# Closeness, trust, and perceived social support in child-robot relationship formation

## Development and validation of three self-report scales

Caroline L. van Straten, Rinaldo Kühne, Jochen Peter, Chiara de Jong, and Alex Barco
University of Amsterdam

Social robots and their interactions with children are becoming increasingly sophisticated, with the emergence of child-robot relationships as a likely result. However, adequate measurement instruments that tap into concepts associated with child-robot relationship formation are scarce. We aimed to develop three measures that can be used to assess children's closeness to, trust in, and perceived social support from, a social robot. We established the validity and reliability of these measures among 87 Dutch children aged 7 to 11 years old. Because of their shortness, the measures can efficiently be applied by scholars aiming to gain insight into the general process of child-robot relationship formation and its specific sub-processes, that is, the emergence of closeness, trust, and perceived social support.

**Keywords:** child-robot relationship formation, closeness, trust, perceived social support, measure development, scale validation

Humans and, in particular, children, tend to form social bonds with both animate entities and inanimate objects (Borenstein & Arkin, 2016). Children's conceptions of robots seem to fall somewhere in between the animate and inanimate (e.g., Kahn, Gary, & Shen, 2013), and their interaction with these robots is inherently social (Salter, Werry, & Michaud, 2008). Given the growing sophistication of the robots children encounter in their everyday lives (Kahn et al., 2013), social relationships between children and robots are likely to become more profound in the near future (Borenstein & Arkin, 2016; Kahn et al., 2013). Judgements about the desirability of children's relationship formation with robots vary greatly. Some

https://doi.org/10.1075/is.18052.str
*Interaction Studies* 21:1 (2020), pp. 57–84. issn 1572-0373 | e-issn 1572-0381

scholars point to potential benefits of child-robot relationship formation (e.g., it might promote children's wellbeing; Pearson & Borenstein, 2014). Others, in contrast, point out potentially detrimental psychological and developmental consequences of child-robot relationships (e.g., in the context of children's interpersonal relationships; Kahn et al., 2013).

Currently, however, adequate measurement instruments to assess the formation of child-robot relationships are scarce, which hinders our understanding of the causes and positive or negative consequences of relationship formation between children and social robots. Scholars in the field of social robotics have repeatedly expressed the need for standardized measurement instruments to assess outcomes of both human-robot interaction (HRI) and child-robot interaction (CRI; e.g., Bartneck, Kulić, Croft, & Zoghbi, 2009; Eyssel, 2017; De Jong, Peter, Kühne, & Barco, in press; Van Straten, Peter, & Kühne, 2019). As pointed out by Belpaeme et al. (2013), due to their specific stage of physiological and mental development, "children are not just small adults" (p. 453), and child-appropriate measurement instruments are required.

## Current practice and measurement issues

Multiple CRI studies have assessed concepts that are relevant to the formation of child-robot relationships, sometimes in the context of a broader topic, such as children's robot perception. Some studies, for example, investigated general constructs, such as relationship building (Hieida et al., 2014; Kruijff-Korbayova et al., 2014) and companionship (Guneysu & Arnrich, 2017; Saint-Aimé, Grandgeorge, Le Pevedic, & Duhaut, 2011), while others focused on more narrow concepts, such as social attraction (Guneysu & Arnrich, 2017; Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009), trust (e.g., Bethel et al., 2016; Looije, Van der Zalm, Neerincx, & Beun, 2012), perceived social support (e.g., Ahmad, Mubin, & Orlando, 2016; Leite, Castellano, Pereira, Martinho, & Paiva, 2014), and attachment (Díaz, Nuño, Saez-Pons, Pardo, & Angulo, 2011).

In methodological terms, these studies mostly took a quantitative approach, using self-report measures with closed-ended items. Some also drew on data from extensive interviews (i.e., Ahmad et al., 2016; Díaz et al., 2011; Leite et al., 2014) or observational findings (i.e., Hieida et al., 2014; Saint-Aimé et al., 2011). Except for Bethel et al. (2016), who addressed both trust and perceived support, the studies typically assessed single aspects of children's relationship formation with robots. Studies often used single-item measures, or did not specify the measurement instrument (for multiple-item measures, see Ahmad et al., 2016; Kose-Bagci et al., 2009; Leite et al., 2014). Response formats varied from dichotomous items (e.g., Guneysu & Arnrich, 2017) to 7-point Likert scales (Kruijff-Korbayova et al., 2014),

some with smiley visualizations (Ahmad et al., 2016; Leite et al., 2014; Saint-Aimé et al., 2011).

There are several shortcomings in the measurement of concepts relevant to the formation of child-robot relationships that mirror measurement issues in CRI research more generally (see, for instance, De Jong et al., in press; Peter, Kühne, Barco, De Jong, & Van Straten, 2019). In line with Eyssel's (2017) recommendations for the field of HRI in general, a recent review of the literature on child-robot relationship formation (Van Straten et al., 2019) identified five challenges that are particularly relevant to the present study. First, clear conceptual definitions of the concepts under investigation are often missing (for a notable exception, see Leite et al., 2014). Second, the transparency and comprehensiveness of studies' reporting on methodological information varies (e.g., concerning measurement instruments and for replication purposes; see also Baxter, Kennedy, Senft, Lemaignan, & Belpaeme, 2016) (for studies that did address at least one of these issues, see Bethel et al., 2016; Kose-Bagci et al., 2009; Leite et al., 2014). Third, ceiling effects are often encountered, which may be the result of social-desirability biases in children's answers (see also Belpaeme et al., 2013). Fourth, studies on child-robot relationship formation that employ both self-report and observational measures are frequently confronted with inconsistent findings. Fifth and finally, tests of measurement validity, such as assessing factorial or concurrent validity, tend to be neglected or can, as in the case of factorial validity, not be conducted due to the use of single-item measures. Consequently, the measures cannot always be guaranteed to accurately reflect the theoretical concept they represent. Whereas the lack of validated scales does not mean that the findings obtained with these instruments are invalid, the use of validated scales would allow for more conclusive tests of the theoretical relationship between constructs (Van Straten et al., 2019).

Given these limitations, a new set of standardized measures seems useful. Such a new set of measures would help pave the way for cumulative research, which in turn would be beneficial to the conclusive power of the field as a whole. We focus on the development of self-report measures because the use of self-report measures, certainly when compared to the implementation of observational measures, is relatively straightforward and less time-consuming. Moreover, for mental concepts, such as those relevant to child-robot relationship formation, the development of observational measures is problematic: Observable behavioral cues are typically related to multiple mental states and are, accordingly, related to mental concepts in different ways in the literature. Observational cues thus do not easily lend themselves for the assessment of mental concepts (e.g., Andrés, Pardo, Díaz, & Angulo, 2015; Van Straten et al., 2019).

Against this background, the present paper reports the results of a study that aimed at developing and validating three self-report scales that, together, can be used to assess child-robot relationship formation. We first identify and define

important elements of interpersonal relationship formation that can reasonably be expected to also play a role in the formation of social relationships between children and robots. Subsequently, we describe the development of three self-report scales – of closeness, trust, and perceived social support – and their validation among 87 children aged 7 to 11 years old.

## Key concepts of interpersonal relationship formation

Given the above identified challenge of linking operationalizations to clear theoretical conceptualizations, we first inspected the literature on interpersonal relationship formation in order to identify, and theoretically define, key concepts of interpersonal relationship formation – as they are also likely to play a crucial role in the emergence of child-robot relationships.

Relationship formation has been defined as a "relationship's progression toward closeness" (Berscheid & Regan, 2005, p. 191), whereas relationship maintenance is generally considered to start when closeness ceases to increase (Berscheid & Regan, 2005). Inherent to this definition is the assumption that closeness constitutes a first key concept of relationship formation. Closeness can be defined as a feeling of connectedness or intimacy that could potentially result in the development of a friendship (Sternberg, 1987), and is generally considered a gradual concept that may occur in various degrees (e.g., Guerrero, 1997; Mehrabian, 1972).

A second concept that figures, like closeness, in most theories of interpersonal relationship formation, is trust (Berscheid & Regan, 2005). According to Larzelere and Huston (1980, p. 596), "trust exists to the extent that a person believes another person […] to be benevolent and honest". Both closeness and trust have been considered primary functions of children's peer-to-peer relationships (Bauminger-Zviely & Agam-Ben-Artzi, 2014), which further justifies their relevance to the development of child-robot relationships.

Third and finally, perceived social support overlaps – but does not coincide – with closeness and trust. It can be defined as "perceived […] instrumental and/or expressive provisions supplied by the community, social networks, and confiding partners" (Lin, 1986, p. 18). As pointed out by Berscheid and Regan (2005), someone's social network does not naturally coincide with someone's social support network. For instance, close friends may not always be supportive and one can be honest and benevolent without actually providing support. Therefore, perceived social support is treated as a distinct key concept.

The centrality of closeness, trust, and perceived social support to interpersonal relationship formation is mirrored in CRI studies that address (elements of) the emergence of child-robot relationships, as outlined above in 'Current

practice and measurement issues'. In sum, then, we focused on closeness, trust, and perceived social support as key concepts of relationship formation and tried to provide a scale for each of these concepts that can be used in empirical research on CRI.

## Method

### Scale development

As a first step in the development of our scales of closeness, trust, and perceived social support, we searched for existing self-report measures of these concepts in both the CRI and interpersonal relationship literature. Below we will discuss, per concept, which existing scales (if any) were used as a starting point as well as the procedure through which we subsequently established the items of our new measurement instruments. For each of the scales, item content, order, formulation, and translation into Dutch were discussed and iteratively refined by the authors until sufficient face validity was achieved. At several points in time, primary school teachers were consulted for their opinion on the use of certain words and grammatical constructions. Pre-final versions of the scales were piloted for intelligibility among four children, and adjustments were made in case issues were encountered.

We focused on children from middle childhood (i.e., 6 to 12 years old; see Cole, Cole, & Lightfoot, 2005). We targeted this age group as in middle childhood, children start to form increasingly close friendships (Cole et al., 2005), in which trust begins to play an important role (Bernath & Feshbach, 1995; Furman & Bierman, 1984; Kahn & Turiel, 1988). Moreover, compared to younger children, children in middle childhood base their friendships upon more fundamental interpersonal criteria (Bigelow, 1977). Thus, middle childhood is a developmental period during which children's relationship formation can be meaningfully investigated.

The selection and development of the items for our scales was guided by three general considerations. First, we aimed at measures that distinctly and comprehensively assess the constructs of interest, with minimal conceptual overlap between their item sets. Second, we tailored our measures toward use with children. In addition to avoiding complex expressions and sentence structures, we refrained from using negations because rating negative statements on a self-report scale can be – especially for younger children – cognitively too demanding (see Marsch, 1986). Similarly, we avoided qualifiers (e.g., 'a bit') because, in a gradual response scale, they would complicate children's responding. For instance, if

a child who perceives the robot to be 'completely like a friend' would have to judge the statement "the robot feels a bit like a friend" on our five-point scale (see below), none of the answer options would fit: Even "applies completely" would mean that the robot 'completely' feels *a bit* like a friend, which would not accurately reflect the child's perception. Third, if items were selected from an existing scale and item statistics were available, we took these into account. Items with high factor loadings or high item-total correlations were preferred in case they did not overlap with other concepts and were child-appropriate.

Appendix A shows the items that made up the final versions of the three self-report scales. Our study was conducted with the social robot Nao (Softbank). Children indicated their answers on a 5-point Likert scale running from "does not apply at all" to "applies completely". Following the approach taken by Severson and Lemm (2016), answer options were accompanied by bars of increasing height that clarified their meanings to the children, but lacked any indications with respect to the valence of the answers (e.g., smileys, colors) to avoid social desirability biases. We used a single response format for all closed-ended items, because children at the age of 7 to 11 have not yet fully developed their language and reading skills (Borgers, De Leeuw, & Hox, 2000). Therefore, these children should not be overburdened (Borgers et al., 2000; Ólafsson, Livingstone, Haddon, 2013; Punch, 2002). In this light, using simple and concrete item formulations and minimizing the number of different question and answer formats can prevent confusion (see Ólafsson et al., 2013). The answering scale can be found in Appendix B.

*Closeness*

As defined above, closeness constitutes a gradually emerging feeling of connectedness or intimacy (Guerrero, 1997; Mehrabian, 1972; Sternberg, 1987). Some pictorial measures of closeness exist (i.e., Aron, Aron, & Smollan, 1992; Popovic, Milne, & Barrett, 2003), of which one has been adjusted and applied in a CRI context (see Kory Westlund, Park, Williams, & Breazeal, 2018). However, for the purpose of our study, these measures are both too narrow, as they do not cover the full conceptual bandwidth of closeness (Dibble, Levine, & Park, 2012), and too broad, as their outcomes are likely influenced by, for instance, feelings of trust and perceived similarity to the self. Therefore, we developed a semantically more distinctive multi-item 5-point scale of closeness. To the best of our knowledge, no scale is currently available that specifically assesses feelings of closeness and reasonably lends itself to be used (or adapted for use) in CRI research.

While Sternberg's (1997) Triangular Love Scale does contain a relevant intimacy dimension, its items overlap with the concept of perceived social support and do not readily lend themselves to be adjusted for use with children. Similarly,

the Unidimensional Relationship Closeness Scale (URCS) by Dibble et al. (2012) uses a rather broad and child-inappropriate item set that does not exclusively tap into feelings of closeness, but also addresses feelings of engagement, for instance. Because we aimed at developing a measure that distinctively addresses the concept of closeness, we did not consider these scales useful for our purposes. Thus, we developed five new items (i.e., "Nao is a friend", "I feel comfortable around Nao", "Nao and I are becoming friends", "Nao and I are a good match", "Nao feels like a friend to me"). As 'closeness' is a rather abstract term, as are 'connectedness' and 'intimacy', we replaced the term by the more easy-to-grasp and child-appropriate term 'friendship.' Moreover, our scale represents the gradual nature of closeness through items of varying intensity.

*Trust*

Larzelere and Huston (1980) defined trust as perceived benevolence and honesty. Although Betts, Rotenberg, and Trueman (2009) developed a trust scale for children, the items of this scale consist of scenario descriptions that would imply an unjust degree of humanlikeness of the robot. As we are not aware of the existence of other child-appropriate self-report scales that distinctively address trust and lend themselves to use in CRI research, we adapted four items of Larzelere and Huston's (1980) eight-item Dyadic Trust Scale for adults. We used an identical item introduction text for each item (i.e., "I feel that…"). The introduction text clarified that the items address one's subjective experiences rather than an objective fact (i.e., "I feel that Nao is trustworthy" rather than "Nao is trustworthy").

*Perceived social support*

Perceived social support refers to the experience of available care and help (see Lin, 1986). The development of our four items was inspired by the peer subscale of Gordon-Hollingsworth et al.'s (2016) Social Support Questionnaire for Children (SSQC), which was validated among children and adolescents aged 8 to 18 years, and Leite et al.'s (2014) adjusted version of the SSQC for use in CRI research, which they tested for reliability (but did not validate) among 16 children aged 8 to 9. Both scales ask participants to indicate how much they agree with factual statements, such as "A peer encourages me" (Gordon-Hollingsworth et al., 2016, p. 135) or "[The robot] encourages me" (Leite et al., 2014, p. 340). We adjusted the formulation of the items such that they referred to hypothetical situations, which a child can evaluate regardless of the number of interactions and the particular behavior of the robot (e.g., "If I were in trouble, I could count on Nao" rather than "I can count on Nao").

## Scale validation

We validated our measures with a confirmatory-factor-analysis (CFA) approach. CFA is an appropriate methodological procedure if latent constructs are to be identified and a substantial body of knowledge exists that can be used to deductively create an a priori model (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Compared to a rather data-driven exploratory-factor-analysis approach, our theory-driven approach allowed us to work with a limited number of items that reflected the definitional cores of the three concepts and, importantly, also prevented that children would be overburdened by large item sets.

We followed a four-step procedure. First, we performed a separate, unidimensional CFA for each of the key concepts. Specifically, we investigated the model fit of the unidimensional CFAs and the factor loadings of the items to assess the factorial validity of each measure as a crucial criterion of measurement quality (Byrne, 2011). Second, as the loadings of a set of indicators on their latent variable are sensitive to the inclusion of additional concepts in the CFA, we conducted a multidimensional CFA of closeness, trust, and perceived social support to further validate the findings of the unidimensional CFAs. We inspected the model fit as well as the correlations between the factors. Third, we estimated the reliability of each scale by calculating Cronbach's alpha and additionally inspected the item fit by computing item difficulty and corrected item-total correlations (i.e., item selectivity; Moosbrugger & Kelava, 2007). High item difficulty values indicate that the items in question are relatively easy to answer and, accordingly, that participants have high scores on the respective items. Fourth, we assessed the concurrent validity of our measures by investigating correlations between outcomes of our newly developed scales and scales of concepts that, based on prior research, can be expected to be related.

We conducted the CFAs in Mplus 7.11 using Maximum Likelihood estimation with standard errors and a mean-adjusted chi-square test statistic that are robust to non-normality (MLM; Muthén & Muthén, 2012). We calculated the chi-square test, the comparative fit index (CFI), and the standardized root mean square residual (SRMR). The root mean square error of approximation (RMSEA) was not taken into account as this measure often falsely indicates poor model fit when the sample size is small and the degrees of freedom of the model are low (Kenny, Kaniskan, & McCoach, 2015). We consider an acceptable model fit to be reflected in the following outcomes: The chi-square test should be non-significant (Byrne, 2010; Schermelleh-Engel, Moosbrugger, & Müller, 2003); the CFI over .95 (Byrne, 2010; Hu & Bentler, 1999); and the SRMR under .08 (Hu & Bentler, 1999). Cronbach's alpha was calculated in SPSS Statistics 24. Items of each scale were averaged, and means, standard deviations, skewness, and kurtosis were inspected. Finally, SPSS was also used to calculate bivariate correlations between our scales and scales of related constructs.

*Validation measures*

We selected agreeableness, social anxiety, attachment, perceived similarity, and fear as validation measures. Our validation measures consisted of items selected from established scales and were slightly adapted, if necessary, to increase their clarity and intelligibility for children. As with the new measurement instruments, children indicated their answers on the visualized 5-point scale discussed above. Although replacing the original answering scales of the validation measures might affect the outcomes of the validation procedure, we decided to use a single response format for all measures in order not to overburden children's cognitive capacities (Borgers et al., 2000; Ólafsson et al., 2013; Punch, 2002).

*Agreeableness*

As defined by Goodboy and Booth-Butterfield (2009, p. 210), "agreeableness is the tendency to get along with others and to sympathize." People with low agreeableness desire less closeness to romantic partners, which according to Goodboy and Booth-Butterfield (2009) can be explained by a need for independence. Similarly, research on interpersonal relationships in organizations showed that agreeableness is positively associated with relationship intimacy (Wu, Foo, & Turban, 2008). People with an agreeable personality moreover tend to trust others more (Dohmen, Falk, Huffman, & Sunde, 2008) and reported higher levels of social support than people low in agreeableness (for an overview, see Swickert, Hittner, & Foster, 2010). Therefore, we expect agreeableness to be positively correlated with closeness, trust, and perceived social support.

Our measure of agreeableness consisted of two items selected from the 13-item agreeableness dimension of the Big Five Questionnaire for Children (BFQ-C), which was developed and validated among children in grades four to eight (Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003; for a Dutch translation, see Muris, Meesters, & Diederen, 2005). Following Rammstedt and John (2007) who successfully assessed the Big Five personality traits among adults with only two items per dimension, we selected the two items from the BFQ-C that were most similar to the agreeableness items used by Rammstedt and John (2007). An overall agreeableness score was computed by averaging the items ($M = 4.07$, $SD = 0.80$). The Spearman-Brown coefficient, which is considered the most appropriate reliability statistic for two-item scales (Eisinga, Te Grotenhuis, & Pelzer, 2013), was .65.

*Social anxiety*

Research has found evidence for a negative relationship between social anxiety and the development of close friendships (Vernberg, Abwender, Ewell, & Beery, 1992). In terms of trust, Muris, Meesters, Van Melick, and Zwambag (2001) have shown that among early adolescents (i.e., aged 12 to 14), social anxiety was nega-

tively correlated with trust in peers and parents. Additionally, a study on adolescents' peer relationships found that social anxiety was negatively correlated with perceived social support (La Greca & Lopez, 1998). Consequently, we expect a negative correlation between social anxiety and closeness, trust, and perceived social support.

Social anxiety was assessed through La Greca and Stone's (1993) six-item Social Avoidance and Distress – Specific to New Peers or Situations (SAD-New) subscale of the revised Social Anxiety Scale for Children (SASC-R). An overall social anxiety score was computed by averaging the items ($M = 3.20$, $SD = 0.88$, $\alpha = .82$).

*Attachment*

Attachment has been defined as "an enduring affectional bond of substantial intensity", and plays an important role in relationship maintenance (Armsden & Greenberg, 1987, p. 428). Even though it is unlikely that attachment develops fully after a single interaction, measures relevant to child-robot relationship formation often lead to ceiling effects (Van Straten et al., 2019). Therefore, we initially aimed at a fourth self-report measure – attachment – as the greater intensity of attachment as compared to, for instance, closeness, might in this context yield more informative outcomes. However, the corresponding scale only exhibits suboptimal psychometric properties. Still, given its centrality to the topic of interpersonal relationships, attachment may serve as a useful validation measure for the concepts of closeness, trust, and perceived social support.

Attachment was assessed by four items that Valkenburg and Peter (2007) had selected, for a study among adolescents, from the peer scale of Armsden and Greenberg's (1987) Inventory of Parent and Peer Attachment (IPPA). Affirmative statements were transformed into conditionals to make them applicable to a one-time CRI setting. One item (i.e., "My friends help me to understand myself better") was considered too difficult and replaced with another item from the IPPA (i.e., "My friends accept me as I am"; Armsden & Greenberg, 1987, p. 452), which was adapted for use in CRI. An overall attachment score was computed by averaging the items ($M = 4.13$, $SD = 0.56$, $\alpha = .49$).

*Perceived similarity*

Perceived similarity is generally considered a primary determinant of interpersonal evaluations (Byrne, 1971). Accordingly, children often develop friendships based on similarities in demographical characteristics, personal preferences, and behavioral similarities (Cole et al., 2005; Furman & Bierman, 1984; Gifford-Smith & Brownell, 2003). Perceived similarity hence plays an important role in the emergence of closeness and trust, which are considered central to relationship formation (Berscheid & Regan, 2005). Moreover, Lakey, Ross, Butler, and Bentley

(1996) found that adults' perceived similarity influenced their impressions of strangers' social support. For these reasons, we expect a positive correlation between perceived similarity and closeness, trust, and perceived social support.

We measured perceived similarity by four items selected from the 15-item Attitude Homophily dimension by McCroskey, McCroskey, and Richmond (2006). We selected affirmative items that had a high item-total correlation and were child-appropriate. We did not adapt the items other than replacing the original 'this person' with 'Nao.' An overall perceived similarity score was computed by averaging the items ($M = 2.86$, $SD = 0.80$, $\alpha = .72$).

*Fear*

Fear is evoked when an individual perceives a situation as threatening or dangerous (see Rapee, 1997). Consequently, if an individual is afraid of another one, it is unlikely that a close and trustful relationship will emerge in which s/he experiences social support from the other. Research for instance found that fear limits both trust itself and engagement in trust-building behaviors (Gambetta, 1988). Moreover, attachment style theory states that fearful people generally refrain from engaging in close relationships and prefer to maintain distance, as well as that they tend to experience less social support from their friends (Ognibene & Collins, 1998). Thus, we expect fear to be negatively correlated with closeness, trust, and perceived social support.

Our measure of fear was based on Spielberger's (1973) State-Trait Anxiety Inventory for Children (STAIC), which was validated among children from grade one to four by Hedl and Papay (1982). We selected four items from the state anxiety-present dimension that had high factor loadings and fit our purposes. The items were adjusted to specify Nao as the source of fear. In two items, more applicable equivalents of fear were used to replace the original wording (i.e., 'uneasy' instead of 'troubled', and 'afraid' instead of 'worried'). An overall fear score was computed by averaging the items ($M = 1.41$, $SD = 0.55$, $\alpha = .67$).

## Data collection

Prior to data collection, ethical approval was obtained from the Ethics Review Board of the Faculty of Social and Behavioral Sciences at the University of Amsterdam. The present paper reports on data that were gathered in a larger investigation that aimed to develop and validate a more encompassing set of standardized measures for CRI research than the subset presented here. The larger investigation assessed (a) children's robot perceptions, (b) their internal states during, and (c) appreciation of, the interaction, and (d) children's personality characteristics and developmental level. From that same investigation, we have already used data

on children's answers to an open-ended question on trust in another publication (see Van Straten, Peter, Kühne, De Jong, & Barco, 2018). Moreover, a paper on a measure of children's intentional acceptance of robots has been submitted for publication (De Jong, Kühne, Peter, Van Straten, & Barco, 2018). The general plan for this data collection has been described elsewhere (De Jong, Van Straten, Peter, Kühne, & Barco, 2018).

*Participants*

Data were collected among 88 children at two Dutch elementary schools. Both the schools and children's parents were asked for active informed consent. One child did not complete the interaction and was not included in the analyses. Our eventual sample thus consisted of 87 children (48 female, 39 male). Children's age ranged from 7 to 11 years ($M = 9.17$, $SD = 0.85$).

*Procedure*

The study was conducted by a female experimenter and a female interviewer, who familiarized the children, at class-level, with the study procedure before the start of the first interaction session to increase children's comfort (Vogt, De Haas, De Jong, Baxter, & Krahmer, 2017). They explained to the children that their participation was voluntary; that no personally identifiable information would be published; and that they could at any point in time withdraw their participation in case they no longer wanted to take part in the study. General questions were answered, while questions about the robot were postponed until the debriefing to prevent the answers from influencing children's initial robot perceptions.

Upon entering the experimental room, in which any distractions were minimized, children were asked to sit down on the floor in front of the robot. The children could freely choose how closely they sat to the robot. Once seated, the experimenter asked the child whether s/he was still willing to participate, and reminded him/her that participation could be withdrawn at any time without providing a reason. After the child had indicated to be ready to begin the interaction, the experimenter started the robot. Video recordings were made during the CRI sessions (but not when administering the questionnaire) if the parents of the child had permitted us to do so.

The study used a Wizard of Oz (WOZ) set-up, such that the experimenter remained present during the interaction to operate the robot from a distance. Children interacted with the robot for approximately eight minutes. After the experimenter introduced the child to the robot, the robot asked the child several questions during a short small-talk phase. Then, the experimenter suggested that the child and the robot play a guessing game, during which the robot made a series of assertions (e.g., "I love to eat fries") of which the child had to guess

whether they were true or false. After each guess, the robot provided the correct answer and a short explanation (e.g., "Like toys and computers, robots do not eat; instead they need electricity to function."). To make the game less repetitive, the robot asked the child some questions (e.g., "What is your favorite color?") throughout the interaction. It never suggested to have any truly human capacities such as feelings or consciousness to prevent false impressions of the robot (Broadbent, 2017). Once the robot had said goodbye to the child, the experimenter brought the child to another room in which the interviewer conducted the questionnaire.

The interviewer orally presented the children with 75 closed-ended and eleven open-ended items (for a discussion of children's answers to an open-ended item on trust, see Van Straten et al., 2018). The items (of which not all are used in this study) addressed children's perception of the robot, their internal states during an interaction with the robot, their appreciation of the interaction with the robot, and children's cognitive development as well as personality. Following Leite, Pereira, and Lehman (2017), children were presented with several practice items (e.g. "I like candy", "I like Brussel's sprouts") before starting the questionnaire, in order to familiarize them with the item format and the answering scale. Once a child indicated to fully understand the procedure, the questionnaire was administered. The entire questionnaire procedure took about 15 minutes ($M = 15.29$, $SD = 3.47$).

After all the children had completed their participation, they were debriefed at class-level through a ten-minute presentation (for a similar approach, see Schadenberg, Neerincx, Cnossen, & Looije, 2017). In this presentation, the experimenter and interviewer addressed the mechanical nature and working of the robot, and emphasized that the interaction was fully scripted and identical for each child. They pointed out some important differences between humans and robots and answered any remaining questions. Finally, children received a small gift to thank them for their attention and/or taking part in the study.

## Results

### Unidimensional confirmatory factor analysis

Using Mplus 7.11, we first conducted CFAs for closeness, trust, and perceived social support separately, to make sure each model had a good fit to the data. As our data were not normally distributed, we used a Maximum Likelihood estimation (i.e., MLM) procedure that employs standard errors and a mean-adjusted chi-square test (Muthén & Muthén, 2010). The CFA for closeness produced a

significant chi-square test, $\chi^2$ (5, $N$ = 87) = 12.254, $p$ = .032. In contrast, the two descriptive fit indices showed a good fit: CFI = .951, SRMR = .045. Overall, then, we found reasonable evidence for an acceptable model fit. Factor loadings ranged from .49 ≤ λ ≤ .84 and were all significant (*p*s < .001). The CFA for trust indicated a good model fit: $\chi^2$ (2, $N$ = 87) = 0.419, $p$ = .811, CFI = 1.000, SRMR = .012. Factor loadings ranged from .49 ≤ λ ≤ .81 and were all significant (*p*s < .001). The CFA for perceived social support indicated a good model fit: $\chi^2$ (2, $N$ = 87) = 2.468, $p$ = .291, CFI = .994, SRMR = .023. Factor loadings ranged from .42 ≤ λ ≤ .77 and were all significant (*p*s < .001 for items one, two, and three, and $p$ = .010 for item four). Table 1 provides a summary of all factor loadings per construct. Overall, the good fit of the CFA models and the significant positive loadings of the items on their corresponding factor demonstrated the factorial validity of our measures of closeness, trust, and perceived social support (see Byrne, 2010; Prudon, 2015).

**Table 1.** Factor loadings

| | Closeness | Trust | Perceived social support |
|---|---|---|---|
| Closeness1 | .71 | | |
| Closeness2 | .49 | | |
| Closeness3 | .76 | | |
| Closeness4 | .84* | | |
| Closeness5 | .76 | | |
| Trust1 | | .49 | |
| Trust2 | | .80 | |
| Trust3 | | .79* | |
| Trust4 | | .81 | |
| Support1 | | | .57 |
| Support2 | | | .77* |
| Support3 | | | .75 |
| Support4 | | | .42 |

*Notes.*
* Indicators with an * were used as marked indicators in the CFA. Only factor loadings over .30 are displayed.

## Multidimensional confirmatory factor analysis

To further assess the factorial validity of the results from the unidimensional CFAs, we conducted a multidimensional CFA with a simple structure, in which the items of closeness, trust, and perceived social support are only allowed to load on their corresponding factor. If this procedure reveals a good fit to the

data, the measures represent distinct concepts. In contrast, if the items do not represent distinct concepts, cross-loadings would occur and, consequently, the proposed simple structure of factor loadings would result in a bad model fit. The combined CFA indicated good model fit: $\chi^2$ (62, $N$=87)=62.277, $p$=.466, CFI=.999, SRMR=.052. This further confirmed that our item sets represent distinct concepts. Closeness was positively correlated with trust ($r$=0.73) and perceived social support ($r$=0.77). Trust correlated positively with perceived social support ($r$=0.67).

## Reliability analysis and item fit

Our five-item scale of children's closeness to robots had good internal consistency ($\alpha$=.84). The items were averaged to create a total score. The mean of the total score of closeness was 4.14 ($SD$=0.69). The distribution of the total score was close to normal: The skewness was −0.42 and kurtosis was −1.05. Likewise, our four-item scale of children's trust in robots had good internal consistency ($\alpha$=.86). The total score of trust had a mean of 4.60 ($SD$=0.54). The total score was negatively skewed (−1.86) with positive kurtosis (4.45). Our four-item scale of children's perceived social support had acceptable internal consistency ($\alpha$=.71). The mean was 4.32 ($SD$=0.59). The distribution of the total score only marginally deviated from the normal distribution: The skewness was −.58 and kurtosis was −0.63. In conclusion, our reliability analyses showed that the three measures had acceptable to good internal consistency.

Furthermore, we inspected the item fit by computing difficulty indices as well as corrected item-total correlations, which reflect item selectivity (see Table 2). The difficulty index ranged from 0.71 to 0.85 for closeness, from 0.85 to 0.94 for trust, and from 0.80 to 0.87 for perceived social support. The high difficulty values indicate that children generally experienced high closeness, had high trust in the robot, and perceived strong social support. The corrected item-total correlations for closeness and trust items ranged from .46 to .73 and from .69 to .74 respectively. The correlations for perceived social support ranged from .34 to .63.

## Concurrent validity

To evaluate the concurrent validity of our newly developed scales, we estimated their bivariate correlations with related concepts (i.e., whether the total score of the relevant concept was correlated with the total scores of related concepts). Closeness to the robot significantly correlated with children's agreeableness ($r$=.39, $p$<.001), attachment ($r$=.58, $p$<.001), perceived similarity ($r$=.28, $p$=.008), and fear ($r$=−.23, $p$=.032). The correlation between closeness and social

**Table 2.** Item difficulty & selectivity

| | Item difficulty | | Item Selectivity |
|---|---|---|---|
| | *M* | Difficulty index | Corrected item-total correlation |
| Closeness1 | 4.13 | 0.78 | .64 |
| Closeness2 | 4.38 | 0.85 | .46 |
| Closeness3 | 4.10 | 0.78 | .66 |
| Closeness4 | 3.85 | 0.71 | .73 |
| Closeness5 | 4.26 | 0.82 | .71 |
| Trust1 | 4.46 | 0.85 | .72 |
| Trust2 | 4.53 | 0.88 | .72 |
| Trust3 | 4.68 | 0.92 | .74 |
| Trust4 | 4.74 | 0.94 | .69 |
| Support1 | 4.18 | 0.80 | .47 |
| Support2 | 4.39 | 0.85 | .63 |
| Support3 | 4.21 | 0.80 | .59 |
| Support4 | 4.48 | 0.87 | .34 |

anxiety was not significant ($r = -.10$, $p = .363$). Trust in the robot significantly correlated with children's agreeableness ($r = .47$, $p < .001$), attachment ($r = .43$, $p < .001$), perceived similarity ($r = .26$, $p = .014$), and fear ($r = -.28$, $p = .010$), but not with social anxiety ($r = -.19$, $p = .073$). Children's perceived social support correlated significantly with agreeableness ($r = .41$, $p < .001$), attachment ($r = .46$, $p < .001$), and perceived similarity ($r = .41$, $p < .001$). The correlations of perceived social support with social anxiety ($r = -.03$, $p = .767$) and fear ($r = .04$, $p = .750$) were not significant. Overall, our newly developed measures correlated with our validation measures in the expected way. This provides evidence for the concurrent validity of our measures of children's closeness to, trust in, and perceived social support from a robot.

## Discussion

The present paper aimed at the development and validation of three self-report scales that can be used to assess child-robot relationship formation. Overall, our findings support the validity and reliability of our new measures of closeness, trust, and perceived social support, which may therefore be used in future studies that aim to investigate these constructs. As our measures are relatively short compared to the existing measures they were based upon, they can more easily be

applied by scholars interested in child-robot relationship formation. Moreover, our results confirm that the measures, although mainly based upon measures of interpersonal relationship formation, are applicable to the context of CRI.

Still, there are several points for improvement. First, we opted for a theory-driven, top-down approach, without first performing exploratory analyses with a larger pool of items. We did so for feasibility reasons: Having children meaningfully answer questions about a social robot presupposes that they have interacted with one, which in turn requires a lot of time and effort. Moreover, we ensured the face validity and child-appropriateness of the items in an extensive process, in which each newly developed item was repeatedly discussed and iteratively refined. Still, it needs to be emphasized that an additional exploratory step would have allowed for the analysis of the psychometric properties of a wider variety of items, which may have led to items superior to the ones we used in this study. Relatedly, exploratory analyses would have allowed for an empirical test of whether indicators included in pre-existing scales, such as the URCS (Dibble et al., 2012), are indeed, as we assumed, not adequate for use in CRI research.

Second, some results of the CFAs were suboptimal. The chi-square test for the CFA of closeness was significant. However, the high CFI and SRMR values of the same model as well as the good values of all three fit indices in the multidimensional CFA led us to retain the theoretically expected model of closeness. Moreover, we observed variations in the sizes of the factor loadings. Generally, the factor loadings of perceived social support were lower than those of closeness and trust. This is also reflected in the lower corrected item-total correlations for the concept of perceived social support. One explanation for the lower loadings of perceived social support items may be that they consisted of conditional statements, which are needed to make the concept of perceived social support applicable to any interaction scenario. Even though children in our age range should be capable of judging such statements, reasoning about conditionals requires advanced thinking (e.g., Janveau-Brennan & Markovits, 1999). The hypothetical nature of the perceived social support items may thus have increased error variance, which in turn may have reduced factor loadings.

Third, whereas most validation patterns were as expected, social anxiety did not correlate significantly with either closeness, trust, or perceived social support. Possibly, socially anxious children may not differ in the degree of closeness, trust, and perceived social support they experience when interacting with a robot, because the sociability of robots has still not reached the level of humans'. Indeed, research has shown that children with Autism Spectrum Disorder (ASD), who often struggle with social interactions, appreciate the simplicity and predictability of interactions with robots (e.g., Broadbent, 2017). Additionally, fear did not significantly correlate with perceived social support, which may

be caused by the fact that children experienced almost no fear at all ($M = 1.41$, $SD = 0.55$). This floor effect may have reduced the variance of fear and thus may have precluded a more substantial correlation of fear with perceived social support. Presumably, low intensity emotional states, such as being nervous as a result of meeting somebody new, better characterize interactions between children and robots, and should be included in future research.

Fourth, the substantial correlations between closeness, trust, and perceived social support – and particularly between closeness and trust on the one hand, and closeness and perceived social support on the other – signal the interrelated status of these concepts. From a theoretical perspective, these correlations make sense. Increasing closeness means increasing interdependence, and "without an accompanying growth of trust in the partner, it is unlikely that [a] relationship will proceed to high levels of interdependence" (Berscheid & Regan, p. 209). Moreover, although someone's social network does not necessarily have to be identical to someone's support network (Berscheid & Regan, 2005), it is likely that the two will overlap. As a result, closeness and perceived social support may overlap conceptually and correlate statistically. At the same time, the multidimensional CFA clearly showed that the items cluster into clean factors, such that the conceptual distinction between the constructs was confirmed.

Next to these shortcomings of the study outcomes, several more general limitations of our study can be identified. First, even though our new measurement instruments generally proved to be both valid and reliable – and are thus appropriate for use with children – using self-report measures with children is known to create various challenges. For instance, children's answers to questionnaire items are often subject to social desirability bias, which can lead to ceiling effects (e.g., Belpaeme et al., 2013; Eyssel, 2017). This issue calls for the triangulation of, for instance, self-report, observational, and psychophysiological measures (e.g., Bethel & Murphy, 2010). To date, however, the interpretation of observational cues often varies across CRI and HRI studies, which hinders the comparability of findings (e.g., Andrés et al., 2015; Van Straten et al., 2019). Given the equivocal and interdependent nature of behavioral cues (Cappella, 1983; Knapp, 1983), we need observational measures for CRI that take into account multiple behavioral cues and weigh these cues according to their relevance to the concept under investigation. It would be valuable for future studies to employ valid behavioral measures to gain further insight into, among other things, children's closeness to, trust in, and perceived social support from robots. Moreover, combining multiple, valid approaches would be promising. The development of validated self-report measures constitutes a first step into this direction.

Second, while we carefully consulted the literature to identify key concepts of interpersonal relationships, there may be other concepts that could be included

in the future investigation of child-robot relationships. Specifically, there may be constructs that are inherently linked to child-robot relationship formation that could be considered. For instance, the more qualitative data resulting from this study (i.e., children's answers to the open question why they [dis]trusted the robot) revealed that children refer to both interpersonal and technological aspects of the robot and its behavior when they explain their level of trust in the robot (see Van Straten et al., 2018). Whereas our measure of trust focused on interpersonal trust, it may thus also be sensible to assess children's technological trust in a robot and see how it correlates with interpersonal trust, closeness, and perceived social support. All in all, a top-down approach rooted in human psychology should be combined with a bottom-up approach which considers the qualitative distinctness of psychological states that emerge in interactions with this new technology (see Krämer, Rosenthal-Von der Pütten, & Hoffman, 2015 for a similar claim in the context of increasing artificial agents' humanlikeness).

Third, although our measures worked properly after a single robot encounter, it still has to be established whether they also work in longitudinal settings. For example, the high difficulty indices reported in Table 1 may reflect a novelty effect, which will wear off after multiple encounters. Due to the novelty of the robot, children's answering patterns may generally have been positive. In turn, this rather uniform answering pattern may have resulted in high item difficulty values, indicating easy-to-answer items. However, in the course of multiple interactions with a social robot, children's answers may become more diverse, which is likely to translate into somewhat lower item difficulty values. The development of CRI over multiple encounters may be particularly important for closeness. As noted by Berscheid and Regan (2005), closeness may at a certain point become stable, such that it may become more informative to assess attachment, which is more concerned with relationship maintenance than formation.

Fourth and finally, the results of the multi-dimensional CFA should be interpreted as preliminary evidence because this analysis estimates a large number of parameters based on a relatively small sample size. Although the findings of the multi-dimensional CFA were in line with our expectations, more evidence from replications seems necessary to corroborate the multidimensional structure of the three measures. Currently, the results of the unidimensional CFAs are more robust.

Apart from the suggestions for future research that can be derived from these specific shortcomings, it would be valuable for future studies to elucidate the generalizability of our measures to different CRI contexts. For instance, it should be assessed whether our measures can equally be used with children from different age ranges. For younger children, abstracts concepts such as trust and some specific item formulations may be too complex. For this age group, implicit mea-

sures or situation-based items might be preferable. Moreover, as closeness and trust play a less prominent role in children's friendships before middle childhood, these concepts may not work properly with younger children. Therefore, we are currently hesitant to generalize the applicability of our scales to use with children from different age groups.

It would also be useful to investigate whether the measures that we tested among developmentally typical children are appropriate for use with developmentally atypical children whose tendencies to form social relationships may deviate from the norm, such as children with ASD. The validation of our measures with different types of robots would further shed light on their generalizability across research contexts. For example, human-human bonding differs from human-animal bonding (Kidd & Kidd, 1987), such that it would be valuable to investigate to what extent our measures work when a zoomorphic robot is used. Finally, it would be valuable to test whether our measures yield different outcomes in different, less-controlled research contexts. When the interaction with the robot is less standardized than in our study, child-robot relationship formation may differ accordingly.

In conclusion, the present study contributes to CRI research in providing child-appropriate measures that can be used individually to gain insight into the manifestation of a specific concept, or together to obtain a more comprehensive understanding of the emergence of child-robot relationships. The scales may help researchers as well as practitioners (e.g., teachers, healthcare professionals) to investigate how the introduction of social robots into our daily environments (Kahn et al., 2013) can affect children's psychological and social responses.

## Acknowledgements

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, under Grant 682733 to the third author.

## References

Ahmad, M. I., Mubin, O., & Orlando, J. (2016). Effect of different adaptations by a robot on children's long-term engagement: An exploratory study. *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, 31–37. https://doi.org/10.1145/3001773.3001803

Andrés, A., Pardo, D.E., Díaz, M., & Angulo, C. (2015). New instrumentation for human robot interaction assessment based on observational methods. *Journal of Ambient Intelligence and Smart Environments*, *7*(4), 397–413. https://doi.org/10.3233/AIS-150331

Armsden, G.C., & Greenberg, M.T. (1987). The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, *16*(5), 427–454. https://doi.org/10.1007/BF02202939

Aron, A., Aron, E.N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality & Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Barbaranelli, C., Caprara, G.V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences*, *34*, 645–664. https://doi.org/10.1016/S0191-8869(02)00051-X

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1):71–81. https://doi.org/10.1007/s12369-008-0001-3

Bauminger-Zviely, N., & Agam-Ben-Artzi, G. (2014). Young friendship in HFASD and typical development: Friend versus non-friend comparisons. *Journal of Autism and Developmental Disorders*, *44*(7), 1733–1748. https://doi.org/10.1007/s10803-014-2052-7

Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., & Belpaeme, T. (2016). From characterising three years of HRI to methodology and reporting recommendations. *Proceedings of the Eleventh International Conference on Human-Robot Interaction*, 391–398. https://doi.org/10.1109/HRI.2016.7451777

Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., ... & Zelati, M.C. (2013). Child-robot interaction: Perspectives and challenges. *Proceedings of the International Conference on Social Robotics*, 452–459. https://doi.org/10.1007/978-3-319-02675-6_45

Bernath, M.S., & Feshbach, N.D. (1995). Children's trust: Theory, assessment, development, and research directions. *Applied & Preventive Psychology*, *4*(1), 1–19. https://doi.org/10.1016/S0962-1849(05)80048-4

Berscheid, E., & Regan, P. (2005). *The psychology of interpersonal relationships*. Upper Saddle River, NJ: Pearson.

Bethel, C.L., Henkel, Z., Stives, K., May, D.C., Eakin, D.K., Pilkinton, M., ... Stubbs-Richardson, M. (2016). Using robots to interview children about bullying: Lessons learned from an exploratory study. *Proceedings of the 25th International Symposium on Robot and Human Interactive Communication*, 712–717. https://doi.org/10.1109/ROMAN.2016.7745197

Bethel, C.L., & Murphy, R.R. (2010). Review of human studies methods in HRI and recommendations. *International Journal of Social Robotics*, *2*(4), 347–359. https://doi.org/10.1007/s12369-010-0064-9

Betts, L.R., Rotenberg, K.J., & Trueman, M. (2009). The early childhood generalized trust belief scale. *Early Childhood Research Quarterly*, *24*(2), 175–185. https://doi.org/10.1016/j.ecresq.2008.10.002

Bigelow, B.J. (1977). Children's friendship expectations: A cognitive-developmental study. *Child Development*, *28*(1), 246–253. https://doi.org/10.2307/1128905

Borenstein, J., & Arkin, R.C. (2016). Robots, ethics, and intimacy: The need for scientific research. *Proceedings of the Conference of the International Association for Computing and Philosophy*. https://doi.org/10.1007/978-3-030-01800-9_16

Borgers, N., De Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique*, *66*(1), 60–75. https://doi.org/10.1177/075910630006600106

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(9), 627–652. https://doi.org/10.1146/annurev-psych-010416-043958

Byrne, D.E. (1971). *The attraction paradigm*. New York, NY: Academic Press.

Byrne, B.M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Oxford, UK: Routledge.

Byrne, B.M. (2011). *Structural equation modeling with Mplus. Basic concepts, applications, and programming*. New York, NY: Routledge.

Cappella, J.N. (1983). Conversational involvement: Approaching and avoiding others. In J.M. Wiemann & R.P. Harrison (Eds.), *Nonverbal interaction* (pp. 113–148). Beverly Hills, CA: Sage.

Cole, M., Cole, S., & Lightfoot, C. (2005). *The development of children* (5th ed.). New York, NY: Worth.

De Jong, C., Kühne, R., Peter, J., Van Straten, C.L., & Barco, A. (2018). Intentional acceptance of social robots: Development and validation of a self-report measure for children. *Manuscript submitted for publication*.

De Jong, C., Peter, J., Kühne, R., & Barco, A. (in press). Children's acceptance of social robots: A narrative review of the research 2000–2017. *Interaction Studies*.

De Jong, C., Van Straten, C.L., Peter, J., Kühne, R., & Barco, A. (2018). Children and social robots: Inventory of measures for CRI research. *Proceedings of the International Workshop on Advanced Robotics and its Social Impact*, 44–45. https://doi.org/10.1109/ARSO.2018.8625764

Díaz, M., Nuño, N., Saez-Pons, J., Pardo, D.E., & Angulo, C. (2011). Building up child-robot relationship for therapeutic purposes: From initial attraction towards long-term social engagement. *Proceedings of the Eleventh International Conference on Automatic Face & Gesture Recognition*, 927–932. https://doi.org/10.1109/FG.2011.5771375

Dibble, J.L., Levine, T.R., & Park, H.S. (2012). The Unidimensional Relationship Closeness Scale (URCS): Reliability and validity evidence for a new measure of relationship closeness. *Psychological Assessment*, *24*(3), 565–572. https://doi.org/10.1037/a0026265

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, *46*(1), 84–90. https://doi.org/10.1111/j.1465-7295.2007.00082.x

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Eyssel, F. (2017). An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems*, *87*, 363–371. https://doi.org/10.1016/j.robot.2016.08.029

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037//1082-989X.4.3.272

Furman, W., & Bierman, K.L. (1984). Children's conceptions of friendship: A multimethod study of developmental changes. *Developmental Psychology*, *20*(5), 925–931. https://doi.org/10.1037//0012-1649.20.5.925

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Padstow, UK: T. J. Press.

Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology*, *41*(4), 235–284. https://doi.org/10.1016/S0022-4405(03)00048-7

Goodboy, A. K., & Booth-Butterfield, M. (2009). Individual differences in romantic partners' desire for less closeness. *North American Journal of Psychology*, *11*(1), 209–218.

Gordon-Hollingsworth, A. T., Thompson, J. E., Geary, M. A., Schexnaildre, M. A., Lai, B. S., & Kelley, M. L. (2016). Social Support Questionnaire for Children: Development and initial validation. *Measurement and Evaluation in Counseling and Development*, *49*(2), 122–144. https://doi.org/10.1177/0748175615596780

Guerrero, L. K. (1997). Nonverbal involvement across interactions with same-sex friends, opposite-sex friends and romantic partners: Consistency or change? *Journal of Social and Personal Relationships*, *14*(1), 31–58. https://doi.org/10.1177/0265407597141002

Guneysu, A., & Arnrich, B. (2017). Socially assistive child-robot interaction in physical exercise coaching. *Proceedings of the 26th International Symposium on Robot and Human Interactive Communication*, 670–675. https://doi.org/10.1109/ROMAN.2017.8172375

Hedl, J. J., & Papay, J. P. (1982). The factor structure of the State-Trait Anxiety Inventory for Children: Kindergarten through the fourth grades. *Personality and Individual Differences*, *3*(4), 439–446. https://doi.org/10.1016/0191-8869(82)90008-3

Hieida, C., Abe, K., Attamimi, M., Shimotomai, T., Nagai, T., & Omori, T. (2014). Physical embodied communication between robots and children: An approach for relationship building by holding hands. *Proceedings of the International Conference on Intelligent Robots and Systems*, 3291–3298. https://doi.org/10.1109/IROS.2014.6943020

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

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*(4), 904–911. https://doi.org/10.1037//0012-1649.35.4.904

Kahn, P. H. & Gary, H. E., & Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Developmental Perspectives*, *7*(1), 32–37. https://doi.org/10.1111/cdep.12011\

Kahn, P. H., & Turiel, E. (1988). Children's conceptions of trust in the context of social expectations. *Merrill-Palmer Quarterly*, *34*(4), 403–419.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kidd, A. H., & Kidd, R. M. (1987). Seeking a theory of the human/companion animal bond. *Anthrozoös*, *1*(3), 140–145. https://doi.org/10.2752/089279388787058489

Knapp, M. L. (1983). Dyadic relationship development. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 179–207). Beverly Hills, CA: Sage.

Kory Westlund, J. M., Park, H. W., Williams, R., & Breazeal, C. (2018). Measuring children's long-term relationships with social robots. *Proceedings of the 17th Conference on Interaction Design and Children*. https://doi.org/10.1145/3202185.3202732

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, *23*(14), 1951–1996. https://doi.org/10.1163/016918609X12518783330360

Krämer, N. C., Rosenthal-Von der Pütten, A. M., & Hoffmann, L. (2015). Social effects of virtual and robot companions. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 137–159). Chicester, UK: John Wiley & Sons.

Kruijff-Korbayova, I., Oleari, E., Baroni, I., Kiefer, B., Zelati, M. C., Pozzi, C., & Sanna, A. (2014). Effects of off-activity talk in human-robot interaction with diabetic children. *Proceedings of the 23rd International Symposium on Robot and Human Interactive Communication*, 649–654. https://doi.org/10.1109/ROMAN.2014.6926326

La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology*, *26*(2), 83–94. https://doi.org/10.1016/S0887-6185(98)00008-5

La Greca, A. M., & Stone, W. L. (1993). Social anxiety scale for children-revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology*, *22*(1), 17–27. https://doi.org/10.1207/s15374424jccp2201

Lakey, B., Thomson Ross, L., Butler, C., & Bentley, K. (1996). Making social support judgments: The role of similarity and conscientiousness. *Journal of Social and Clinical Psychology*, *15*(3), 283–304. https://doi.org/10.1521/jscp.1996.15.3.283

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, *42*(3), 595–604. https://doi.org/10.2307/351903

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic robots for long-term interaction: Evaluating social presence, engagement, and perceived support in children. *International Journal of Social Robotics*, *6*(3), 329–341. https://doi.org/10.1007/s12369-014-0227-1

Leite, I., Pereira, A., & Lehman, J. F. (2017). Persistent memory in repeated child-robot conversations. *Proceedings of the Conference on Interaction Design and Children*, 238–247. https://doi.org/10.1145/3078072.3079728

Lin, N. (1986). Conceptualizing social support. In N. Lin, A. D. Walter, & M. Ensel (Eds.), *Social support, life events, and depression* (pp. 17–30). Orlando, FL: Academic Press. https://doi.org/10.1016/B978-0-12-450660-2.50008-2

Looije, R., Van der Zalm, A., Neerincx, M. A., & Beun, R. J. (2012). Help, I need some body: The effect of embodiment on playful learning. *Proceedings of the International Workshop on Robot and Human Interactive Communication*, 718–724. https://doi.org/10.1109/ROMAN.2012.6343836

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37–49. https://doi.org/10.1037//0012-1649.22.1.37

McCroskey, L. L., McCroskey, J. C., & Richmond, V. P. (2006). Analysis and improvement of the measurement of interpersonal attraction and homophily. *Communication Quarterly*, *54*(1), 1–31. https://doi.org/10.1080/01463370500270322

Mehrabian, A. (1972). *Nonverbal communication*. New Brunswick, NJ: AldineTransaction.

Moosbrugger, H., & Kelava, A. (Eds.). (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg, Germany: Springer.

Muris, P., Meesters, C., & Diederen, R. (2005). Psychometric properties of the Big Five Questionnaire for Children (BFQ-C) in a Dutch sample of young adolescents. *Personality and Individual Differences*, *38*(8), 1757–1769. https://doi.org/10.1016/j.paid.2004.11.018

Muris, P., Meesters, C., Van Melick, M., & Zwambag, L. (2001). Self-reported attachment style, attachment quality, and symptoms of anxiety and depression in young adolescents. *Personality and Individual Differences*, *30*(5), 809–818. https://doi.org/10.1016/S0191-8869(00)00074-X

Muthén, L.K., & Muthén, B.O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ognibene, T.C., & Collins, N.L. (1998). Adult attachment styles, perceived social support and coping strategies. *Journal of Social and Personal Relationships*, *15*(3), 323–345. https://doi.org/10.1177/0265407598153002

Ólafsson, K., Livingstone, S., & Haddon, L. (2013). *How to research children and online technologies? Frequently asked questions and best practice*. London, UK: EU Kids Online, LSE.

Pearson, Y., & Borenstein, J. (2014). Creating "companions" for children: The ethics of designing esthetic features for robots. *AI & Society*, *29*(1), 23–31. https://doi.org/10.1007/s00146-012-0431-1

Peter, J., Kühne, R., Barco, A., De Jong, C., & Van Straten, C.L. (2019). Asking today the crucial questions of tomorrow: Social robots and the Internet of Toys. In G. Mascheroni & D. Holloway (Eds.), *The Internet of Toys: Practices, affordances and the political economy of children's smart play* (pp. 25–46). Cham, Switzerland: Palgrave Macmillan (Springer Nature).

Popovic, M., Milne, D., & Barrett, P. (2003). The Scale of Perceived Interpersonal Closeness (PICS). *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *10*(5), 286–301. https://doi.org/10.1002/cpp.375

Prudon, P. (2015). Confirmatory factor analysis: A brief introduction and critique. *Comprehensive Psychology*, *4*. https://doi.org/10.2466/03.CP.4.10

Punch, S. (2002). Research with children. The same or different from research with adults? *Childhood*, *9*(3), 321–341. https://doi.org/10.1177/0907568202009003005

Rammstedt, B., & John, O.P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rapee, R.M. (1997). Perceived threat and perceived control as predictors of the degree of fear in physical and social situations. *Journal of Anxiety Disorders*, *11*(5), 455–461. https://doi.org/10.1016/S0887-6185(97)00022-4

Saint-Aimé, S., Grandgeorge, M., Le Pevedic, B., & Duhaut, D. (2011). Evaluation of Emi interaction with non-disabled children in nursery school using Wizard of Oz technique. *Proceedings of the International Conference on Robotics and Biomimetics*, 1147–1152. https://doi.org/10.1109/ROBIO.2011.6181442

Salter, T., Werry, I., & Michaud, F. (2008). Going into the wild in child-robot interaction studies: Issues in social robotic development. *Intelligent Service Robotics 1*(2), 93–108. https://doi.org/10.1007/s11370-007-0009-9

Schadenberg, B.R., Neerincx, M.A., Cnossen, F., & Looije, R. (2017). Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive Systems Research*, *43*, 222–231. https://doi.org/10.1016/j.cogsys.2016.08.003

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Severson, R. L., & Lemm, K. M. (2016). Kids see human too: Adapting an individual differences measure of anthropomorphism for a child sample. *Journal of Cognition and Development*, *17*(1), 122–141. https://doi.org/10.1080/15248372.2014.989445

Spielberger, C. D. (1973). *State-Trait Anxiety Inventory for Children (STAIC): Preliminary Manual*. Palo Alto, CA: Consulting Psychologists Press.

Sternberg, R. J. (1987). Liking versus loving: A comparative evaluation of theories. *Psychological Bulletin*, *102*(3), 331–343. https://doi.org/10.1037/0033-2909.102.3.331

Sternberg, R. J. (1997). Construct validation of a triangular love scale. *European Journal of Social Psychology*, *27*(3), 313–335. https://doi.org/10.1002/(SICI)1099-0992(199705)27:3<313::AID-EJSP824>3.3.CO;2-W

Swickert, R. J., Hittner, J. B., & Foster, A. (2010). Big Five traits interact to predict perceived social support. *Personality and Individual Differences*, *48*(6), 736–741. https://doi.org/10.1016/j.paid.2010.01.018

Valkenburg, P. M., & Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology*, *43*(2), 267–277. https://doi.org/10.1037/0012-1649.43.2.267

Van Straten, C. L., Peter, J., & Kühne, R. (2019). Child-robot relationship formation: A narrative review of empirical research. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00569-0

Van Straten, C. L., Peter, J., Kühne, R., De Jong, C., & Barco, A. (2018). Technological and interpersonal trust in child-robot interaction: An exploratory study. *Proceedings of the Sixth International Conference on Human-Agent Interaction*, 253–259. https://doi.org/10.1145/3284432.3284440

Vernberg, E. M., Abwender, D. A., Ewell, K. K., & Beery, S. H. (1992). Social anxiety and peer relationships in early adolescence: A prospective analysis. *Journal of Clinical Child Psychology*, *21*(2), 189–196. https://doi.org/10.1207/s15374424jccp2102_11

Vogt, P., De Haas, M., De Jong, C., Baxter, P. & Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Frontiers in Human Neuroscience*, *11*(73). https://doi.org/10.3389/fnhum.2017.00073

Wu, P.-C., Foo, M.-D., & Turban, D. B. (2008). The role of personality in relationship closeness, developer assistance, and career success. *Journal of Vocational Behavior*, *73*(3), 440–448. https://doi.org/10.1016/j.jvb.2008.08.005

## Appendix A. Indicators of closeness, trust, and perceived social support

**Table A1.** Closeness

| Item | Dutch | Backtranslations into English |
|---|---|---|
| 1 | Nao is een vriendje. | Nao is a friend. |
| 2 | Ik voel me op mijn gemak als ik met Nao ben. | I feel comfortable around Nao. |
| 3 | Nao en ik zijn vriendjes aan het worden. | Nao and I are becoming friends. |
| 4 | Nao en ik passen goed bij elkaar. | Nao and I are a good match. |
| 5 | Nao voelt als een vriendje voor mij. | Nao feels like a friend to me. |

**Table A2.** Trust

| Item | Dutch | Backtranslations into English |
|---|---|---|
| 1 | Ik heb het gevoel dat ik Nao kan vertrouwen. | I feel that I can trust Nao. |
| 2 | Ik heb het gevoel dat Nao een geheim van mij kan bewaren. | I feel that Nao can keep one of my secrets. |
| 3 | Ik heb het gevoel dat Nao eerlijk is. | I feel that Nao is honest. |
| 4 | Ik heb het gevoel dat te vertrouwen is. | I feel that Nao is trustworthy. |

**Table A3.** Perceived social support

| Item | Dutch | Backtranslations into English |
|---|---|---|
| 1 | Als ik in de problemen zat zou ik op Nao kunnen rekenen. | If I were in trouble I could rely on Nao. |
| 2 | Als ik in de problemen zat zou Nao mij willen helpen. | If I were in trouble Nao would be willing to help me. |
| 3 | Als ik in de problemen zat zou Nao voor mij opkomen. | If I were in trouble Nao would stand up for me. |
| 4 | Als ik in de problemen zat zou Nao mij opvrolijken. | If I were in trouble Nao would cheer me up. |

## Appendix B. Answering scale

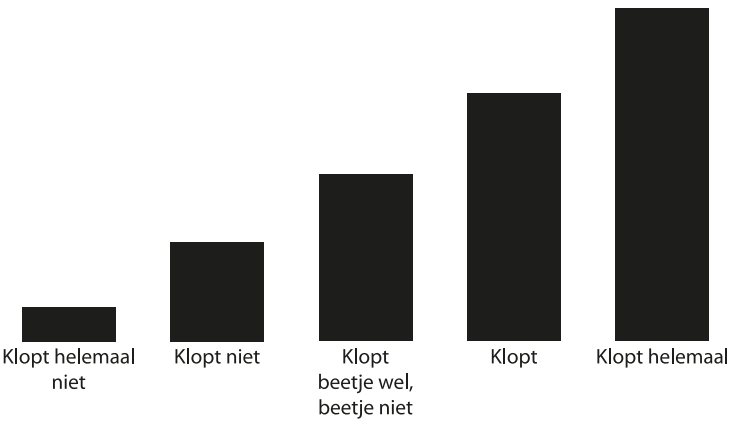


**Figure B1.** Answering scale Dutch labels (adapted from Severson & Lemm, 2016)

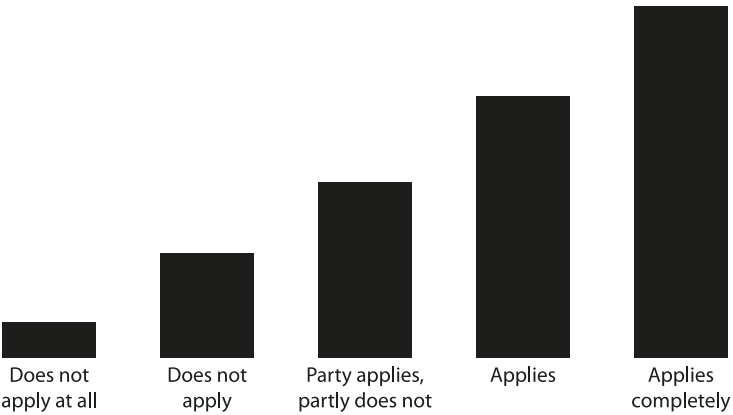


**Figure B2.** Answering scale English labels (adapted from Severson & Lemm, 2016)

## Address for correspondence

Caroline L. van Straten
Amsterdam School of Communication Research (ASCoR)
University of Amsterdam
P.O. box 15791
1001 NG Amsterdam
The Netherlands
c.l.vanstraten@uva.nl

## Biographical notes

**Caroline L. van Straten** (MSc, 2016, Utrecht University) is a PhD candidate at the Amsterdam School of Communication Research.

**Rinaldo Kühne** (Ph.D., 2015, University of Zürich) is an Assistant Professor at the Amsterdam School of Communication Research.
r.j.kuhne@uva.nl

**Jochen Peter** (Ph.D., 2003, University of Amsterdam) is a Full Professor at the Amsterdam School of Communication Research.
j.peter@uva.nl

**Chiara de Jong** (MA, 2017, Tilburg University) is a PhD candidate at the Amsterdam School of Communication Research.
c.deJong@uva.nl

**Alex Barco** (Ph.D., 2017, La Salle – Ramon Llul University) is a Postdoctoral Researcher at the Amsterdam School of Communication Research.
a.barcomartelo@uva.nl