

This document is the author deposited version.

You are advised to consult the publisher's version if you wish to cite from this version. Published article: Takamatsu, R., Tsou, Y-T., Kusumi, T., & Rieffe, C. (2021). The Japanese empathy questionnaire (EmQue) for preschool children: Psychometric properties and measurement invariance across gender. *International Journal of Behavioral Development*. Advanced online publication. <https://doi.org/10.1177/01650254211022852>

The Japanese Empathy Questionnaire (EmQue) for Preschool Children:  
Psychometric properties and Measurement Invariance Across Gender

Reina Takamatsu

Kyoto University

Yung-Ting Tsou

Leiden University

Takashi Kusumi

Kyoto University

Carolien Rieffe

Leiden University

Author note

Reina Takamatsu, Graduate School of Education, Kyoto University, Kyoto, Japan

Yung-Ting Tsou, Unit of Developmental and Educational Psychology, Institute of Psychology, Leiden University, Leiden, The Netherlands

Takashi Kusumi, Graduate School of Education, Kyoto University, Kyoto, Japan

Carolien Rieffe, Unit of Developmental and Educational Psychology, Institute of

Psychology, Leiden University, Leiden, The Netherlands; Department of Psychology and Human Development, Institute of Education, University College London, London, UK;  
Department of Human Media Interaction, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, The Netherlands

Correspondence concerning this article should be addressed to Reina Takamatsu, Graduate School of Education, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan. Email: [takamatsu.reina.4c@kyoto-u.ac.jp](mailto:takamatsu.reina.4c@kyoto-u.ac.jp)

Declarations of interest: none

### **Acknowledgments**

Funding: This study was supported in part by research grants from the start-up program for young scholars at the Kyoto University

### Abstract

Empathy is assumed to be a universal human motivation to act altruistically toward others. Developmental models of empathy explaining when and how children acquire the capacity to empathize have been proposed. However, the existing knowledge is largely built upon studies conducted in the Western context. To fill this gap, a cross-culturally validated measure of empathy for children is needed. The purpose of this study was to assess the Japanese version of the Empathy Questionnaire (EmQue), a parent-reported measure of empathy in preschool children, including its construct validity, measurement invariance across genders, and reliability. A total of 550 children aged 1–6 years participated in this study ( $M_{age} = 4.17$  years,  $SD = 1.21$ ). Their mothers completed the Japanese EmQue. Confirmatory factor analysis confirmed the hypothesized three-factor structure (emotional contagion, attention to others' feelings, and prosocial actions) in the 13-item Japanese EmQue. The internal consistencies of the three scales were high. Measurement invariance across gender groups was also supported. Overall, the results demonstrate that the Japanese EmQue is a reliable and valid measure of the empathy of Japanese preschool children. It can serve as a tool in future studies to elucidate the role of culture in shaping empathy in early childhood.

*Keywords:* empathy, preschool children, parent-report questionnaire, confirmatory factor analysis, measurement invariance

## **The Japanese Empathy Questionnaire (EmQue) for Preschool Children: Psychometric Properties and Measurement Invariance Across Gender**

### **Introduction**

Empathy is essential for making emotional connections with others. From infancy to late adulthood, empathy affects many aspects of our lives, including enriching interpersonal experiences and strengthening group cohesion (de Waal, 2009). It has been a challenge for psychologists to operationalize this multifactorial construct and to create an assessment tool that reliably measures its main aspects. In developmental and educational psychology, studies on the development of empathy have relied mainly on observations, peer nominations, children's responses to distressed others, and parental reports of empathy-related behaviors (e.g., Spinrad & Stifter, 2000). The utilization of a convenient questionnaire has been in high demand. In this study, we assessed the Japanese version of the Empathy Questionnaire (EmQue), a parent-report questionnaire developed by Rieffe et al. (2010) for measuring empathy in early childhood.

### **The Empathy Questionnaire (EmQue)**

The EmQue (Rieffe et al., 2010) is the first validated parental rating measure that encompasses three dimensions of empathy in preschool children: emotional contagion, attention to others' feelings, and prosocial actions. The questionnaire was built upon Hoffman's (1987) four developmental stages of empathy in childhood. According to Hoffman's model, empathy is observed as an emotional arousal during infancy (Stage 1). During this period, a caregiver may see one infant crying and then another infant spontaneously starting to cry. The emotional state of the infant (observer) is the direct result of perceiving the crying child's emotional state because at this stage, infants do not see the self as separate from others and experience the other's emotions as their own. Emotional contagion is an effortless process, and infants have little control over their

reactions because of poor emotional regulation skills. The arousal triggered by perceiving others' emotions could be overwhelming because infants experience the same level of distress. At this stage, empathy is directed toward the self, and infants have little or no understanding of the causes and consequences of others' distress (Hoffman, 1987; Zahn-Waxler & Radke-Yarrow, 1990).

After the first year of life (Stage 2), children become better equipped with emotional regulation strategies and undergo a gradual separation-individuation process through which they identify their own and others' feelings separately. In this stage, children become less affected by emotional contagion and start to pay attention to others' feelings (Hoffman, 1987). When another person is in a strong positive or negative emotional state, the child stops playing with their toys and watches attentively. However, the child has no resources to reduce the distress of others yet. In parallel with motor developments and with increasing awareness of others' emotions, prosocial actions emerge around the age of two years (Stage 3; Zahn-Waxler et al., 1992). Children start to provide some form of help to distressed others. In the last stage (Stage 4), the child foresees the long-term impact of their prosocial acts and shows empathic concern for distant others (Hoffman, 1987; Rieffe et al., 2010; Wellman et al., 2001).

The EmQue is designed to measure the first three developmental milestones of empathy in Hoffman's (1987) model that emerge before or during preschool years: emotional contagion, attention to others' feelings, and prosocial actions. The original EmQue has 20 items (Rieffe et al., 2010), and a 13-item short form in which the three-factor structure remained was developed at a later stage (Grazzani et al., 2016; Lucas-Molina et al., 2018; see Table 1). In fact, the short form achieved a better model fit than the original 20-item version in Italian and Spanish samples (Grazzani et al., 2016; Lucas-Molina et al., 2018). Across samples, the assumed sequence of the first three stages of empathy was reflected by the EmQue. The levels of prosocial actions positively and most strongly correlated with age, while the levels of emotional contagion and attention to

others' feelings were either unrelated or less strongly correlated with age in preschool children (Grazzani et al., 2016; Lucas-Molina et al., 2018; Rieffe et al. 2010).

The EmQue has not been validated in the Japanese context, nor is there any other validated parent-reported measure of early childhood empathy available in Japan. Furthermore, research on the development of empathy in non-Western cultures is limited. Cross-cultural findings on empathy in adulthood have suggested that empathy may be expressed or construed differently in East Asian context (Atkins et al., 2016; Cassels et al., 2010; Chopik et al., 2017). Compared to those who have been socialized in a Western context, East Asians tend to show an inward focus on the distress triggered by others and to experience higher levels of personal emotional arousal when witnessing others in distress (Atkins et al., 2016). One possible explanation for the differences in empathic responses is the differences in cultural values. While Western cultures are more strongly individualistic-oriented and emphasize individual uniqueness and responsibilities, in East Asian collectivistic-oriented cultures, the self is strongly tied to their social group (Markus & Kitayama, 1991). Thus, East Asians typically may try not to share negative emotions to avoid upsetting other people and hampering group harmony (Matsumoto et al., 2008). As a result of infrequent negative emotion expression, seeing other people's distress can lead to higher personal arousal (Trommsdorff, 1995). Despite the accumulating evidence that has shown how culture influences empathic responses, how such variations emerge in empathic maturation is still unknown. To conduct empirical studies to ascertain the development of empathy among East Asian children and cultural variations, a cross-culturally validated measurement scale is needed.

### **Present Study**

The aim of this present study was to assess the Japanese version of the EmQue for measuring empathy in preschool children in Japan and to understand whether the three-factor

structure of empathy (emotional contagion, attention to others' feelings, and prosocial actions) is also present in a Japanese sample. To achieve this aim, we first evaluated the construct validity by testing the hypothesized three-factor structure. We assessed the structure using the original 20-item model and the short-form model with 13 items to examine which model best fits the Japanese data. In addition, we assessed the measurement invariance across genders to verify whether the mothers of boys and girls provide equivalent responses to the items. Second, we evaluated the internal consistency of the three scales in the EmQue. Third, we examined the correlations of the three scales with children's age. Given the sequential model proposed by Hoffman (1987), we expected that age would be correlated most strongly with the levels of prosocial actions and least strongly with the levels of emotional contagion.

## Methods

### Participants and Procedures

A total of 550 children (273 boys and 279 girls) aged 1–6 years ( $M_{age} = 4.17$  years,  $SD = 1.21$ ) participated in this study. Their mothers ( $n = 500$ ;  $M_{age} = 34.16$ ,  $SD = 4.41$ , range = 22–48 years) took part in an online survey in which they were asked to respond to the 20-item EmQue after providing demographic information. Among the responding mothers, 482 (96.4%) were married, and 18 (3.6%) were single or divorced. The average age of the mothers and the proportion of mothers who were single or divorced were comparable to the national figures (Ministry of Health, Labor and Welfare, 2017). If a mother indicated that she had two or more children in the studied age range, she was asked to fill out the EmQue for each child.

The recruitment and fieldwork were administered by Cross Marketing Inc., a market research company in Tokyo, Japan. A survey invitation email was sent to the mothers who registered themselves on an existing online panel managed by the company and met the inclusion criteria (over 20 years old and with at least one preschool aged child). The survey programmer

ensured that the participants were from every geographic area of Japan by monitoring responses and controlling the survey distribution based on the respondents' preregistered demographic information. The geographic distribution of respondents was comparable to the population density in Japan (Hokkaido: 5.0%, Tohoku: 5.1%, Kanto: 37.2%, Chubu: 18.5%, Kansai/Kinki: 16.7%, Chugoku: 6.3%, Shikoku: 1.7%, Kyusyu including Okinawa: 9.6%).

After giving their informed consent to participate, the mothers filled out the Japanese EmQue on the online portal operated by the company. To maintain the anonymity, privacy, and confidentiality of the process (Mustanski, 2001), we assured the mothers that their responses would be anonymous. In addition, the introduction screen also explained to the mothers that the data would be used only for research and analyzed only by the principal investigators. Ethical approval for this research was obtained from the Research Ethics Committee of Kokoro Research Center, Kyoto University prior to data collection (ID: 1-P-19).

## **Materials**

The original EmQue (Rieffe et al., 2010) contains 20 items categorized into three dimensions of empathy in preschool children: emotional contagion, attention to others' feelings, and prosocial actions. Mothers rated their child's behaviors over the previous two months on a 3-point scale (0 = *no*, 1 = *sometimes*, and 2 = *often*). Its short-form version includes 13 items with the same tridimensional structure (Grazzani et al., 2016; Lucas-Molina et al., 2018; see Table 1).

The EmQue was translated into Japanese using the back-translation method (Brislin, 1986) and professional translators. We then adjusted the items using a committee approach (Harkness, 2003). A team of researchers worked with the translators, and any inconsistencies between the original EmQue and the back-translation were resolved by discussions. Instead of the original 3-point rating scale (Rieffe et al., 2010), a 5-point scale (1 = *never* to 5 = *always*) was used for the



Japanese EmQue to increase the reliability, internal consistency, and discriminative power (Preston & Colman, 2000).

### **Statistical Analysis**

We split the data into two subsets, the training set ( $n = 284$ ) and the testing set ( $n = 266$ ), by generating random samples of approximately 50% of the cases. The two sets of data did not differ in the age of the children,  $t(548) = 1.61, p = .10$ ; or in the gender distribution,  $\chi^2(1, N = 550) = 1.04, p = .31$ .

We started by conducting confirmatory factor analysis (CFA) on the training set to examine the model fit indices of the 20-item and 13-item models. To correct the skewness, the bootstrapping method was used (Ishiguro et al., 1997). For model standardization, the marker variable method was used, where the first variable of each latent variable was selected as the marker and fixed to 1. Given that having fewer items per scale automatically improves model fit (Little et al., 2002), we applied the parceling approach to the training data set in addition to the item-level analyses to compare the 20-item and 13-item EmQue scales with the same degree of freedom (Little, 2013; Kishton & Widaman, 1994). Three parcels per scale were created using the single-factor method. In this method, highest loaded items were matched with lowest loaded items to achieve an item-to-construct balance (Little et al., 2002; Matsunaga, 2008).

Next, we performed CFA on the testing data set at the item level to validate the factor structure of the better-fitting model. This cross-validation method has been recommended for rigorously testing a hypothetical model (Arlot & Celisse, 2010).

Afterwards, we combined the subsamples to perform multigroup CFA and evaluated whether the measurement properties were invariant across gender. Following the standard procedure (Brown, 2014; Little, 2013), we tested five levels of measurement invariance sequentially: configural, metric, scalar, structural (homogeneity of variances and covariances), and

latent means. In the first step, we tested the configural invariance in which the model structure was examined in both gender groups without constraints. Configural invariance indicates that the overall structure of the measurement model is equivalent across gender. In the next step, all the factor loadings on the latent construct were constrained to be equal across gender (metric invariance). If the assumption of metric (weak) invariance is met, the measurement scale has the same operational definition across groups. In the third step, we tested the scalar (strong) invariance by constraining the item intercepts to be equal across gender. For gender, scalar invariance indicates that boys and girls with the same actual level of empathy would be rated similarly by their mothers.

In the fourth and fifth steps, we tested homogeneity of variances, covariances, and latent means across the groups. To test homogeneity of variances and covariances, factor variances and covariances were constrained to be equal. If the assumption of homogeneity is met, the range of scores on a latent construct and the relations among the constructs are similar across gender groups. Lastly, we tested homogeneity of means. To estimate the difference between group means, the female group was selected as the reference, while the means of the male group were freely estimated. If the mean-level test is passed, the latent construct means are considered equal across the groups. The chi-square difference test was used to assess the tenability of homogeneity of variances, covariances, and latent means across gender.

To interpret the model fits, a set of absolute and relative goodness of fit indices were used. A model was preferred when the normed chi square (NC) < 3.0 (Bollen, 1989), the comparative fit index (CFI) > .90 (an acceptable fit) or > .95 (a good fit; Hu & Bentler, 1999; MacCallum et al., 1996), and the root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) < .08 (Cheung & Rensvold, 2002; Little, 2013; MacCallum, et al., 1996). Because the chi squared ( $\chi^2$ ) statistics are sensitive to minor deviations from the conceptual model and the sample size,  $\chi^2$  tests may be statistically significant even when the model fits well. The NC

is a relative  $\chi^2$  value computed by dividing the  $\chi^2$  index by the degrees of freedom, which reflects overall discrepancy between the observed data and the fitted covariance matrices and is less sensitive to sample size than the absolute  $\chi^2$  (Schweizer, 2010).

To test the weak and strong measurement invariance hypotheses, changes in the CFI was used. The hypothesis of equivalence between the gender groups was accepted when a decrease in CFI was  $< .01$  (Cheung & Rensvold, 2002). To test the homogeneity of variances, covariances, and means, the chi-square difference test was performed.

To estimate the internal consistency, Cronbach's alpha coefficients, McDonald's omega, and interitem correlations were used. We computed alphas for comparison with other language versions of the EmQue. A Cronbach's alpha coefficient of .70 or higher is considered acceptable (Ponterotto & Ruckdeschel, 2007). McDonald's omega ( $\omega$ ) was computed as a model-based reliability estimate for the EmQue scales. In multidimensional models, McDonald's omega provides a more accurate estimate of reliability than Cronbach's alpha (Dunn et al., 2014; Flora, 2020). An omega coefficient of .70 or higher was considered adequate. The average interitem correlations should be between .15 and .50 (Clark & Watson, 1995). Finally, the correlations between age (in years) and the EmQue scores were examined by including age as a continuous variable in the CFA model.

SPSS AMOS version 26 (SPSS Inc., Chicago, IL, USA) was used to test the factor structure and measurement invariance. The other analyses were conducted using SPSS version 25 (SPSS Inc., Chicago, IL, USA) and R version 4.0.2 (R Core Team, 2020).

## **Results**

### ***Factor Structure***

CFA was performed to test the three-factor structure of the original model (20 items) and the short-form model (13 items), respectively, using the training data set. At the item level, the

original model showed a mediocre fit ( $\chi^2(167, N = 284) = 480.72, p < .001$ ; NC = 4.21, CFI = .86; RMSEA = .08; and SRMR = .07), suggesting that the model may be further improved. The 13-item model indicated a good fit ( $\chi^2(62, N = 284) = 135.81, p < .001$ ; NC = 2.19; CFI = .94; RMSEA = .07; and SRMR = .05). In the additional CFA using parcels, both the 20-item and 13-item models showed an adequate fit, while the 13-item model was better (20-item:  $\chi^2(24, N = 284) = 63.06, p < .001$ ; NC = 2.63; CFI = .97; RMSEA = .08; and SRMR = .05; 13-item ( $\chi^2(24, N = 284) = 44.74, p = .006$ ; NC = 1.86; CFI = .98; RMSEA = .06; and SRMR = .04). The results of CFA using parcels are presented in Supplemental Table 1.

Next, we conducted CFA on the testing set to validate the three-factor structure of the 13-item model. The good model fit ( $\chi^2(62, N = 266) = 137.87, p < .001$ ; NC = 2.22; CFI = .93; RMSEA = .07; and SRMR = .06) confirmed the appropriateness of the model. The three scales were moderately correlated with one another ( $r = .32-.58$ ). Figure 1 shows the standardized factor loadings of the 13 EmQue items for the entire sample (training and testing data sets combined) and for boys and girls separately. The standardized factor loadings of the 20-item model are provided in Figure 2.

### ***Measurement Invariance Across Gender***

Multigroup CFA was performed to test the equality of the measurement parameters between gender groups for the 13-item model (Table 3). First, the fit statistics obtained from the configural (baseline) model showed a good fit ( $\chi^2(124, N = 550) = 289.44, p < .001$ ; NC = 2.33; CFI = .93; RMSEA = .05; and SRMR = .06). In the next step, the testing of the metric (weak) invariance showed a nonsignificant change in the model fit ( $\Delta\text{CFI} = .003$ ), indicating that the hypothesis of metric invariance was tenable. Testing scalar (strong) yielded  $\Delta\text{CFI}$  of .003, indicating that scalar invariance could be assumed.

Further, the model for testing homogeneity of variances and covariances yielded a non-significant change in  $\chi^2$  ( $\Delta\chi^2 = 11.22$ ,  $\Delta df = 6$ ,  $p = .08$ ). Lastly, the model for testing homogeneity of means yielded a non-significant change in  $\chi^2$  ( $\Delta\chi^2 = 6.64$ ,  $\Delta df = 3$ ,  $p = .08$ ). These outcomes indicated that variance, covariance, and latent means were homogeneous across gender.

### ***Internal Consistency***

As Table 4 shows, the Cronbach's alphas showed good internal consistencies: .75 for Emotional Contagion, .74 for Attention to Others' Feelings, and .84 for Prosocial Actions. The alpha values were comparable to the other EmQue versions in Dutch ( $\alpha = .58-.80$ ; Rieffe et al., 2010), Italian ( $\alpha = .73-.80$ ; Grazzani et al., 2016), Spanish ( $\alpha = .60-.83$ ; Lucas-Molina et al., 2018), and Lithuanian (.70-.83; Lazdauskas, & Nasvytienė, 2021). The omega values were also good: .76 for Emotional Contagion, .74 for Attention to Others' Feelings, and .85 for Prosocial Actions. The average interitem correlations were within the acceptable range for Emotional Contagion (.42) and Attention to Others' Feelings (.36). The average interitem correlation for Prosocial Actions (.57) was slightly higher than the suggested range (i.e., .15 to .50; Clark & Watson, 1995), indicating potential redundancy in this scale.

### ***Correlations with Age***

A significant positive correlation was observed for Prosocial Actions ( $r = .20$ ). However, age was not significantly correlated with Emotional Contagion ( $r = -.04$ ), nor with Attention to Others' Feelings, ( $r = .06$ ). The correlations between age and the three scales were comparable among the gender groups (Emotion Contagion:  $Z = .35$ ,  $p = .36$ ; Attention to Others' Feelings:  $Z = .35$ ,  $p = .36$ ; Prosocial Actions:  $Z = .61$ ,  $p = .27$ ). Both boys and girls had positive correlations between age and Prosocial Actions. Given that the latent parameters were found homogeneous across gender, a significant correlation between age and an EmQue scale can be interpreted as not attributable to any gender differences. Table 2 presents means and standard deviations of the three

EmQue scales by age category. Table 4 shows the statistical descriptors of the three scales in the 13-item model and the correlations between the scales.

### **Discussion**

The aim of this study was to examine the psychometric properties of the Japanese EmQue. The results confirmed the validity of its tridimensional factor structure (Emotional Contagion, Attention to Others' Feelings, and Prosocial Actions) and high internal consistencies of the scales, which has also been found in Western samples (Grazzani et al., 2016; Lucas-Molina et al., 2018; Rieffe et al., 2010). These findings suggest that the three-factor EmQue with 13 items can be a useful tool for measuring empathy in Japanese preschool children. Furthermore, the results supported measurement invariance across genders. Therefore, we can conclude that any differences in the EmQue scores between preschool boys and girls can be interpreted as actual differences in empathy among the participants. Prosocial Actions increased in older children while Emotional Contagion and Attention to Others' Feelings were unrelated to age. Overall, the Japanese EmQue is a suitable instrument to be used in Japanese-speaking populations and for cross-cultural comparisons.

While empathy has been pivotal in explaining the foundation of human sociality (Decety & Svetlova, 2012), little is known about whether it is construed and developed in the same manner cross-culturally. This study provides the first piece of empirical evidence that the three components of empathy noted in Hoffman's (1987) model can be observed and well differentiated in Japanese preschool children. Moreover, similar to previous studies that used the EmQue on Western children, we found a positive correlation between age and Prosocial Actions, while Emotional Contagion and Attention to Others' Feelings were unrelated to age (Grazzani et al., 2016; Lucas-Molina et al., 2018; Rieffe et al., 2010). These findings suggest the sequential nature of the model that Prosocial Actions develop after Emotional Contagion and Attention to Others' Feelings. For children to

conduct prosocial actions and to understand which action is appropriate given a certain situation, awareness of others' emotions, motor skills, and cognitive maturation are prerequisites (Hoffman, 1987; Imuta et al., 2016). Note that the average age of the children in this study was four years (range = 1-6 years). At this age, children have learned the skills for regulating and understanding basic emotions (Thompson, 1991; Vaish et al., 2009), yet their ability to understand the intention and beliefs of others and causes of others' emotions is just emerging (Eggum et al., 2011; Nicholas et al., 2009).

Noteworthy, both the 13-item short-form model and the original 20-item model showed an acceptable to good fit, while the 13-item model fitted better, according to the CFAs using individual items and parcels. Little and colleagues (1999) has pointed out that items with low loadings may not necessarily be poor items if they help the construct be better represented. The slightly higher average interitem correlation observed in the Prosocial Actions scale in the 13-item version also indicates a potential lack of breadth in this scale. Taken together, our findings showed that the three factors of empathy can be well identified in Japanese preschool children using both versions. Researchers are suggested to choose the version that fits their needs (e.g., covering a greater breadth vs. reducing the completion time for participants) while being aware of the potential limitations discussed below.

Although it is difficult to conclude why the original 20-item model did not achieve a good fit at the item level, one possibility could be that we recruited older children than the original study (Rieffe et al., 2010;  $M_{age} = 2.50$  years,  $SD = 1.08$ ; range = 1-5 years). Therefore, different developmental stages might be reflected by the models. There might also be cultural influences, considering that children are socialized to acquire behavioral patterns in line with what is valued in the cultural context (Trommsdorff, 2015). However, to evaluate the appropriateness of the items

deleted from the original model, qualitative feedback on the items and cross-cultural investigations may be needed to understand the extent to which these items represent the construct.

### **Limitations and Future Directions**

Some limitations should be noted for future studies. First, the original study (Rieffe et al., 2010) included mothers and fathers to validate the three separate scales within the EmQue while this study only included mothers to achieve homogeneity in the sample. However, as Grazzani et al. (2016) suggested, future studies should test the reliability and validity with a sample of mothers and fathers at an equivalent ratio.

Second, as previously mentioned, the Prosocial Actions scale in the short-form version showed a slightly higher average interitem correlation. This indicates that the breadth of this scale can be further improved. Future studies are suggested to explore the prosocial behaviors in preschool children and include qualitative feedback sessions, to allow for better interpretation of the appropriateness of the items in the two EmQue versions.

Third, the original EmQue has 20 items that concern both negative and positive emotions of others in each scale; but in the 13-item version, only the attention to others' feelings scale has three items that measure empathic responses to others' positive emotions (e.g., "When my child sees other children laughing, he/she starts laughing too"). Positive empathy can be dissociated from negative empathy and is related to higher levels of social competence and greater relationship strength (Telle & Pfister, 2016). An observational study by Sallquist et al. (2009) found that positive empathy is present early in life and can be reliably reported by mothers. Researchers using the 13-version EmQue should keep in mind that most items in the short form measure empathic responses to others' negative emotions.

Last, future studies are suggested to test the concurrent validity of the Japanese EmQue to examine the relation of the three scales with other parental questionnaires or observational tasks



that also assess the social-emotional development in preschool years, such as mind theory (Rieffe et al., 2010) and moral judgments (Ball et al., 2016). In addition, some temperaments have been found to be associated with empathy, such as inhibition, perceptual sensitivity, and fearfulness (Rajhans et al., 2015; Young et al., 1999). Moreover, some theorists have proposed that empathy develops in the context of the mother-child relationship, and the quality of parental attachment, maternal empathy, and family socialization influence development (Robinson et al., 1994; Trommsdorff, 2008; Van der Mark et al., 2002). The extent to which these areas are associated with the three components of empathy in an East Asian context is a topic worth further investigation.

### **Conclusion**

This study aimed to validate a tool for measuring empathy in preschool children in an Eastern Asian context and cross-culturally, which is currently lacking. The Japanese EmQue is proven appropriate for investigating the three components (Emotional Contagion, Attention to Others' Feelings, and Prosocial Actions) involved in the development of empathy in Japanese preschool girls and boys. It can be reliably compared to the other language versions of the EmQue, making it a suitable instrument for cross-cultural studies.

### References

- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79. <https://doi.org/10.1214/09-SS054>
- Atkins, D., Uskul, A. K., and Cooper, N. R. (2016). Culture shapes empathic responses to physical and social pain. *Emotion, 16*(5), 587–601. <https://doi.org/10.1037/emo0000162>
- Ball, C. L., Smetana, J. G., & Sturge-Apple, M. L. (2016). Following my head and my heart: Integrating preschoolers' empathy, theory of mind, and moral judgments. *Child Development, 88*(2), 597–611. <https://doi.org/10.1111/cdev.12605>
- Bollen, K. A. (1989). *Wiley series in probability and mathematical statistics. Applied probability and statistics section. Structural equations with latent variables*. John Wiley & Sons. <https://doi.org/10.1002/9781118619179>
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. Lonner and J. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Beverly Hills, CA: Sage.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Byrne, M. B. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates, Inc, Publishers.
- Cassels, T. G., Chan, S., & Chung, W. (2010). The role of culture in affective empathy: Cultural and bicultural differences. *Journal of Cognition and Culture, 10*(3), 309–26. <https://doi.org/10.1163/156853710X531203>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)

- Chopik, W. J., O'Brien, E., & Konrath, S. H. (2017). Differences in empathic concern and perspective taking across 63 countries. *Journal of Cross Cultural Psychology, 48*(1), 23–38. <http://dx.doi.org/10.1177/0022022116673910>.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience, 2*(1), 1–24. <https://doi.org/10.1016/j.dcn.2011.05.003>
- de Waal, F. B. M. (2009). *The age of empathy: Nature's lessons for a kinder society*. New York: Harmony Books.
- Dunn, T. J., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399–412. <http://dx.doi.org/10.1111/bjop.12046>
- Eggum, N. D., Eisenberg, N., Kao, K., Spinrad, T. L., Bolnick, R., Hofer, C., Kupfer, A. S., & Fabricius, W. V. (2011). Emotion understanding, theory of mind, and prosocial orientation: relations over time in early childhood. *The Journal of Positive Psychology, 6*(1), 4–16. <https://doi.org/10.1080/17439760.2010.536776>
- Flora, D. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practice in Psychological Science, 3*(4), 484–501. <https://doi.org/10.1177/2515245920951747>
- Grazzani, I., Ornaghi, V., Pepe, A., Brazzelli, E., & Rieffe, C. (2016). The Italian Version of the Empathy Questionnaire for 18- to 36-months-old children: Psychometric properties and

measurement invariance across gender of the EmQue-I13. *European Journal of Developmental Psychology*, 14(1), 118–126.

<https://doi.org/10.1080/17405629.2016.1140640>

Harkness, J. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35-56). New York, NY: Wiley.

Hoffman, M. L. (1987). The contribution of empathy to justice and moral judgment. In Eisenberg, N., and Strayer, J. (eds.), *Empathy and Its Development*, University of Cambridge Press, Cambridge, U.K.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

Imuta, K., Henry, J. D., Selcuk, B., & Slaughter, V. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology*, 52(8), 1192–1205. <https://doi.org/10.1037/dev0000140>

Ishiguro, M., Sakamoto, Y., & Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC. *Annals of the Institute of Statistical Mathematics*, 49, 411–434. <https://doi.org/10.1023/A:1003158526504>

Kishton, B. M. & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, 54, 757–765. <https://doi.org/10.1177/0013164494054003022>

Lazdauskas, T., & Nasvytienė, D. (2021). Psychometric properties of Lithuanian versions of empathy questionnaires for children. *European Journal of Developmental Psychology*, 18(1), 144–159. <https://doi.org/10.1080/17405629.2020.1772048>

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford

- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151-173. [https://doi.org/10.1207/S15328007SEM0902\\_1](https://doi.org/10.1207/S15328007SEM0902_1)
- Little, T. D., Lindenberger, U., & Nesselrode, J. R. (1999). On selecting indicators for multivariate measurement and modeling latent variables: When “good” indicators are bad and “bad” indicators are good. *Psychological Methods*, 4, 192–211. <https://doi.org/10.1037/1082-989X.4.2.192>
- Lucas-Molina, B., Sarmiento, R., Quintanilla, L., & Giménez-Dasí, M. (2018). The Spanish version of the Empathy Questionnaire (EmQue): Evidence for longitudinal measurement invariance and relationship with emotional regulation. *Early Education and Development*, 29(4), 467–476. <https://doi.org/10.1080/10409289.2018.1427929>
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
- Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, 94, 925-937. <https://doi.org/10.1037/0022-3514.94.6.925>
- Matsunaga M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures*, 2(4):260–293. <https://doi.org/10.1080/19312450802458935>
- Ministry of Health, Labour and Welfare (2017). The overview of 2016 data for single/divorced mothers (in Japanese). Retrieved from <https://www.mhlw.go.jp/file/04-Houdouhappyou->

[11923000-Kodomokateikyoku-Kateifukishika/0000188136.pdf](https://doi.org/10.1192/3000-Kodomokateikyoku-Kateifukishika/0000188136.pdf)

- Mustanski, B. S. (2001). Getting wired: Exploiting the Internet for the collection of valid sexuality data. *Journal of Sex Research, 38*, 292–301.  
<https://doi.org/10.1080/00224490109552100>
- Nicholas, S. R., Svetlova, M., & Brownell, C A. (2009). The role of social understanding and empathic disposition in young children's responsiveness to distress in parents and peers. *Cognition, Brain, Behavior: An Interdisciplinary Journal, 13*(4), 449–478. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3359013/pdf/nihms369944.pdf?tool=EBI>
- Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills, 105*, 997–1014.  
<https://doi.org/10.2466/PMS.105.7.997-1014>
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*, 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Rajhans, P., Missana, M., Krol, K. M., & Grossmann, T. (2015). The association of temperament and maternal empathy with individual differences in infants' neural responses to emotional body expressions. *Development and Psychopathology, 27*(4), 1205–1216.  
<https://doi.org/10.1017/S0954579415000772>
- R Core Team (2020). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- 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>

- Robinson, J. L., Zahn-Waxler, C., & Emde, R. N. (1994). Patterns of development in early empathic behavior: Environmental and child constitutional influences. *Social Development*, 3(2), 125–145. <https://doi.org/10.1111/j.1467-9507.1994.tb00032.x>
- Sallquist, J., Eisenberg, N., Spinrad, T. L., Eggum, N. D., & Gaertner, B. M. (2009). Assessment of preschoolers' positive empathy: Concurrent and longitudinal relations with positive emotion, social competence, and sympathy. *The Journal of Positive Psychology*, 4, 223–233. <https://doi.org/10.1080/17439760902819444>
- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, 26(1), 1–2. <https://doi.org/10.1027/1015-5759/a000001>
- 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\\_1](https://doi.org/10.1207/s15327078in1002_1)
- Telle, N. T., & Pfister, H. R. (2016). Positive empathy and prosocial behavior: A neglected link. *Emotion Review*, 8(2), 154–163. <https://doi.org/10.1177/1754073915586817>
- Thompson, R. A. (1991). Emotional regulation and emotional development. *Educational Psychology Review*, 3, 269–307. <https://doi.org/10.1007/BF01319934>
- Trommsdorff, G. (1995). Person context relations as developmental conditions for empathy and prosocial action: A cross-cultural analysis. In T.A. Kindermann & J. Valsiner (Eds.), *Development of person-context relations* (pp. 113–146). Hillsdale, NJ: Erlbaum.
- Trommsdorff, G. (2008). Child-rearing and children's empathy. *Perceptual and Motor Skills*, 72(2), 387–390. <https://doi.org/10.2466/PMS.72.2.387-390>
- Trommsdorff, G. (2015). Development of moral values in cultural contexts. In L. A. Jensen (Ed.) *The Oxford handbook of human development and culture: An interdisciplinary perspective*.

New York, NY: Oxford University Press.

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. <https://doi.org/10.1037/a0014322>

Van der Mark, I. L., Van IJzendoorn M. H., & Bakermans-Kranenburg, M. J. (2002).

Development of empathy in girls during the second year of life: Associations with parenting, attachment, and temperament. *Social Development, 11*(4), 451–468.

<https://doi.org/10.1111/1467-9507.00210>

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development:

The truth about false belief. *Child Development, 72*(3), 655–684.

<https://doi.org/10.1111/1467-8624.00304>

Young, S. K., Fox, N. A., & Zahn-Waxler, C. (1999). The relations between temperament and empathy in 2-Year-olds. *Developmental Psychology, 35*(5), 1189–1197.

<https://doi.org/10.1037//0012-1649.35.5.1189>

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and*

*Emotion 14*, 107–130. <https://doi.org/10.1007/BF00991639>

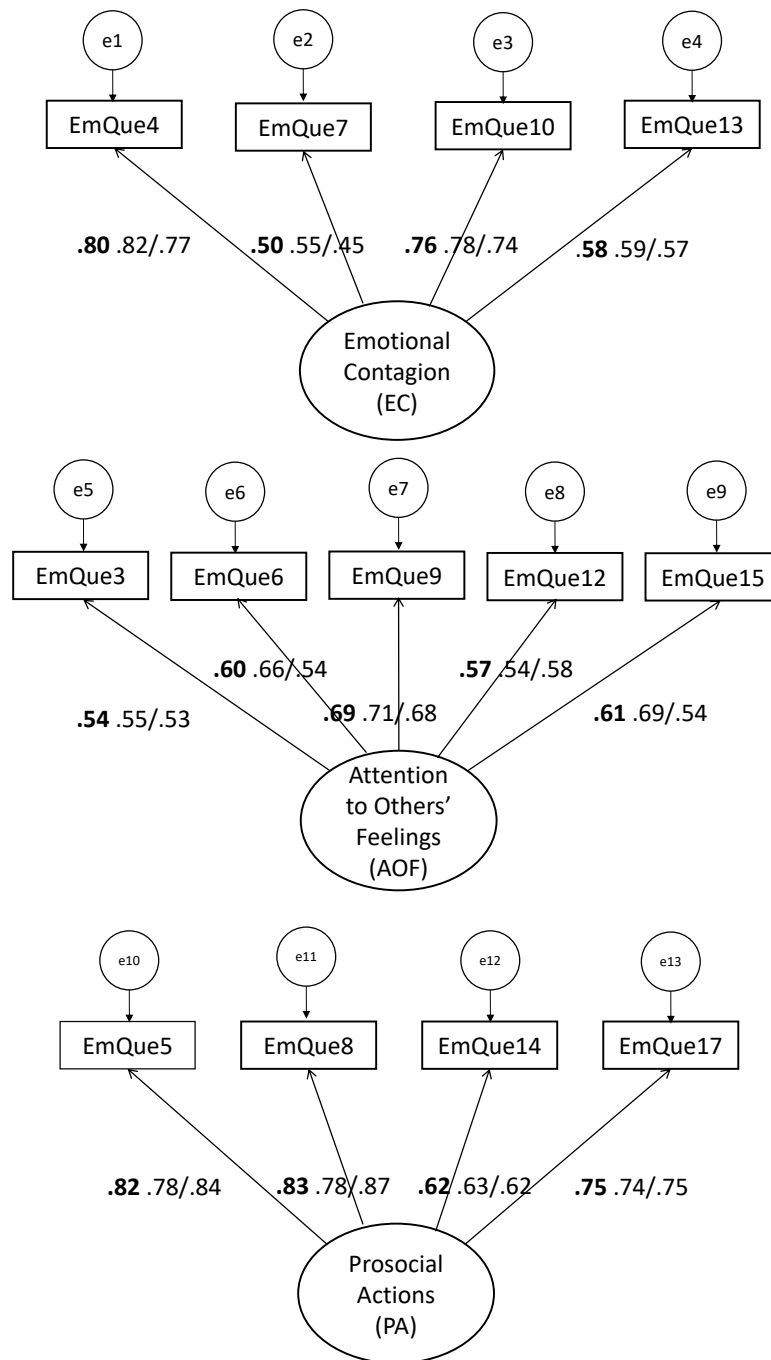
Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of

concern for others. *Developmental Psychology, 28*(1), 126–136.

<https://doi.org/10.1037/0012-1649.28.1.126>



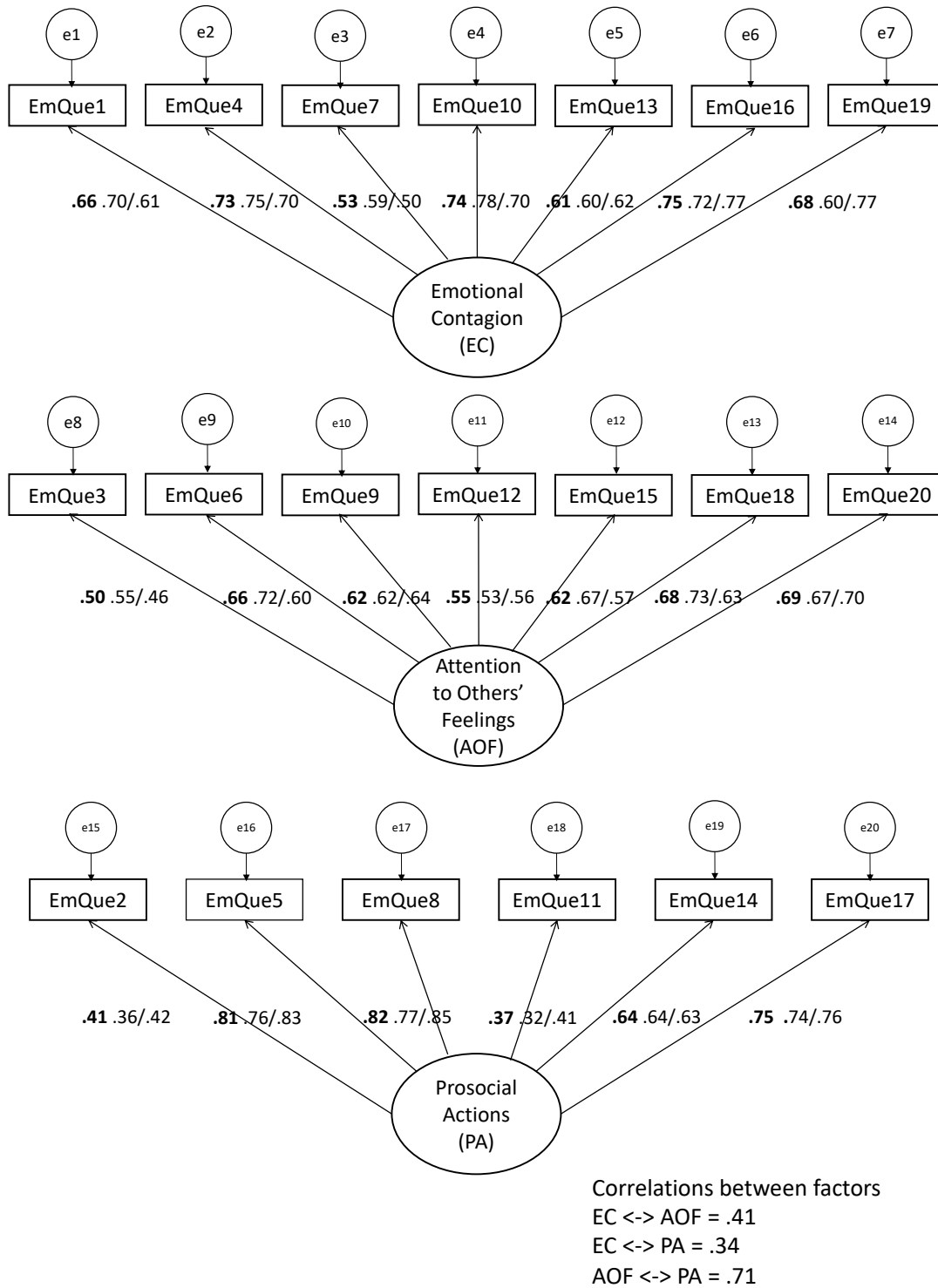
**Figure 1**  
*Factor Structure and Loadings of the 13-item Model*



Correlations between factors  
 EC <-> AOF = **.32** .30/.34  
 EC <-> PA = **.32** .32/.33  
 AOF <-> PA = **.58** .55/.62

*Note.*  $N = 266$ . Factor loadings are presented as **overall**, boys/girls.

**Figure 2**  
*Factor Loadings of the Original 20-item Model*



*Note.*  $N = 266$ . Factor loadings are presented as **overall**, boys/girls.

Table 1

*Empathy Questionnaire (EmQue) Items*Emotional Contagion (EC)

1. When another child cries, my child gets upset too.
4. **My child also needs to be comforted when another child is in pain.**
7. **When another child makes a bad fall, shortly after my child pretends to fall too.**
10. **When another child is upset, my child needs to be comforted too.**
13. **When another child gets frightened, my child freezes or starts to cry.**
16. When other children argue, my child gets upset.
19. When another child cries, my child looks away.

Attention to Others' Feelings (AOF)

3. **When my child sees other children laughing, he/she starts laughing too.**
6. **When an adult gets angry with another child, my child watches attentively.**
9. **My child looks up when another child laughs.**
12. **When adults laugh, my child tries to get near them.**
15. **My child looks up when another child cries.**
18. When another child is angry, my child stops his own play to watch.
20. When other children quarrel, my child wants to see what is going on.

Prosocial Actions (PA)

2. When I make clear that I want some peace and quiet, my child tries not to bother me.
5. **When another child starts to cry, my child tries to comfort him/her.**
8. **When another child gets upset, my child tries to cheer him/her up.**
11. When I make clear that I want to do something by myself (e.g. read), my child leaves me alone for a while.
14. **When two children are quarrelling, my child tries to stop them.**
17. **When another child gets frightened, my child tries to help him/her.**

*Note.* Items in bold are included in the 13-item EmQue. The number in brackets indicates the item number in the original 20-item EmQue (Rieffe et al., 2010).

Table 2  
*Fit Indices for Measurement Invariance Models Across Gender*

Parameter	Model fit indices					Indices of model fit differences		
	$\chi^2$	<i>df</i>	NC ( $\chi^2/df$ )	CFI	RMSEA [90% CI]	SRMR	$\Delta$ CFI	$\Delta\chi^2$ ( <i>p</i> )
Null model	2525.687*	156	-	-	-	-	-	-
Configural	289.438*	124	2.334	.930	.049 [.042, .057]	.058	-	-
Metric	293.638*	134	2.191	.933	.047 [.039, .054]	.059	.003	4.20 (.938)
Scalar	311.797*	147	2.121	.930	.045 [.038, .052]	.063	.003	18.16 (.156)
Homogeneity of variances and covariances	323.013*	153	2.111	.928	.045 [.038, .052]	.063	.002	11.216 (.082)
Homogeneity of means	316.372*	150	2.109	.930	.045 [.038, .052]	.054	.002	6.641 (.082)

*Note.*  $N = 550, p^* < .001$ .

Configural (baseline) = one factor loading constrained to unity and other items freely estimated.

Metric (weak) = factor loadings are equal across gender.

Scalar (strong) = item intercepts are equal across gender.

Homogeneity of variances/covariances (structural) = variances and covariances are equal across gender.

Homogeneity of means (factorial) = factorial means are equal across gender. Girls' means were set to 0.

Table 3

*Means, Variances, Standard Deviations, Skewness, and Correlation Coefficients of the EmQue Subscales and Age*

	1	2	3	4	Skewness	$\alpha$	$\omega$	$r$
1. Age	-	<b>-.04</b> -.03/-.06	<b>.06</b> .05/.08	<b>.20</b> .18/.23	.05	-	-	-
2. Emotional Contagion		-	<b>.43</b> .40/.47	<b>.38</b> .39/.38	.77	.75	.76	.42
3. Attention to Others' Feelings			-	<b>.72</b> .66/.77	-.17	.74	.74	.36
4. Prosocial Actions				-	.17	.84	.85	.57
Means	<b>4.06</b> 4.11/4.01	<b>1.93</b> 1.92/1.93	<b>2.94</b> 2.92/2.94	<b>2.56</b> 2.47/2.64	-	-	-	-
Variances	<b>1.47</b> 1.46/1.48	<b>0.68</b> 0.71/0.64	<b>0.39</b> 0.40/0.38	<b>0.83</b> 0.58/0.94	-	-	-	-

*Note.*  $N = 550$ . Boys = 273, Girls = 277.

Age = Age in years.  $\alpha$  = Cronbach's alpha;  $\omega$  = McDonald's omega;  $r$  = average inter-item correlation. Correlations, means, and variances are presented as **overall**, boys/girls.

Table 4  
*Means and Standard Deviations of the Three EmQue Subscales by Age Category*

EmQue subscales	Age (in year)					
	1 ( <i>n</i> = 1)	2 ( <i>n</i> = 118)	3 ( <i>n</i> = 141)	4 ( <i>n</i> = 127)	5 ( <i>n</i> = 138)	6 ( <i>n</i> = 25)
Emotional	1.75	1.95	1.97	1.93	1.83	2.09
Contagion	(-)	(0.77)	(0.79)	(0.70)	(0.73)	(0.91)
Attention to	2.20	2.89	3.01	2.88	2.95	3.13
Others' Feelings	(-)	(0.85)	(0.76)	(0.71)	(0.77)	(0.95)
Prosocial Actions	1.75	2.30	2.46	2.68	2.72	2.80
	(-)	(0.87)	(0.85)	(0.88)	(0.93)	(0.90)

*Note.* 5-point scale (1 = *never* to 5 = *always*)

Supplemental Table 1

*Empathy Questionnaire (EmQue) Items, Parcels, and Standardized Factor Loadings*

	20-item	13-item
<u>Emotional Contagion (Parcel 1)</u>	.73	.53
<b>7. When another child makes a bad fall, shortly after my child pretends to fall too</b>		
16. When other children argue, my child gets upset		
<u>Emotional Contagion (Parcel 2)</u>	.81	.83
<b>10. When another child is upset, my child needs to be comforted too</b>		
<b>13. When another child gets frightened, my child freezes or starts to cry</b>		
<u>Emotional Contagion (Parcel 3)</u>	.89	.78
1. When another child cries, my child gets upset too		
<b>4. My child also needs to be comforted when another child is in pain</b>		
19. When another child cries, my child looks away (R)		
<u>Attention to Others' Feelings (Parcel 1)</u>	.77	.63
<b>3. When my child sees other children laughing, he/she starts laughing too</b>		
20. When other children quarrel, my child wants to see what is going on		
<u>Attention to Others' Feelings (Parcel 2)</u>	.77	.62
<b>12. When adults laugh, my child tries to get near them</b>		
18. When another child is angry, my child stops his own play to watch		
<u>Attention to Others' Feelings (Parcel 3)</u>	.72	.69
<b>6. When an adult gets angry with another child, my child watches attentively</b>		
<b>9. My child looks up when another child laughs</b>		
<b>15. My child looks up when another child cries</b>		
<u>Prosocial Actions (Parcel 1)</u>	.88	.87

**8. When another child gets upset, my child tries to cheer him/her up**

11. When I make clear that I want to do something by myself (e.g. read), my child leaves me alone for a while

Prosocial Actions (Parcel 2) .83 .83

2. When I make clear that I want some peace and quiet, my child tries not to bother me

**5. When another child starts to cry, my child tries to comfort him/her**

Prosocial Actions (Parcel 2) .72 .76

14. **When two children are quarrelling, my child tries to stop them**

17. **When another child gets frightened, my child tries to help him/her**

---

*Note.* Items in bold are included in the 13-item EmQue. The number in brackets indicates the item number in the original 20-item EmQue. R = reversely scored.