological Bulletin*, *103*(3), 324-344. https://doi.org/10.1037/0033-2909.103.3.324

Murphy, B. A. (2019). The Griffith Empathy Measure does not validly distinguish between cognitive and affective empathy in children. *Australian Psychologist*, *54*(3), 159-164. https://doi.org/10.1111/ap.12336

Muthén, L. K., & Muthén, B. O. (2001). M*plus* (Version 8.2) [Computer software]. Muthén and Muthén. https://www.statmodel.com/index.shtml

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, *142*(2), 71-89. https://doi.org/10.1080/00221309.2014.994590

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402. https://doi.org/10.3758/bf03200807

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184-188. https://doi.org/10.1177/0963721414531598

Pasalich, D. S., Dadds, M. R., & Hawes, D. J. (2014). Cognitive and affective empathy in children with conduct problems: Additive and interactive effects of callous-unemotional traits and autism spectrum disorders symptoms. *Psychiatry*

Research, *219*(3), 625-630. https://doi.org/10.1016/j.psychres.2014.06.025

Peebles, A., Bonus, J. A., & Mares, M.-L. (2018). Questions + answers + agency: Interactive touchscreens and children's learning from a socio-emotional TV story. *Computers in Human Behavior*, *85*, 339-348. https://doi.org/10.1016/j.chb.2018.03.039

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

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

Rieffe, C., Ketelaar, L., & Wiefferink, C. H. (2010). Assessing empathy in young children: Construction and validation of an Empathy Questionnaire (EmQue). *Personality and Individual Differences*, *49*(5), 362-367. https://doi.org/10.1016/j.paid.2010.03.046

Spinrad, T. L., & Stifter, C. A. (2006). Toddlers' empathy-related responding to distress: Predictions from negative emotionality and maternal behavior in infancy. *Infancy*, *10*(2), 97-121. https://doi.org/10.1207/s15327078in1002

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

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, *45*(2), 534-543.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor-analysis: Some further observations. *Multivariate Behavioral Research*, *25*(1), 97-114. https://doi.org/10.1207/s15327906mbr2501_12

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

Wagner, N. J., Waller, R., Flom, M., Ronfard, S., Fenstermacher, S., & Saudino, K. (2020). Less imitation of arbitrary actions is a specific developmental precursor to callous–unemotional traits in early childhood. *Journal of Child Psychology and Psychiatry*, *61*(7), 818-825. https://doi.org/10.1111/jcpp.13182

Waller, R., Wagner, N. J., Barstead, M. G., Subar, A., Petersen, J. L., Hyde, J. S., & Hyde, L. W. (2020). A meta-analysis of the associations between callous-unemotional traits and empathy, prosociality, and guilt. *Clinical Psychology Review*, *75*(February), Article 101809. https://doi.org/10.1016/j.cpr.2019.101809

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94. https://doi.org/10.20982/tqmp.09.2.p079

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and Emotion*, *14*(2), 107-130. https://doi.org/10.1007/BF00991639

Zahn-Waxler, C., Radke-Yarrow, M., & King, R. A. (1979). Child rearing and children's prosocial initiations toward victims of distress. *Child Development*, 319-330.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*(1), 126-136. https://doi.org/10.1037/0012-1649.28.1.126

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology*, *28*(6), 1038-1047. https://doi.org/10.1037/0012-1649.28.6.1038