Intentional acceptance of social robots: Development and validation of a self-report measure for children

de Jong, C.; Kühne, R.; Peter, J.; van Straten, C.L.; Barco, A.

DOI
10.1016/j.ijhcs.2020.102426

Publication date
2020

Document Version
Final published version

Published in
International Journal of Human-Computer Studies

License
Article 25fa Dutch Copyright Act

International Journal of Human-Computer Studies 139 (2020) 102426

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Intentional acceptance of social robots: Development and validation of a self-report measure for children

Chiara de Jong*, Rinaldo Kühne, Jochen Peter, Caroline L. van Straten, Alex Barco

Amsterdam School of Communication Research (ASCoR), University of Amsterdam· P.O. box 15791, Amsterdam 1001 NG, the Netherlands

ARTICLE INFO

Keywords:
Intentional acceptance
Social robots
Child-robot interaction
Human-robot interaction
Technology acceptance
Intention to use

ABSTRACT

As the field of child-robot interaction (CRI) is still young, little consensus exists on the conceptual and methodological approach to the study of children's intentional acceptance of social robots, an important antecedent and an outcome of CRI. Against this background, the goal of this study was to develop and validate a self-report measure of intentional social robot acceptance (i.e., intention to use) for children. Partly based on measures for adults, we developed a four-item scale for children's intentional acceptance of social robots. We administered the measure along with the validation measures (i.e., enjoyment, social presence, and social anxiety) among 87 Dutch children aged 7 to 11 years. Our measure reliably and validly assessed children's intentional acceptance of social robots. As children are increasingly likely to encounter social robots in their lives, our measure may help to improve the study of children's acceptance of social robots and its pertinent processes.

1. Introduction

The field of child-robot interaction (CRI), and in particular research on children's acceptance of social robots, is still in its infancy. As a result, little consensus exists on the conceptual and methodological approach to the study of children's acceptance of social robots. A narrative review of 34 articles published between 2000–2017 (De Jong et al., 2019), for example, found that studies differ greatly in their theoretical conceptualization and operationalization of robot acceptance (see also, in a broader context of CRI and HRI research Baxter et al., 2016; Van Straten et al., 2019). Due to this heterogeneity, the comparability across studies is limited and consistent results about children's acceptance of social robots and its antecedents are largely lacking. Consequently, we know little about why children accept, or reject, social robots in the first place.

Research on technology acceptance suggests that two aspects of acceptance can be differentiated (e.g., Davis, 1989; Heerink et al., 2010; Venkatesh et al., 2003): the intention to use a technology repeatedly – *intentional acceptance* – and the actual repeated or longitudinal use of a technology – *behavioral acceptance*. In CRI research, operationalizations that refer to the intentional or behavioral acceptance of social robots are prevalent (De Jong et al., 2019). However, the theoretical distinction between the two forms of acceptance is not well-established in CRI. Moreover, operationalizations are often not linked to theoretical definitions (De Jong et al., 2019).

The suboptimal state of affairs in research on children's acceptance of social robots is particularly problematic in the case of children's intentional acceptance of social robots. A theoretically grounded and validated measure of the concept is required for theoretical as well as methodological reasons. First, intentional acceptance is regarded as a crucial antecedent and outcome of CRI. On the one hand, intentions are generally regarded as pivotal predictors of behavior (Ajzen, 1991; Davis, 1989; Venkatesh et al., 2003). Accordingly, if children do not have the intention to use a robot, there is little to no chance that there will be a successful CRI. On the other hand, children's interactions with robots feed back into behavioral intentions. Positive experiences with robots, in particular, increase children's willingness to have continued or repeated interactions (see for example Kędzierski et al., 2013; Kruijff-Korbayová et al., 2014), which, in turn, may elicit further effects, such as the formation of social relationships with robots (for an overview see Van Straten et al., 2019). From a theoretical point of view, children's intentional acceptance of social robots is thus key to understand how children behave and interact with social robots.

Second, the validity of ad-hoc measures of children's intentional acceptance of social robots is likely to be limited because the measurement of the concept is methodologically challenging. Because intentional acceptance – in contrast to behavioral acceptance – is a latent concept, it cannot be observed but only indirectly measured through

* Corresponding author.
E-mail addresses: c.dejong@uva.nl (C. de Jong), r.j.kuhne@uva.nl (R. Kühne), j.peter@uva.nl (J. Peter), c.l.vanstraten@uva.nl (C.L. van Straten), a.barcomartelo@uva.nl (A. Barco).

https://doi.org/10.1016/j.ijhcs.2020.102426
Received 26 November 2018; Received in revised form 3 October 2019; Accepted 27 February 2020
Available online 29 February 2020

indicators and, more specifically, through children's self-reports. Without a clear theoretical conceptualization and an operational definition that clarifies the link between indicators and concept, the validity of an ad-hoc measure of children's intentional acceptance of social robots may thus be questionable. Moreover, administering self-reports to children is challenging and a variety of problems can arise (De Leeuw and Otter, 1995; Read and MacFarlane, 2006). Most importantly, children may struggle to understand questions about their intentional acceptance of social robots and, thus, provide incorrect responses, which may give rise to systematic measurement errors (Read and MacFarlane, 2006).

Against this background, the goal of this study is to develop and validate a self-report measure of intentional social robot acceptance for children. We focus on the development of a measure for children in their middle childhood (typically 6 to 12 years old; Cole et al., 2005) as this is the age group for which a self-report measure is most relevant. Children in their early childhood are hardly able to process surveys, as language ability is particularly important when assessing a questionnaire (De Leeuw et al., 2004; Read and MacFarlane, 2006). Conversely, adolescents (i.e., over 12 years old) developmentally converge more with adults and have less difficulty with ambiguous questions (De Leeuw et al., 2004; De Leeuw and Otter, 1995), so that the employment of standard measurement instruments is less problematic with this age group compared to children.

2. Existing measures of intentional robot acceptance

2.1. Children's intentional robot acceptance

The CRI literature has paid little attention to children's intentional acceptance of social robots. Studies that did analyze children's intentional acceptance of social robots or related concepts, have mostly used single-item measures (e.g., Kędzierski et al., 2013), which precludes an assessment of reliability. Among the studies on children's intentional acceptance of social robots, we can distinguish between studies that explicitly deal with the concept of acceptance (e.g., Al-Taee et al., 2016; Kanda et al., 2012; Looije et al., 2017; Saint-Aimé et al., 2011) and studies that do not explicitly focus on the concept, but use measures that correspond in their operationalization to intentional acceptance (e.g., Abe et al., 2012; Breazeal et al., 2016; Ferraz et al., 2016; De Haas et al., 2016). For example, some studies (Abe et al., 2012; Blanson Henkemans et al., 2017; De Haas et al., 2016) asked children whether they want to play again with the robot, but did not link this question to intentional acceptance. This procedure is plausible if studies are not explicitly designed to investigate intentional acceptance. Still, for researchers who consider to use such a measure (possibly because it seems to have a sound face validity), the question arises whether the measure shows, for example, satisfactory criterion validity.

The small number of CRI studies (Al-Taee et al., 2016; Kanda et al., 2012; Looije et al., 2017; Saint-Aimé et al., 2011) that explicitly studied the concept of intentional acceptance have provided meaningful insights into the interaction between children and robots (for an overview see De Jong et al., 2019). However, some studies did not clarify how intentional acceptance was precisely operationalized (Al-Taee et al., 2016; Looije et al., 2017; Saint-Aimé et al., 2011). Moreover, even if a measure contained multiple items, typically neither the reliability (except for Kanda et al., 2012; Kose-Bagci et al., 2009) nor the validity of the measures were assessed. As a result, it has not been demonstrated satisfactorily whether the measurement instruments used in previous studies represent children's intentional acceptance of social robots.

2.2. Adults' intentional robot acceptance

In the field of human-robot interaction (HRI), several scales for the measurement of acceptance of social robots have been validated among adults (Eyssel et al., 2011; Heerink et al., 2010; Shin and Choo, 2011). Notably, the scale by Heerink et al., 2010 has been successfully used with 11- to 12-year-old children and was found to be reliable (Kanda et al., 2012). However, the scale has not been validated among children. The missing validation among children is problematic because children have less advanced language skills and may not have the cognitive capability to correctly understand the indicators included in scales for adults (see, for instance, De Leeuw et al., 2004; Read and MacFarlane, 2006). Moreover, children in middle childhood, especially those in early middle childhood (i.e., 7 to 10), may have difficulty with abstract wording and vague definitions in survey questions (De Leeuw et al., 2004). For example, indicators used in adult scales do not specify how individuals intend to use the robot in the future (e.g., "I plan to use the robot during the next few days"; Heerink et al., 2010, p. 364). Although Kanda et al. (2012) have tried to make the items more child-appropriate by applying them to a specific use-context (i.e., use in a Lego class or in another class), the type of use (e.g., playing with the robot) remains unspecified. As a result, the items used by Kanda et al. (2012) may be too abstract for the use among children in (early) middle childhood, and children's answers are likely to depend on the activity they imagine doing with the robot (Al-Taee et al., 2016; Banthia et al., 2016). All of this may result in systematic measurement error (Carpenter, 2018).

In sum, to our knowledge, no validated scale exists to measure children's intentional acceptance of social robots. Existing scales for adults have not been validated among children and are, due to their abstractness, not suitable for use with children. In contrast, measures used with children lack information about crucial psychometric properties. In this study, we therefore aim to develop a measure of intentional acceptance of social robots for children which fulfills the following requirements: a) it provides a clear and theoretically grounded definition of the concept; b) it provides a valid operationalization of the concept; c) it uses child-appropriate indicators; and d) its reliability and validity among children are empirically established.

2.3. Theoretical definition

In order to develop a reliable and valid scale, we first need a clear, theoretically grounded definition of children's intentional acceptance of social robots. We define children's intentional acceptance of social robots as *children's intention to use a social robot repeatedly and/or on a long-term in their daily life* (see De Graaf et al., 2018). Social robots are robots that are capable of having a social interaction that approaches interpersonal interaction (Broadbent, 2017; Zhao, 2006). A social robot does often have an anthropomorphic appearance, but it can also have an alternative morphology (e.g., zoomorphic or caricatured; Fong et al., 2003).

In the definition of intentional acceptance, we refer to children in middle childhood, that is, the period roughly from 6 to 12 years of age (Cole et al., 2005). A scale for intentional social robot acceptance is particularly relevant for this age group. Due to their limited language skills and cognitive development, younger children may struggle with questionnaires, especially in the case of ambiguous questions (De Leeuw et al., 2004; Read and MacFarlane, 2006), whereas children in adolescence have the cognitive skills to understand and respond better to adult scales than children (De Leeuw et al., 2004; De Leeuw and Otter, 1995).

The specific developmental preferences and needs associated with middle childhood as well as the specific affordances of social robots may give rise to particular uses of social robots and are thus relevant to the conceptualization of intentional social robot acceptance. This is reflected in CRI research that shows that children's acceptance of certain functions of a social robot is related to their age (Al-Taee et al., 2016; Banthia et al., 2016). For example, adolescents are, compared to children, less likely to accept companionship from a social robot (Al-Taee et al., 2016). Compared to adults, children are expected to react differently to robots (Belpaeme et al., 2013). Children focus more on the hedonic aspects (i.e., is it enjoyable?) and functions (i.e., can I play with it?) of a social robot (De Jong et al., 2019), whereas adults typically take into account also utilitarian aspects of the technology, such as the ease of use (e.g., Heerink et al., 2010).

Finally, following De Graaf et al., 2018, our definition of intentional acceptance emphasizes that acceptance refers to the repeated use and the integration of social robots in one's daily life. Looking at the intention to use a robot only once, without including repetitive or longitudinal use, would not fully capture intentional acceptance, not least because empirical research has shown that children have a high initial enthusiasm for robots that wears off after some time (Baxter et al., 2017; Kanda et al., 2004). Accordingly, De Graaf et al., 2018 argue that social robot acceptance "[…] ideally, ends with that individual [the user] embracing the technology and incorporating its use in his or her everyday life." (p. 4). We, therefore, include the intention to incorporate a robot into one's daily life into the conceptualization of social robot acceptance.

3. Method

3.1. Development of indicators

As a starting point, we carefully inspected existing measures for intentional social robot acceptance, both in the HRI and CRI literature. As the scale by Heerink et al. (2010) is the only one to date that has been successfully used with children (see Kanda et al., 2012), and broadly aligns with our theoretical definition of intentional social robot acceptance, we took this measure as a starting point. Next, we adjusted the indicators to make them suitable for children. More specifically, to facilitate children's responses, we followed the approach by Eyssel et al., 2011 and adapted the indicators such that they did not refer to the general intention to use the robot, but to specific activities that children may aim at doing with the robot (see De Leeuw et al., 2004). Additionally, we avoided using negations – as this might be too cognitively demanding for younger children (Marsh, 1986) – and qualifiers (e.g., a bit) – as this might be confusing combined with a gradual answering scale. The formulation of the indicators was iteratively refined through discussions among the authors of this manuscript. Whenever the authors were unsure about whether certain words or grammatical constructions were suitable for the language skills of children, primary school teachers were asked for advice. As we aimed to validate the measures by means of data collected from Dutch primary schools, we translated the items to Dutch and adjusted them where necessary to match Dutch children's language ability. The indicators were then piloted among four children, which led to further adjustments of the indicators in case we encountered any issues. This procedure resulted in the following four items: "I would like to play again with NAO," "I would like to see NAO again," "It would be nice if NAO and I could meet again," and "I would like to take NAO home." (see Appendix 1 for Dutch items).

3.2. Scale validation

The newly developed scale was validated in three steps. First, we conducted a confirmatory factor analysis (CFA) to evaluate the dimensionality of the measure as well as the factorial validity of the scale by inspecting the factor loadings of each indicator on the concept. The identification of a one-dimensional structure with substantive factor loadings would be indicative of factorial validity (Byrne, 2010). In addition, in the CFA we also checked for metric and scalar invariance across boys and girls (Kühne, 2013). Second, to test the reliability of the scale, we estimated Cronbach's alpha (Cronbach, 1951). Third, criterion validity was assessed by inspecting Pearson's correlations between the newly developed scale and a series of validation measures – i.e., measures of concepts which on theoretical and empirical grounds can be assumed to be correlated with children's intentional acceptance of social robots (Carmines and Zeller, 1979). Significant correlations between our new measure and the validation measures, which are neither too low nor too high, suggest that the measures are associated, but not redundant, which indicates criterion validity. In contrast, very small, non-significant correlations suggest that the theoretical relationship does not exist, which is indicative of low criterion validity. Very high correlations suggest that the measures are redundant, which is also indicative of low criterion validity.

3.2.1. Confirmatory factor analysis

The CFA was conducted in Mplus 7.4 (Muthén and Muthén, 2012). We estimated four models. The first model – the full sample model – was estimated in the full sample to assess the dimensionality and factorial validity of the measure. The model fit was assessed by inspecting the chi-square test, the comparative fit index (CFI), and the standardized root mean square residual (SRMR). We did not use the root mean square error of approximation (RMSEA) to assess model fit in our analyses as RMSEA in models with low degrees of freedom and small sample is prone to falsely indicating bad model fit (Kenny et al., 2015). A non-significant result of the chi-square test (Byrne, 2010; Schermelleh-Engel et al., 2003), a CFI larger than 0.95 (Byrne, 2010; Hu and Bentler, 1999), and a SRMR smaller than 0.08 indicate an acceptable model fit (Hu and Bentler, 1999).

Next, we tested whether the scale had stable psychometric properties among boys and girls (i.e., measurement invariance). In that way, we can confirm that the indicators of the latent construct operate in a similar way across different groups (Byrne, 2010). More specifically, we tested whether factor loadings (i.e., metric variance), and intercepts of indicators (i.e., scalar variance) are similar for boys and girls (Byrne, 2010; Kühne, 2013; Steenkamp and Baumgartner, 1998). The second model – the (multi-group) baseline model – was the same as the full sample model, but estimated for boys and girls separately. In the third model – the metric invariance model – we constrained the loadings of like indicators to be equal across the two groups. Finally, in the fourth model – the scalar invariance model – we additionally constrained the intercepts of like indicators to be equal across groups. In order to check metric and scalar invariance, we compared the metric invariance model to the baseline model, and the scalar invariance model to the metric invariance model (Kühne, 2013). If the fit of the model does not decrease (i.e., a non-significant χ^2-test and ΔCFI ≥ -0.005) after adding constraints, we can confirm metric or scalar invariance (Kühne, 2013). For the change in SRMR we used different cutoff values for the test of metric invariance (ΔSRMR $\leq$ 0.025) and the test of scalar invariance (ΔSRMR $\leq$ 0.005), because the SRMR is more sensitive to violations of loading invariance than violations of intercept invariance (Chen, 2007).

3.2.2. Reliability analysis and scale construction

To check the internal consistency of the scale, we estimated Cronbach's alpha in SPSS 25. Next, we separately estimated Cronbach's alpha for boys and girls to inspect whether the reliability is comparable between them. We averaged the indicators to create a total score and assessed its distribution by calculating the range, mean, standard deviation, skewness, and kurtosis.

3.2.3. Assessment of criterion validity

To assess the criterion validity of the scale, we estimated bivariate correlations in SPSS 25 between the scale (i.e., the mean index) and the following concepts: enjoyment, social presence, and social anxiety. The main purpose of hedonic systems, such as social robots, is the experience of fun (De Jong et al., 2019; Van der Heijden, 2004). Thus, more enjoyment during an interaction with a robot is likely to lead to a higher intentional acceptance. Accordingly, empirical research has shown that enjoyment increases both adults' (Heerink et al., 2010; Shin and Choo, 2011) and children's (Kędzierski et al., 2013) willingness to interact (again) with a social robot .

As social presence plays an important role "[…] in mediating technology users' attitudes, evaluations, and social responses toward the technology" (Lee et al., 2006, p. 759), a higher perception of social presence is likely to result in a higher intentional acceptance. Empirically, social presence has been found to both indirectly (Heerink et al., 2010) and directly (Shin and Choo, 2011) predict intentional robot acceptance in adult robot users.

Finally, as social anxiety is a fear of social situations in which an individual is confronted with unfamiliar people or the possible scrutiny of others (Beidel et al., 1995; La Greca and Stone, 1993), children with social anxiety are more likely to withdraw from social situations (La Greca and Stone, 1993). Given that social robots are designed to interact in a social manner with its user, we expect a negative relationship between social anxiety and intentional acceptance of social robots.

3.3. Data collection

The data used in the present study come from a larger study in which we aim to develop and validate a set of standardized measures for CRI research. Data from this larger study are also used in four other publications (De Jong et al., 2019; De Jong et al., 2020; Van Straten et al., 2020; Van Straten et al., 2018). The full study was approved by the Ethics Review Board of the Faculty of Social and Behavioural Sciences at the University of Amsterdam (2018-YME-8706).

3.3.1. Sample

The initial sample consisted of 88 children from two elementary schools in the Netherlands. One child did not finish the interaction and thus no data was collected. The final sample consisted of 87 children between 7 and 11 years old ($M = 9.17$, $SD = 0.85$), with 48 girls and 39 boys.

3.3.2. Procedure

Prior to conducting the study, we received active written consent from schools and parents or guardians, and active verbal consent from children themselves. In order for children to feel comfortable during the interaction (Vogt et al., 2017), we introduced the study at class-level before the CRI sessions took place. Additionally, we emphasized that their participation was voluntary; that no personal information would be published; and that they could withdraw from the study at any point without giving reasons. Children were allowed to ask questions, but robot related questions were postponed until the debriefing to prevent any influence on children's perception of the robot.

The interaction took place in a quiet room in the presence of a female researcher. The child was asked to sit on the floor across from the robot at a distance they felt comfortable with. Before the start of the interaction, children were asked whether they still wanted to participate in the study and were reminded of the possibility to refrain from participation at any point in time without any negative consequences. After the child gave verbal consent for their participation, the interaction was started. If parents gave their consent, the interaction was videotaped. The researcher controlled the robot with a laptop from a small distance by means of Wizard of Oz.

Children interacted with a NAO robot (SoftBank) for an average of 8 min. NAO is a small humanoid robot capable of social interaction. During the interaction, the robot asked the child various questions (e.g., "How old are you", "Is there a game that you like to play? Or a sport?") and showed some of its functions (e.g., telling a joke, dancing). Moreover, the child and the robot played a game of 'True or False'. The robot told the child various statements about its abilities (e.g., "My eyes can change color"), after which the child had to guess whether the statement was true or false. After each guess, the robot provided the correct answer (e.g., "My eyes have lights with different colors. Just have a look!"). To avoid the child from getting bored, the robot also asked the child several questions in-between (e.g., "Do you also like fries, or do you prefer pancakes?"). At the end of the interaction, the robot and the child said goodbye and the child was led to the interview room. During the interaction, the robot never made any statements suggesting that it might have human capabilities in order to prevent children from feeling deceived (Broadbent, 2017).

In the interview room, a second female researcher administered a questionnaire. The interviewer presented the questions to the child orally. The questionnaire included various closed-ended questions on children's perception of the robot, their internal states during the interaction with the robot, their appreciation of the interaction, and on their cognitive development and personality. Additionally, to give children the opportunity to elaborate on some of their answers, several open-ended questions were included. In line with earlier research (Leite et al., 2017), we first presented children with practice items (e.g., "I like candy"), to familiarize them with the item and answering format. Once children fully understood the procedure, the questionnaire was started.

After all children participated, the experimenters debriefed them at class-level (for a similar approach see Wood et al., 2013). Supported by a power-point presentation, the experimenters explained children the mechanical nature of the robot, its workings, and the procedure of programming the robot. Additionally, the Wizard of Oz paradigm was explained and it was emphasized that the robot said and did the same things with every child, to prevent children from feeling deceived. Finally, the experimenters pointed out some differences between humans and robots, and answered any remaining questions children had. All children, also the ones that did not participate, received a small gift to thank them.

3.3.3. Measures

Our scale of intentional acceptance of social robots for children included the aforementioned four items (see Appendix 1 for the Dutch translations). Children had to indicate their agreement or disagreement on a five-point rating scale with the response options ranging from "Does not apply at all" to "Applies completely." A bar chart adapted from Severson and Lemm (2016) (see Appendix 2) was used to visualize the degree of agreement and facilitate children's responding. This bar chart was chosen because it lacked any indications of valence of the items (e.g., smileys, colors) and, as such, avoids the elicitation of socially desirable answers as much as possible. We also asked children for a motivation for their answer to the last of the four items, that is, we asked them why they would/would not like to take the robot home. These data are not analyzed in the present study because they are not directly relevant to the validation of the scale.

We adapted the scale by Oliver and Bartsch (2010) to measure enjoyment. We asked children to indicate their agreement – on the same five-point rating scale as mentioned above – with four items (e.g., "It was fun for me to talk with NAO"). The items were averaged to form a total score for enjoyment during the interaction ($M = 4.78$, $SD = 0.34$, $\alpha = 0.71$). The scale had a skewness of -1.73 ($SE = 0.26$) and kurtosis of 2.58 ($SE = 0.51$).

Social presence was measured by means of a 4-item scale in which two items were adapted from Heerink et al., 2010. Children were asked to state their agreement to four statements (e.g., "When I was with NAO, I felt like I was talking to a person"). Again, children could indicate their agreement to these statements on the five-point rating scale. The items were averaged to form the total score of social presence ($M = 3.86$, $SD = 0.92$, $\alpha = 0.88$) with a skewness of -0.60 ($SE = 0.26$) and kurtosis of -0.20 ($SE = 0.51$).

Finally, social anxiety was measured by means of the six-item Social Avoidance and Distress – Specific to New Peers or Situations sub-scale of the Social Anxiety scale by La Greca and Stone (1993). An adjusted version of the subscale has previously been used by Valkenburg and Peter (2007) to measure social anxiety in adolescents. Children were asked to indicate their agreement - on the same five-point rating scale - with six items (e.g., "I worry about doing something new in front of other kids"). The items were averaged to form the total social anxiety scale ($M = 3.20$, $SD = 0.88$, $\alpha = 0.82$). The skewness was -0.22 ($SE = 0.26$) and the kurtosis -0.22 ($SE = 0.51$).

4. Results

4.1. Confirmatory factor analysis and factorial validity

The CFAs were estimated in Mplus 7.4. As the univariate distributions of the indicators deviated from normality – which implies that the assumption of multivariate normality is violated – we employed the MLM estimator (i.e., maximum likelihood estimation with standard errors and a mean-adjusted chi-square test that are robust to non-normality; Muthén and Muthén, 2012). The CFA in the full sample revealed a good fit of the data: χ^2 (2, $N = 87$) = 3.562, $p = .168$, CFI = 0.978, SRMR = 0.043. Additionally, the analysis showed that all four indicators loaded positively and significantly on the latent construct. The first three indicators had a high loading (respectively $\lambda = 0.77$, $p < .001$; $\lambda = 0.99$, $p < .001$; $\lambda = 0.73$, $p < .001$), whereas the last item had a lower loading ($\lambda = 0.33$, $p = .006$). The overall substantive factor loadings and good model fit are a sign of the measure's factorial validity (Byrne, 2010).

4.2. Tests of measurement invariance

To test for measurement invariance, we first estimated a multi-group baseline model, and, subsequently, two constrained models – a metric invariance model, and a scalar invariance model – which we compared to the baseline model. In the baseline model, the residual variance of the second indicator was negative in the female group (i.e., a Heywood case occurred), which is not uncommon in small samples (Chen et al., 2001; Dillon et al., 1987). Following the recommendations by Chen et al., 2001, we checked whether the Heywood case was the result of sampling fluctuation or of model misspecification before fixing the negative error variance. By means of a two-tailed z-test, we evaluated the alternative hypothesis that the negative error variance is different from zero (i.e., a result of model misspecification) versus the null hypothesis that the negative error variance is not different from zero (i.e., a result of sampling fluctuation).The z-test is equivalent to the Wald test (Chen et al., 2001). Notably, both tests are more liberal or sensitive than the likelihood ratio and Lagrange multiplier test in finite samples (Engle, 1983). Put differently, the z-test and Wald test lead more quickly to a rejection of the null hypothesis, and a non-significant finding thus provides stronger evidence in favor of the null hypothesis (albeit the asymmetry of statistical tests has to be considered; e.g., Hagen, 1997). The z-test showed that the null hypothesis should be accepted ($z = -0.006$, $p = .995$), which means that the negative error variance is most likely not due to model misspecification. Consequently, following the literature (Chen et al., 2001), we fixed the residual variance of the second indicator to 0.01 in the female group. The resulting CFA revealed a good fit of the data: χ^2 (5, $N = 87$) = 6.800, $p = .236$, CFI = 0.979, SRMR = 0.038.

To assess metric invariance, we estimated a constrained model in which we set the loadings of like indicators to be identical across the female and male group. We compared the fit of this model with the fit of the baseline model (see Table 1). MLM estimation required us to use the Satorra-Bentler scaled chi-square difference test (χ^2-TRd) in the comparison of model fit (Mplus, n.d.; Satorra and Bentler, 2010). The results of the model comparison were inconclusive: The Satorra-Bentler scaled chi-square difference test was not significant: χ^2-TRd (3) = 5.673, $p = .129$. However, the changes in the CFI and SRMR indicated a substantial decrease of the fit of the metric invariance model compared to the baseline model (ΔCFI = -0.031, ΔSRMR = 0.088).

As mentioned by Byrne (2010), it is possible – as measures are often group specific – that a measurement instrument does not function completely the same in different groups (see also Byrne et al., 1989). Consequently, the constraints on the loadings may lead to a bad model fit. In such a case, an option is to test a partially invariant measurement model, in which unequal parameters are freely estimated in the different groups (Byrne et al., 1989). Upon inspection of the loadings and modifications indices, we found that the loadings of the third and fourth indicator substantially differed across groups. Consequently, we estimated a partial metric invariance model in which the constraints on the third and fourth indicator were released. The fit of this model was good (see Table 1) and comparable to the fit of the baseline model: χ^2-TRd(1) = 0.615, $p = .433$, ΔCFI = 0.000, ΔSRMR = 0.007). Thus, partial metric invariance can be assumed.

Finally, we estimated the scalar invariance model. This model included the same constraint as the partial metric invariance plus constraints of the indicators' intercepts, which were set to be equal across groups. The test of scalar invariance was inconclusive (see Table 1): The Satorra–Bentler scaled chi-square difference test was not significant: χ^2-TRd(3) = 5.067, $p = .167$. However, the change in CFI and SRMR showed a decrease in model fit (ΔCFI = -0.025, ΔSRMR = 0.081). Upon inspection of the intercepts and modification indices, we found that the intercepts of the same two indicators with unequal loadings (i.e., the indicators three and four) substantially differed across groups. We estimated the partial scalar invariance model by releasing the constraints on the intercepts of the third and fourth indicator. The partial scalar invariance model showed a good fit (see Table 1) and did not substantially decrease in model fit: χ^2-TRd(1) = 0.036, $p = .850$, ΔCFI = 0.006, ΔSRMR = 0.000). Partial scalar invariance could thus

Table 1
Model comparisons: tests of measurement invariance among boys and girls.

Model	χ^2	df	p	SCF*	CFI	SRMR
1 Baseline model, unconstrained	6.800	5	.236	1.821	.979	.038
2 Metric invariance model	12.386	8	.135	1.763	.948	.126
3 Metric invariance model, partial	7.819	6	.252	1.643	.979	.045
4 Scalar invariance model	12.954	9	.165	1.693	.954	.126
5 Scalar invariance model, partial	8.308	7	.306	1.550	.985	.045

* SCF: Scaling correction factor for MLM.

be established. Based on the model with partial scalar invariance, we compared the latent means of boys' (M = 4.75, SE = 0.06) and girls' (M = 4.75, SE = 0.08) intentional acceptance of social robots. The Wald-test showed that there existed no significant difference (W_T = 0.003, p = .953).

4.3. Reliability analysis and scale construction

The internal consistency of the 4-item intentional acceptance of social robots' scale for children was acceptable (overall α = 0.72; boys α = 0.78; girls α = 0.67). When deleting the last item, which had a lower factor loading than the other three items, the alpha improved (overall α = 0.85; boys α = 0.82; girls α = 0.87). Despite the increase in internal consistency, we calculated a total score by averaging the four indicators because the fourth item created more variance in the scale (s^2 = 0.21 for the 4-item scale versus s^2 = 0.17 for the 3-item scale).

An inspection of the distribution of the scale revealed that children's scores ranged from 3.25 to 5.00, with a mean of 4.72 (boys M = 4.72; girls M = 4.71) and a standard deviation of 0.45 (respectively SD = 0.45; 0.46). The skewness was −1.62 and the kurtosis 1.79.

4.4. Criterion validity

To assess the criterion validity of the intentional acceptance of social robots' scale for children, we inspected the bivariate Pearson correlations between the scale and enjoyment, social presence, and social anxiety. In line with our expectations, we found that the scale positively correlated with enjoyment (r = 0.494, p < .001) and social presence (r = 0.244, p = .023). The correlation between children's intentional acceptance of a social robot and their social anxiety was, as expected, negative, but it failed to reach conventional levels of significance (r = −0.204, p = .059). This pattern of results was also confirmed when we applied a p-value correction for multiple comparisons, namely the False Discovery Rate (FDR) by Benjamini and Hochberg (1995). Using an online calculator (Carbocation Corporation, 2016) with a false discovery rate of 0.05, the adjusted p-values for the correlations were p = .003 for the correlation with enjoyment; p = .035 for the correlation with social presence; and p = .059 for the correlation with social anxiety.

5. Discussion

Currently, we still know little about what predicts children's acceptance of social robots. One reason for this is that there is little consensus on the conceptual and methodological approach in the field. Notably, although children's intentional acceptance of social robots is a crucial concept in CRI research, no validated measurement instrument exists to date. Therefore, the main aim of the current paper was to develop and validate such a measurement instrument. Partly based on measures designed for adult populations, we developed a four-item scale of intentional acceptance of social robots for children, which we administered to children aged 7 to 11 years.

The results showed that our new measure reliably and validly assesses children's intentional acceptance of social robots. However, one indicator (i.e., "I would like to take NAO home") had a relatively low loading. Removing this indicator would improve the reliability of the scale. An explanation for this relatively low loading could be that the indicator centered on using the robot in children's personal environment, whereas the other indicators did not specify the location of use. As the current study involved only one interaction between the child and the robot, taking the robot home might have been a step too far for some children. Moreover, children's open answers on a follow-up question in the survey suggested that they also considered their family members and external constraints – such as time, space and costs – when determining their willingness to take the robot home (De Jong et al., 2020). This potentially led the indicator to correlate less strongly with the other indicators and reduced its factor loading. Nevertheless, incorporation of the robot in one's daily life is conceptually an important aspect of acceptance, which distinguishes acceptance from mere use of the robot (De Graaf et al., 2018). As the fourth item still loaded significantly on the latent construct and created more variance in the scale, we recommend that the item should be included initially in the scale to validly represent the concept of intentional acceptance of social robots.

To test whether the psychometric properties (i.e., the factor loadings and intercepts of indicators) are stable across boys and girls, we assessed measurement invariance. Although the multi-group baseline model showed a good fit to the data, we were not able to establish full metric and scalar invariance. Still, we were able to establish partial metric and scalar invariance. According to the literature (Byrne et al., 1989; Meredith, 1993), comparisons across groups are still possible as long as at least two indicators (i.e., the marker plus an additional indicator) are invariant. In our case, even though not optimal, the comparison between boys and girls is thus still admissible. As the reliability of the measure was also slightly lower for girls, future research should address the partial discrepancy in measurement functioning between boys and girls, also in the long run, as this might lead to a slightly different functioning measure in both groups.

The scale had an acceptable reliability. Consequently, we constructed a total score by averaging the scores of the four indicators. Descriptive statistics revealed that the total score had a high mean and that there were moderate deviations from normality (i.e., the skewness was negative and the kurtosis positive). These findings are likely the result of a ceiling effect of children's first interaction with a robot. Previous research has demonstrated that children typically form very positive impressions in first-time interactions with social robots because of the novelty effect, leading to ceiling effects (e.g., Belpaeme et al., 2013; Kanda et al., 2004; Sung et al., 2009). However, despite the high mean, moderate non-normality of the data and the small sample size, the expected correlations between intentional acceptance of social robots and two of the three validation measures were still found, showing criterion validity. Children who experienced more enjoyment during the interaction and more social presence of the robot also had a higher intention to accept the social robot.

There was also a negative correlation between social anxiety and children's intentional acceptance of the social robot (Pearson's r = −0.204), which failed to reach conventional levels of significance (p = .059). The absence of a significant correlation probably results from the relatively small sample size. Against this background, we consider the correlations reasonable and the analyses to overall confirm the criterion validity of the scale.

These correlations between intentional acceptance and the validation measures indicate that, even after first-time interactions between children and a social robot, there is enough variance in our measure to identify relationships between intentional acceptance of social robots and other theoretically related concepts. Presumably, the measure might function even better in longitudinal research, as the novelty effect is likely to wear off after some time. However, future research should confirm the reliability and validity of the scale in a longitudinal perspective.

5.1. Limitations and future research

Our study has at least three limitations. First, we tested the scale only at one specific point in time, after children's first interactions with a social robot. Collecting the same data also at a later point is both costly and time-consuming, especially in the context of CRI research where children need to directly interact with a robot. Nevertheless, as mentioned before, research on CRI is often subject to the novelty effect in first interactions, and it would thus be important to validate the usefulness of the measure in a longitudinal perspective. Moreover, conceptually, intentional acceptance requires the intention to use a robot repeatedly or longitudinally to distinguish it from mere use. Empirically, this would mean that children have to interact with a robot repeatedly or at least imagine to interact with it repeatedly, to actually investigate intentional acceptance. Presumably, the psychometric properties of the scale (e.g., its distribution) would improve in longitudinal research in which novelty effects fade away. Thus, future research should assess the functioning of the scale in long-term interactions and the consequences of repeated interactions for the scale's psychometric properties (such as its reliability and distribution).

Second, self-report measures, such as the one we developed in this study, suffer from certain shortcomings. For example, children may give socially desirable answers (De Leeuw and Otter, 1995) or have difficulty in general to formulate a correct answer (De Leeuw et al., 2004; De Leeuw and Otter, 1995; Read and MacFarlane, 2006). Even though we tried to minimize these shortcomings as much as possible, for example by formulating concrete items and affirming, during the questionnaire, that every child is different and may provide different answers, response biases are difficult to completely eliminate. Therefore, it would be good, for triangulation reasons (Bethel and Murphy, 2010), to also develop and validate a measure of behavioral acceptance of social robots for children. Ideally, this measure would be an observational measure as children are bad at reporting the frequency of their own behavior over time (De Leeuw and Otter, 1995). Such a measure could, for example, be based on tracking children's duration of use of the robot (e.g., Sandygulova and O'Hare, 2016), the number of rounds of a game children play with the robot (Ribi et al., 2008), or the choice between playing with the robot and something else (Wigdor et al., 2016), all presuming that the CRI is a free interaction.

Third and finally, the measure was validated in a specific context with a specific, anthropomorphic robot. Future research should confirm whether these results can be generalized to other contexts, for example with an education-oriented interaction, and with other robot morphologies, such as a zoomorphic robot. Additionally, as this measure is specifically designed for children in middle childhood (more specifically 7- to 11-year-olds), future research should investigate the usability of the measure with other age groups. Children in adolescence would be of specific interest, as their language capability is sufficient to understand the scale, but the particular uses of social robots that are referred to by the indicators might not align with their interests as well as it does with children in middle childhood (Al-Taee et al., 2016).

6. Conclusion

With child-robot interactions increasingly moving outside the lab and to more natural environments (Michaelis and Mutlu, 2017), it is essential to study why children accept or reject social robots in the first place. To date, no validated measure existed for children's intentional acceptance of social robots, and children's intentional acceptance has rarely been the main focus of a study (except for Al-Taee et al., 2016). With the establishment of this reliable and valid measure for intentional acceptance of social robots by children, we hope to contribute to future research into the investigation of the effects of various child, robot, and interaction characteristics on intentional acceptance. As this new measure solely focuses on intentional acceptance, without including potential predictors of intentional acceptance, it allows future research to disentangle psychological responses, such as enjoyment, during the interaction from acceptance of social robots after the interaction. Our measure also allows future research to replicate and validate models of adult's social robot acceptance (e.g., De Graaf et al., 2019; Heerink et al., 2010; Shin and Choo, 2011) with children to see whether the same predictors and mechanisms are at stake when predicting acceptance of social robots.

Finally, the new scale of intentional acceptance of social robots for children may facilitate the investigation of the apparent discrepancy between children's intentional acceptance and behavioral acceptance of social robots. A review study of the literature on children's acceptance of social robots (De Jong et al., 2019) showed that there seems to be more variation in children's behavioral social robot acceptance than children's intentional social robot acceptance. With this new measure, we can reliably and validly measure children's intentional acceptance of social robots and compare it to their behavioral acceptance of robots, which may help to see whether there is, in contrast to adults, a difference between the two concepts for children.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the schools that participated in our research for their collaboration. Additionally, we wish to give our special thanks to the children that participated in this research. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. [682733]) to the third author.

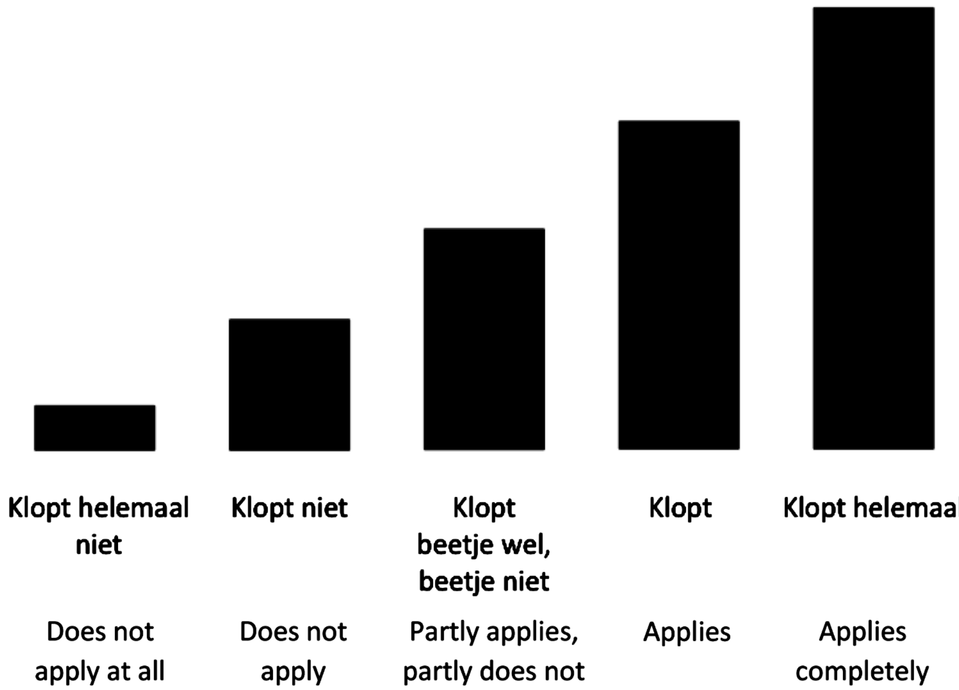
Appendix 1

Items of Intentional Acceptance of Social Robots

Item	Dutch	English
1	Ik wil NAO graag opnieuw zien.	I would like to see NAO again.
2	Ik wil graag opnieuw met NAO spelen.	I would like to play again with NAO.
3	Het zou leuk zijn als NAO en ik opnieuw iets kunnen doen samen.	It would be nice if NAO and I could do something together again.
4	Ik zou NAO mee naar huis willen nemen.	I would like to take NAO home.

Appendix 2

Bar scales (adapted from Severson and Lemm, 2016) used to indicate agreement or disagreement in Dutch and English respectively.



References

Abe, K., Iwasaki, A., Nakamura, T., Nagai, T., Yokoyama, A., Shimotomai, T., Okada, H., Omori, T., 2012. Playmate robots that can act according to a child's mental state, in: Proceedings of the International Conference on Intelligent Robots and Systems. Vilamoura, pp. 4660–4667. 10.1109/IROS.2012.6386037.

Ajzen, I., 1991. The theory of planned behavior. Org. Behav. Hum. Decis. Process. 50, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T.

Al-Taee, M.A., Kapoor, R., Garrett, C., Choudhary, P., 2016. Acceptability of robot assistant in management of type 1 diabetes in children. Diabetes Technol. Ther. 18, 551–554. https://doi.org/10.1089/dia.2015.0428.

Banthia, V., Maddahi, Y., May, M., Blakley, D., Chang, Z., Gbur, A., Tu, C., Sepehri, N., 2016. Development of a graphical user interface for a socially interactive robot: a case study evaluation, in: Proceedings of the 7th Annual Information Technology, Electronics and Mobile Communication Conference. Vancouver, pp. 1–8. 10.1109/IEMCON.2016.7746294.

Baxter, P., Ashurst, E., Read, R., Kennedy, J., Belpaeme, T., 2017. Robot education peers in a situated primary school study: personalisation promotes child learning. PLoS ONE 12, 1–23. https://doi.org/10.1371/journal.pone.0178126.

Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., Belpaeme, T., 2016. From characterising three years of HRI to methodology and reporting recommendations, in: Proceedings of the 11th International Conference on Human-Robot Interaction. Christchurch, pp. 391–398. 10.1109/HRI.2016.7451777.

Beidel, D.C., Turner, S.M., Morris, T.L., 1995. A new inventory to assess childhood social anxiety and phobia: the social phobia and anxiety inventory for children. Psychol. Assess. 7, 73–79. https://doi.org/10.1037/1040-3590.7.1.73.

Belpaeme, T., Baxter, P., Greeff, J.De, Kennedy, J., Read, R., Looije, R., Neerincx, M., Baroni, I., Zelati, M.C., 2013. Child-robot interaction: perspectives and challenges. In: Herrmann, G., Pearson, M.J., Lenz, A., Bremner, P., Spiers, A., Leonards, U. (Eds.), International Conference on Social Robotics: Lecture Notes in Computer Science. Springer, Bristol, pp. 452–459. https://doi.org/10.1007/978-3-319-02675-6_45.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bethel, C.L., Murphy, R.R., 2010. Review of human studies methods in HRI and recommendations. Int. J. Soc. Robot. 2, 347–359. https://doi.org/10.1007/s12369-010-0064-9.

Blanson Henkemans, O.A., Bierman, B.P.B., Janssen, J., Looije, R., Neerincx, M.A., van Dooren, M.M.M., de Vries, J.L.E., Burg, G.J., der, van, Huisman, S.D., 2017. Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. Int. J. Hum. Comput. Stud. 106, 63–76. https://doi.org/10.1016/j.ijhcs.2017.06.001.

Breazeal, C., Harris, P.L., DeSteno, D., Kory, J.M., 2016. Young children treat robots as informants. Top. Cogn. Sci. 8, 481–491. https://doi.org/10.1111/tops.12192.

Broadbent, E., 2017. Interactions with robots: the truths we reveal about ourselves. Annu. Rev. Psychol. 68, 627–652. https://doi.org/10.1146/annurev-psych-010416-043958.

Byrne, B.M., 2010. Structural Equation Modeling With AMOS: basic concepts, applications, and programming, 2nd ed. Routledge, Oxford.

Byrne, B.M., Shavelson, R.J., Muthén, B., 1989. Testing for the equvalence of fator covariance and mean structures: the issue of partial measurement in variance. Psychol. Bull. 105, 456–466. https://doi.org/10.1037/0033-2909.105.3.456.

Carbocation Corporation, 2016. False discovery rate online calculator [WWW document]. URLhttps://tools.carbocation.com/FDR.

Carmines, E.G., Zeller, R.A., 1979. Reliability and validity assessment. Sage, Beverly Hills, CA. 10.4135/9781412985642.

Carpenter, S., 2018. Ten steps in scale development and reporting: a guide for researchers. Commun. Methods Meas. 12, 25–44. https://doi.org/10.1080/19312458.2017.1396583.

Chen, F.F., 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct. Equ. Model. 14, 464–504. https://doi.org/10.1080/10705510701301834.

Chen, F.F., Bollen, K.A., Paxton, P., Curran, P.J., Kirby, J.B., 2001. Improper solutions in structural equation models. Sociol. Methods Res. 29, 468–508. https://doi.org/10.1177/0049124101029004003.

Cole, M., Cole, S.R., Lightfoot, C., 2005. The Development of children, 5th ed. Worth Publishers, New York, NY.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555.

Davis, F.D., 1989. Perceived usefulness, rerceived ease of use, and user acceptance of social robots. MIS Q 13, 319–340. https://doi.org/10.2307/249008.

De Haas, M., Mois Arayo, A., Barakova, E., Haselager, W., Smeekens, I., 2016. The effect of a semi-autonomous robot on children, in: Proceedings of the 8th International Conference On Intelligent Systems. Sofia, pp. 376–381. 10.1109/is.2016.7737448.

De Graaf, M.M.A., Ben Allouch, S., Van Dijk, J.A.G.M., 2018. A phased framework for long-term user acceptance of interactive technology in domestic environments. New Media Soc. 20, 2582–2603. https://doi.org/10.1177/1461444817727264.

De Graaf, M.M.A., Ben Allouch, S., van Dijk, J.A.G.M., 2019. Why would I use this in my home? A model of domestic social robot acceptance. Human-Computer Interact 34, 115–173. https://doi.org/10.1080/07370024.2017.1312406.

De Jong, C., Kühne, R., Peter, J., Van Straten, C.L., Barco, A., 2019 What do children want from a social robot? Toward gratifications measures for child-robot interaction, in: Proceedings of the 28th IEEE International Conference On Robot and Human Interactive Communication. New Delhi, pp. 1-8. 10.1109/RO-MAN46459.2019.8956319.

De Jong, C., Peter, J., Kühne, R., Barco, A., 2019. Children's acceptance of social robots: a narrative review of the research 2000-2017. Interact. Stud 20, 393–425. https://doi.org/10.1075/is.18071.jon.

De Jong, Chiara, Peter, Jochen, Barco, Alex, Kühne, Rinaldo, Van Straten, Caroline L., 2020. Exploring children's beliefs that underlie their intended domestic adoption or rejection of social robots. Interaction Design and Children conference Submitted for publication.

De Leeuw, E., Borgers, N., Smits, A., 2004. Pretesting questionnaires for children and adolescents. In: Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J. (Eds.), Methods for Testing and Evaluating Survey Questionnaires. John Wiley & Sons, New Jersey, pp. 409–429. https://doi.org/10.1002/0471654728.ch20.

De Leeuw, E.D., Otter, M.E., 1995. The reliability of children's responses to questionnaire items; question effects in children's questionnaire data. In: Hox, J.J., Van der Meulen, B.F., Janssens, J.M.A.M., Ter Laak, J.J.F., Tavecchio, L.W.C. (Eds.), Hearing Children's Voices. Thesis Publishers, Amsterdam, pp. 251–257.

Dillon, W.R., Kumar, A., Mulani, N., 1987. Offending estimates in covariance structure analysis: comments on the causes of and solutions to Heywood cases. Psychol. Bull. 101, 126–135. https://doi.org/10.1037/0033-2909.101.1.126.

Engle, R.F., 1983. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: Grilliches, Z., Intrilligator, M.D. (Eds.), Handbook of Econometrics. Elsevier Science, Amsterdam, pp. 796–801. https://doi.org/10.1002/0471142727.mb1300s82.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., 2011. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism, in: Proceedings of the 6th International Conference on Human-Robot Interaction. Lausanne, pp. 61–67. 10.1145/1957656.1957673.

Ferraz, M., Câmara, A., O'Neill, A., 2016. Increasing children's physical activity levels through biosymtic robotic devices, in: Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology. Osaka, p. no. 2. 10.1145/3001773.3001781.

Fong, T, Nourbakhsh, I, Dautenhahn, K, 2003. A survey of socially interactive robots. Robotics and Autonomous Systems 42, 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X.

Hagen, R.L., 1997. In praise of the null hypothesis statistical test. Am. Psychol. 52, 15–24. https://doi.org/10.1037/0003-066X.52.1.15.

Heerink, M., Kröse, B., Evers, V., 2010. Assessing acceptance of assistive social agent technology by older adults: the Almere model. Int. J. Soc. Robot. 2, 361–375. https://doi.org/10.1007/s12369-010-0068-5.

Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55. https://doi.org/10.1080/10705519909540118.

Kanda, T., Hirano, T., Eaton, D., Ishiguro, H., 2004. Interactive robots as social partners and peer tutors for children: a field trial. Human-Comput. Interact 19, 61–84. https://doi.org/10.1207/s15327051hci1901&2_4.

Kanda, T., Shimada, M., Koizumi, S., 2012. Children learning with a social robot, in: Proceedings of the 7th International Conference on Human-Robot Interaction. Boston, pp. 351–358. 10.1145/2157689.2157809.

Kędzierski, J., Muszyński, R., Zoll, C., Oleksy, A., Frontkiewicz, M., 2013. EMYS-Emotive head of a social robot. Int. J. Soc. Robot. 5, 237–249. https://doi.org/10.1007/s12369-013-0183-1.

Kenny, D.A., Kaniskan, B., McCoach, D.B., 2015. The performance of RMSEA in models with small degrees of freedom. Sociol. Methods Res. 44, 486–507. https://doi.org/10.1177/0049124114543236.

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. Adv. Robot. 23, 1951–1996. https://doi.org/10.1163/016918609X12518783330360.

Kruijff-Korbayová, 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, in: Proceedings of the 23rd International Symposium on Robot and Human Interactive Communication. Edinburgh, pp. 649–654. 10.1109/ROMAN.2014.6926326.

Kühne, R., 2013. Testing measurement invariance in media psychological research. J. Media Psychol. 25, 153–159. https://doi.org/10.1027/1864-1105/a000096.

La Greca, A.M., Stone, W.L., 1993. Social anxiety scale for children-revised: factor structure and concurrent validity. J. Clin. Child Psychol. 22, 17–27. https://doi.org/10.1207/s15374424jccp2201.

Lee, K.M., Peng, W., Jin, S.A., Yan, C., 2006. Can robots manifest personality?: an empirical test of personality recognition, social responses, and social presence in human-robot interaction. J. Commun. 56, 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x.

Looije, R., Neerincx, M.A., Hindriks, K.V., 2017. Specifying and testing the design rationale of social robots for behavior change in children. Cognit. Syst. Res. 43, 250–265. https://doi.org/10.1016/j.cogsys.2016.07.002.

Marsh, H.W., 1986. Negative item bias in ratings scales for preadolescent children: a cognitive-developmental phenomenon. Dev. Psychol. 22, 37–49. https://doi.org/10.1037/0012-1649.22.1.37.

Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58, 525–543. https://doi.org/10.1007/bf02294825.

Leite, I., Pereira, A., Lehman, J.F., 2017. Persistent memory in repeated child-robot conversations, in: Proceedings of the 2017 Conference on Interaction Design and Children. Stanford, pp. 238–247. 10.1145/3078072.3079728.

Michaelis, J.E., Mutlu, B., 2017. Someone to read with: design of and experiences with an in-home learning companion robot for reading, in: Proceedings of the Conference on Human Factors in Computing Systems. Denver, pp. 301–312. 10.1145/3025453.3025499.

Mplus, n.d. Chi-square difference testing using the Satorra-Bentler scaled chi-square. URLhttps://www.statmodel.com/chidiff.shtml.

Muthén, L.K., Muthén, B.O., 2012. Mplus User's Guide, 7th. Muthén & Muthén, Los Angeles, CA.

Oliver, M.B., Bartsch, A., 2010. Appreciation as audience response: exploring entertainment gratifications beyond hedonism. Hum. Commun. Res. 36, 53–81. https://doi.org/10.1111/j.1468-2958.2009.01368.x.

Read, J.C., MacFarlane, S., 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction, in: Proceedings of the 2006 Conference on Interaction Design and Children. Tampere, pp. 81–88. 10.1145/1139073.1139096.

Ribi, F.N., Yokoyama, A., Turner, D.C., 2008. Comparison of children's behavior toward Sony's robotic dog Aibo and a real dog. A pilot study. Anthrozoos 21, 245–256. https://doi.org/10.2752/175303708X332053.

Saint-Aimé, S., Le Pévédic, B., Duhaut, D., 2011. Preliminary study to evaluate EMI emotional interaction with two young children, in: Proceedings of the International Conference on Mechatronics and Automation. Beijing, pp. 1309–1314. 10.1109/ICMA.2011.5985851.

Sandygulova, A., O'Hare, G.M.P., 2016. Investigating the impact of gender segregation within observational pretend play interaction, in: Proceedings of the 11th International Conference on Human-Robot Interaction. Christchurch, pp. 399–406. 10.1109/HRI.2016.7451778.

Satorra, A., Bentler, P.M., 2010. Ensuring positiveness of the scaled difference chi-square test statistics. Psychometrika 75, 243–248. https://doi.org/10.1007/s11336-009-9135-y.

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 Psychol. Res. Online 8, 23–74. https://doi.org/10.1002/0470010940.

Severson, R.L., Lemm, K.M., 2016. Kids see human too: adapting an individual differences measure of anthropomorphism for a child sample. J. Cogn. Dev. 17, 122–141. https://doi.org/10.1080/15248372.2014.989445.

Shin, D.-.H., Choo, H., 2011. Modeling the acceptance of socially interactive robotics: social presence in human–robot interaction. Interact. Stud. 12, 430–460. https://doi.org/10.1075/is.12.3.04shi.

Steenkamp, J.E.M., Baumgartner, H., 1998. Assessing measurement invariance in cross-national consumer research. J. Consum. Res. 25, 78–107. https://doi.org/10.1086/209528.

Sung, J.Y., Christensen, H.I., Grinter, R.E., 2009. Robots in the wild: understanding long-term use, in: Proceedings of the 4th International Conference On Human-Robot Interaction. La Jolla, pp. 45–52. 10.1145/1514095.1514106.

Valkenburg, P.M., Peter, J., 2007. Preadolescents' and adolescents' online communication and their closeness to friends. Dev. Psychol. 43, 267–277. https://doi.org/10.1037/0012-1649.43.2.267.

Van der Heijden, H., 2004. User acceptance of hedonic information systems. MIS Q. 28, 695–704. https://doi.org/10.2307/25148660.

Van Straten, Kühne, R., Peter, J., De Jong, C., Barco, A., 2020. Closeness, trust, and perceived social support in child-robot relationship formation: development and validation of three self-report scales. Interact. Stud. 21, 57–84. https://doi.org/10.1075/is.18052.str.

Van Straten, C.L., Peter, J., Kühne, R., 2019. Child-robot relationship formation: a narrative review of empirical research. Int. J. Soc. Robot 1–20. 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, in: Proceedings of the 6th International Conference on Human-Agent Interaction. 10.1145/3284432.3284440.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. MIS Q. 27, 425–478. https://doi.org/10.2307/30036540.

Vogt, P., De Haas, M., De Jong, C., Baxter, P., Krahmer, E., 2017. Child-robot interactions for second language tutoring to preschool children. Front. Hum. Neurosci. 11. https://doi.org/10.3389/fnhum.2017.00073.

Wigdor, N., De Greeff, J., Looije, R., Neerincx, M.A., 2016. How to improve human-robot interaction with conversational fillers, in: Proceedings of the 25th International Symposium on Robot and Human Interactive Communication. New York, pp. 219–224. 10.1109/ROMAN.2016.7745134.

Wood, L.J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., Syrdal, D.S., 2013. Robot-mediated interviews - How effective is a humanoid robot as a tool for interviewing young children? PLoS ONE 8, e59448. https://doi.org/10.1371/journal.pone.0059448.

Zhao, S., 2006. Humanoid social robots as a medium of communication. New Media Soc. 8, 401–419. https://doi.org/10.1177/1461444806061951.