Measuring Young Children's Long-Term Relationships with Social Robots

Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal
MIT Media Lab, 20 Ames St.
Cambridge, MA 02139, USA
{jakory,haewon,randiw12,cynthiab}@media.mit.edu

ABSTRACT
Social robots are increasingly being developed for long-term interactions with children in domains such as healthcare, education, therapy, and entertainment. As such, we need to deeply understand how children's relationships with robots develop through time. However, there are few validated assessments for measuring young children's long-term relationships. In this paper, we present a pilot test of four assessments that we have adapted or created for use in this context with children aged 5–6: the Inclusion of Other in Self task, the Social-Relational Interview, the Narrative Description, and the Self-disclosure Task. We show that children can appropriately respond to these assessments with reasonably high internal reliability, and that the proposed assessments are able to capture child-robot relationship adjustments over a long-term interaction. Furthermore, we discuss gender and population differences in children's responses.

ACM Classification Keywords
J.4 Social and Behavioral Sciences: Psychology; K.4.2 Computers and Society: Social Issues; I.2.9 Robotics

Author Keywords
assessments; children; long-term interaction; human-robot interaction; social robots

INTRODUCTION
Social robots are increasingly being developed for use with children in application domains such as education, entertainment, healthcare, and therapy [11, 14, 21, 28, 41]. In these domains, because learning and behavior change may take weeks or months to achieve, the robot interactions must necessarily move toward longer-term encounters. Because children will not simply have a one-off interaction, we need to deeply understand how children think about the robots through time. In prior research, we have seen that children treat robots as more than mere artifacts, e.g., ascribing them mental states, psychological attributes, and moral standing [17, 30]. Furthermore, in long-term interactions, social robots are taking on a *relational* role—i.e., they are situated as agents that actively attempt to build and maintain long-term social-emotional relationships [4]. They are introduced as peers, tutors, and learning companions [21, 37, 41, 43]. While children's relationships with robots may not be like the relationships they have with their parents, pets, imaginary friends, or smart devices, they will form relationships of some kind, and as such, we need to find ways to characterize and measure these relationships.

However, most existing assessments for measuring relationships have targeted older children or adults [1, 2, 13, 34]. There is a dearth of assessments for measuring young children's relationships. In an effort to remedy this, we have created four assessments for measuring the relationships that children aged 5–6 years form with social robots, which we have tested in a two-month field study (early results in [25]). This paper makes two contributions. First, we provide reliability and validity information about the assessments. Second, we report the results of the field study. Our rationale for taking this two-pronged approach is that we wanted to show that the assessments were valid and could be used by others, and also to show what kind of results might be obtained with the assessments during a long-term child-robot interaction study. We hope that including the specific study results will provide a point of comparision for future work, so we can see whether children develop relationships or rate robots similarly across different studies, robots, and contexts.

Measuring children's relationships with robots will not only give us insight into how children think about robots through time, but will also lead us toward developing autonomous systems that can model and manage the ongoing relationship. This could, e.g., allow a robot to determine whether it still needs to gain a child's friendship before it can effectively administer an intervention, or whether the child is too attached, and the robot needs to recommend that the child seek a person for help instead. Prior work has accomplished this with adults [19], using relationship assessments to assess, model, maintain, and repair a relationship over repeated encounters to achieve the long-term goal of being a weight-loss coach.

IDC '18 June 19–22, 2018, Trondheim, Norway

ACM ISBN 978-1-4503-5152-2/18/06.

DOI: https://doi.org/10.1145/3202185.3202732

BACKGROUND

Relationship Behaviors

Prior work has found evidence that children readily treat robots as social agents [6, 16, 18, 23, 24, 36]. When children are given time to develop relationships with social agents, such as in long-term human-robot interaction studies, children display social behaviors such as sharing gaze, mirroring emotions, affection, helping behaviors, turn-taking, and disclosing information [11, 22, 26, 29, 41, 42]. Developmental psychology research suggests that these behaviors are related to children's friendships and close relationships [10, 12, 33, 40]. Friends are perceived as social beings with psychological attributes [10]. Children recognize that affection, empathy, feeling close, and wanting companionship are part of friendship [10, 40]. They solve conflicts more equitably with friends [10, 12], and share secrets and disclose more personal information [10, 39].

When studying children's long-term relationships with social robots, it will be important to assess whether children display these kinds of friendship behaviors, as well as the extent to which they see the robot as responding in kind. The assessments we present focus on measuring the following behaviors: children's perception of closeness to the robot, self-disclosure to the robot, perception of the robot as a social-relational agent like themselves, and a comparison of children's descriptions of a human best friend versus of the robot.

Assessing Relationships

Existing work on long-term child-robot interaction has covered a wide range of ages, including younger children 4–6 years [11, 22, 42], older children (e.g., 10–13 years) [41], or a wide age range (e.g., 5–12 years) [26, 29]. These studies have used a variety of assessments to measure children's engagement, the robot's social presence, and the effectiveness of the interaction (e.g., learning gains). Commonly, multiple-choice questionnaires were administered, often using smiley faces instead of numerical values. Some were variations on the Fun Toolkit, a set of assessments for measuring children's engagement and fun that has been validated with children aged 7–12 years [38]. Some behavioral measures were also used, such as children's gaze, affect, and speech patterns. However, no assessments explicitly measured children's relationships with the robot—such as feelings of closeness.

There are numerous existing ways of measuring people's relationships, most of which can be adapted to measure people's relationships with robots. For example, the Working Alliance Inventory [13] measures the quality of the alliance between two individuals and was used to measure an adult-robot alliance in [19]. Most assessments for adults involve self-report questionnaires, which necessarily involve reading, comprehending, and answering questions [1, 2, 34]. Many could fairly easily be adapted for older children who can both read and self-reflect. However, younger children may be pre-reading. They may have shorter attention spans, and may not be able to fill out a standard Likert-style questionnaire [8]. Thus, the assessments that work well for older children and adults may not work for this age group.

With younger children, one common way of assessing their relationships with others or their friendships with peers has been through observational methods. In these methods, children's behavior is coded for various dimensions of relationships, such as features of friendship (e.g., companionship, aid, exclusivity), connectedness, conflict, and physical proximity [12, 27, 44]. Few studies have asked children directly about their relationships or used behavioral methods to probe their relationships [10, 35, 39].

METHODOLOGY

Participants

We performed a pilot test of four relational assessments during a long-term child-robot interaction study at three Boston-area schools. We recruited from multiple schools because it was not possible to recruit sufficient children from a single school. This had the benefit of allowing us to recruit a diverse population of children (one higher-socioeconomic status (SES) school and two lower-SES schools). Forty-four children aged 4–7 (M = 5.4, SD = 0.66) participated in the study. There were 3 four-year-olds, 21 five-year-olds, 19 six-year-olds, and 1 seven-year-old. We included the 4-year-olds and the 7-year-old because they were in the same classrooms as the other children. Given their birthdates, the 4-year-olds were nearly 5 and the 7-year-old had only recently turned 7. Thus, during the analyses reported below, we grouped the 4-year-olds with the 5-year-olds, and the 7-year-old with the 6-year-olds.

There were 16 children from school A, 13 from school B, and 15 from school C. While there were no statistically significant differences in mean age between schools, school C had a median age of 5 and schools A and B had a median age of 6. Children at all schools were English and Indo-European language bilinguals (mostly Spanish), out of which five children were English Language Learners. Fifteen children were from English-dominant families (ED), 24 from Spanish-dominant families (SD), and 5 used another language as a dominant language at home (OD). Table 1 lists demographic information.

Assessments

The full instructions and materials for each assessment described below as well as our revisions are available on figshare: 10.6084/m9.figshare.5419102.

Social-Relational Interview (SRI)

We created a set of questions targeting children's perceptions of the robot as a social, relational agent. These questions move away from how children feel about a robot—such as whether children attribute certain properties to robots [15, 17, 20]—and toward how children think robots feel. Five questions targeted provisions of children's friendship: conflict, instrumental help, sharing secrets / disclosure, wanting companionship, and empathy / affection [10, 27]; these questions are somewhat similar to those asked in the McGill Friendship Questionnaires [31]. Two questions asked whether the robot was genuine, i.e., whether what it felt was real or whether it was just pretending (i.e. coudln't really feel that way). Each question offered three responses: "yes, the robot would feel something (e.g., sad or happy)", "maybe / don't know", and

Table 1. Demographic information about the participants by school. M. = Mean; Med. = Median; ED = English-dominant families; SD = Spanish-dominant; OD = Other- dominant. The median income and median mother's education are only from parents who agreed to disclose this information.

School	M. Age (SD)	Med. Age	Girls	Boys	ED	SD	OD	Median income	Median mother education
A	5.56 (0.51)	6	8	8	7	6	3	over 150k USD	graduate or professional training
B	5.54 (0.78)	6	9	4	3	9	1	10k-30k USD	community college or similar
C	5.13 (0.64)	5	7	8	5	9	1	30k-50k USD	college

"no, the robot wouldn't mind (coded as 2, 1, 0)". Each question was followed by asking the child to explain their choice, and whether they would feel the same way as the robot. This way, we have some context for understanding children's response. Thus, if a child chose "no, the robot wouldn't mind," but said that the robot would not be sad because it was happy playing by itself, then we could adjust the coding of the child's response to reflect the child's perception of the robot as a social, relational individual. We also computed a composite score consisting of the sum of children's responses to get an overall picture of the child's perception of the robot as a social, relational agent.

Inclusion of Other in Self (IOS) Task

The Inclusion of Other in Self scale is a single item pictorial measure of closeness and interconnectedness [1, 9]. Participants are shown pictures of seven pairs of increasingly overlapping circles, and asked to point to the circles that best describe their relationship with someone. We have adapted it for use with preschool children. Each child is asked about their relationship with their best friend, a bad guy they saw in the movies that they do not like, a parent, the robot, and a pet or favorite toy. We include the non-robot items as a comparison, so we can see where the robot stands in relation to these other characters in the child's life.

Narrative Description Task

In this task, a puppet asks the child to help it learn about people and robots. We used a puppet because it could plausibly need to learn this information, while the adult/experimenter would not—the experimenter is a person themselves, and, because they are facilitating an activity with a robot, likely know something about robots as well. The child is then asked to describe both their best friend and the robot that they played with. The goal is to see how the child describes the robot in relation to how they describe their friend. We expected that each description would include a mix of physical attributes (e.g., the robot is red and blue, my friend is tall) and psychological/relational characteristics or activities performed together (e.g., we play together, she's nice), and that children might include more psychological/relational elements for their friend and for the robot with whom they have a closer relationship (e.g., at the posttest vs. at the pretest).

Self-disclosure Task

Because self-disclosure is one of the features of children's friendships [7, 10, 39], we had the robot disclose information and prompt for information disclosure in return. The protocol was adapted from [39]. The robot twice disclosed a piece of information, paused to allow children time to spontaneously disclose in return, and prompted for a disclosure. For example, the robot would say "Did you know, I'm afraid of messing up. One time I was trying to sing a song but I forgot the words and I got it all wrong. I think I'm bad at singing. (wait for response) What about you? Can you tell me things that you are not so good at, or things you tried but didn't go so well? (wait for response) Don't be shy, tell me what you think. (wait for response)". As per [39], the amount of disclosure can be measured by counting the number of utterances made. We also examined the kind of information disclosed, as we expected both that children would disclose more total information as well as personal or sensitive information to a robot with whom they have a closer relationship (e.g., more during a posttest than a pretest).

Figure 1. A child listens to the autonomous robot Tega tell a story during the study. The story pictures are shown on the tablet.

Procedure

During the study, the children interacted one-on-one with a fully autonomous social robot, Tega, approximately 1–2 times a week, for a total of 7 sessions (Figure 1). The robot was introduced as a peer who likes stories. Each session, it told stories and children were asked to retell the stories, in effect acting as a storytelling tutor. The robot called the child by name and occasionally referenced shared experiences such as stories told together.

We administered the IOS task, Narrative Description, and SRI after children's first and last sessions with the robot. Due to its length, the Self-disclosure Task was implemented as part of a conversation at the start of the second session, and a second time during the start of the final session.

Data

We recorded children's responses to the IOS task and SRI in a spreadsheet. Children's speech during all tasks was recorded with a microphone and transcribed for analysis.

Data Analysis

We took a two-pronged approach to data analysis. First, relating to the first goal of this paper, we examined reliability and validity, including ease and appropriateness of responding. Then, we examined the results of the study with respect to pretest-posttest differences, and differences by age, gender, and school. These analyses support the second goal of this paper—i.e., to provide a point of comparison for future work—as well as showing that these assessments do reflect individual differences in relationships and do have sufficient variability in responses. We were particularly interested in gender differences because prior work has shown that boys and girls develop friendships differently—girls often rate intimacy and alliance in their friendships more highly [7, 10]. Because SES may be related to the development of social-emotional skills and friendships, we also expected that children from the higher-SES school A might have stronger existing relationships as indicated by the IOS task, and might be more likely to form strong relationships with the robot [5, 32].

As part of our analysis of ease and appropriateness of responding, we coded children's explanations of their SRI responses to find out why children chose the answers they did. One author first performed open coding to identify distinct concepts and themes in the data. The themes identified were: (1) explicit references to the robot's feelings (e.g., "She would feel sad"), (2) references to the robot's attributes (e.g., "He will not because he's too little"), (3) references to the robot's actions (e.g., "Because Tega said nice to meet you"), (4) references to the child/self as explanation (e.g., "Just like when Judah snatched the train track from my hand"), (5) references to others (e.g., "Because someone was mean", "so the kid can't be sad"), (6) references to the situation (e.g., "Because he took her toy", "it's an emergency"), (7) references to consequences ("Because then she can't read a story"), and (8) references to moral judgments or obligations (e.g., "Because that's not nice"). Two other authors then performed a second axial coding to confirm that these concepts reflected the breadth of children's responses. Disagreements in coding were resolved via discussion.

Similarly, children's descriptions of their friend and of the robot in the Narrative Description were coded into the following categories: (1) mention of the agent's name (e.g., "My friend's name is…"), (2) description of physical attributes (e.g., "He is really tall", "When you stand up, she has three blue hairs"), (3) description of social/cognitive attributes (e.g., "She's nice, and she listens, and she's kind", "Tega is smart"), (4) mention of other facts (e.g., "He went to another school", "She has a sister and brother"), or (5) activities performed together (e.g., "I play with him a lot in recess", "She tells me stories").

We also coded children's disclosures during the Self-disclosure Task to learn what kind of information children tended to disclose to the robot. These were coded into the following categories: (1) physical skills (e.g., "stand on one foot", "singing", "ride a bike"), (2) fine motor skills (e.g., "writing", "draw a ball", "coloring"), (3) social skills (e.g., "teaching", "I don't know how to share"), (4) cognitive skills (e.g., "reading", "math", "doing puzzles"), and (5) not specific (e.g., "I'm not good at many things", "a lot of things").

Finally, a few children refused to answer individual questions on an assessment or did not complete the full assessment (e.g., one child was not available to complete the posttest). These children were excluded from the relevant analyses.

RESULTS

Ease and appropriateness of response

Social-Relational Interview

One-sample t-tests were used to compare the mean number of "positive" responses (i.e., indications that the robot was more friend-like and not "just pretending") for each SRI question in the pretest and in the posttest to chance levels of responding (i.e., mean of 1 for each question, or mean of 7 for the composite score). The results are shown in Table 2. Children's responses differed from chance in the expected ways: children said the robot had friend-like qualities, in that it would be sad if a child was mean to it or if it had no friends, help a child who needed help, and cheer up a child who was sad. Furthermore, children tended to say the robot really did want to make friends (it was not just pretending), and really did like them. Children's responses to the question about sharing secrets did not differ from chance levels.

Regarding childen's justifications for their responses, most children provided explanations, with a mean of 30.1 of the 44 children (*SD* = 2.66) doing so for each SRI item. A mean of 5.5 children (*SD* = 2.56) per item used more than one justification type in their response, e.g., refering both to the robot's feelings and to consequences: "Because she is not going to have no one to play with or talk with. That would make her really, really, really sad." The frequency of each justification type by SRI item is listed on in a Table we have put on figshare: 10.6084/m9.figshare.6128054. These frequencies are summed across the pretest and posttest because children often used the same justification types at both times, and thus, the more interesting feature to examine is what type of justifaction children used when explaining that *yes*, the robot did feel something, or *no*, the robot did not mind or did not care.

Children most commonly cited the robot's feelings, e.g., "If you just left him here and nobody came to play with him, he might be sad," and "Because he likes sharing stuff like stories". They were more likely to reference the robot's feelings when giving a positive (i.e., social/relational) response. They were more likely to talk about consequences when considering why the robot might be sad, e.g., "Because if you don't have any friends, you won't have anybody to play with.".

Children referenced the robot's attributes most often when explaining why it would help or cheer up a child, e.g., "Because she's nice," or "Because she would be a good friend." Two children said the robot would not help because it had no legs and could not move; when probed further, they decided the robot would help if the situation did not require moving. Children also cited moral reasons in these cases, such as "Because it's nice to help." However, moral reasons were cited most often when discussing secrets, e.g., "Because you are

Table 2. Summary of children's overall SRI responses. All but the sharing secrets question differed significantly from chance (mean = 1), as shown by one-sample t-tests. Here, "df" = "degrees of freedom", "t" = "t-value", and "p" = "p-value".

Question	Time	Mean *(SD)*	df	*t*	*p*
Sad if mean	*pre*	1.61 *(0.78)*	43	5.19	<0.001
	post	1.71 *(0.72)*	40	6.33	<0.001
Sad no friend	*pre*	1.73 *(0.66)*	43	7.31	<0.001
	post	1.71 *(0.68)*	40	6.66	<0.001
Help child	*pre*	1.59 *(0.82)*	43	4.80	<0.001
	post	1.71 *(0.68)*	40	6.66	<0.001
Share secret	*pre*	1.00 *(1.00)*	42	0.00	1.00
	post	0.98 *(0.99)*	40	-0.16	0.875
Cheer child	*pre*	1.53 *(0.85)*	42	4.10	<0.001
	post	1.80 *(0.60)*	40	8.58	<0.001
Wants friends	*pre*	1.44 *(0.91)*	42	3.19	0.003
	post	1.56 *(0.84)*		4.29	<0.001
Likes you	*pre*	1.76 *(0.66)*	41	7.53	<0.001
	post	1.37 *(0.92)*	40	2.56	0.014
Total	*pre*	10.8 *(3.32)*	41	7.35	<0.001
	post	10.6 *(3.5)*	41	6.69	<0.001

not supposed to hide lots of secrets." One child disclosed that her teacher did not want kids to tell secrets in class. When asked if the robot really liked them, most children spoke about themselves or gave no explanation, e.g., " Because I told her a story," and "Because I'm nice." They also mentioned actions the robot took, such as "Because she read me a story," and "Because he always says [name], hi [name]".

Finally, nearly all children said they would feel the same way as the robot. Those who did not also tended to be the children who said the robot would not care.

IOS Task

One-sample t-tests were used to compare the mean of children's responses to chance levels of responding (i.e., mean of 3.5) for each IOS question. Descriptive statistics and the t-test results are shown in Table 3. Children's responses differed from chance in the expected directions: children rated their best friend, a parent, and a pet or toy as closer. They rated a bad guy from the movies that they didn't like as farther. The robot was also rated as closer.

Narrative Description

During the pretest, 38 children provided descriptions of their best friend and 34 provided descriptions of the robot. At the posttest, 36 children described their friend and the robot. During the pretest, children's descriptions of their friends were a mean of 42.6 words (*SD* = 42.8; *sentences M* = 4.42, *SD* = 3.62) and of the robot were a mean of 26.4 words (*SD* = 25.6; *sentences M* = 3.26, *SD* = 2.75). During the posttest, children's descriptions of their friends were a mean of 38.8 words (*SD* = 37.6; *sentences M* = 4.22, *SD* = 2.74) and of the robot were a mean of 35.6 words (*SD* = 29.3; *sentences M* = 3.72, *SD* = 2.81).

Table 4 lists the types of descriptions used at the pretest and posttest. Children's descriptions of their friends most often involved activities performed together, e.g., "He plays outside with me." They also described things their friends liked, e.g., "I know that he likes to play road blocks and Minecraft." Children shared similar information about the robot, e.g., "We read a story, two stories." Children rarely described their friends' physical characteristics, e.g., "She had long hair, cute shoes." They did so more often for the robot, e.g., "Tega's so cute, she's soft." They also were somewhat more likely to share facts about the robot, e.g., "Her favorite color is red and blue. She likes to talk about stories."

Self-disclosure Task

During the pretest, 9 children spoke to the robot after each of the robot's two disclosures (before the prompt); 40 children spoke after each of the robot's prompts with 29 disclosing information. At the posttest, 4 children spoke to the robot after the robot's first disclosure and 2 after the second; 37 children spoke following each of the robot's prompts, with 30 disclosing information. The mean word and sentence counts are listed in Table 5.

When prompted to disclose what they were good or bad at, children generally disclosed physical and cognitive skills, like the robot did, but did not say much more, e.g., "I'm good at singing", "I don't know how to draw a ball," and "I tried riding a bike, but I would fall." Table 6 lists the types of disclosures made at the pretest and posttest.

Reliability

Social-Relational Interview

The reliability of the SRI during the pretest and the posttest was determined by measuring the internal consistency of the seven core questions using Cronbach's alpha. An alpha coefficient of 0.70 (95% CI: 0.57–0.89) was found for the pretest. Item reliability was calculated through an item analysis, which revealed that all seven questions were correlated with the total score, with *r* values between 0.54–0.80 for all but one item. If we dropped the item about sharing secrets (r = 0.32), the reliability would improve to 0.78. For the posttest, an alpha coefficient of 0.73 (95% CI: 0.61–0.86) was found when the sharing secrets question was reverse-scored. Item analysis showed that reliability would improve to 0.78 if the sharing secrets item (r = 0.45) was removed; the *r* values were between 0.61–0.75 for all other items.

Because the items' reasonably high internal reliability, we computed the sum of the SRI items as a composite score. We computed test-retest reliability for each item and for the composite score. Test-retest reliability was poor overall, which was reasonable since we expected children to change their opinions of the robot over time. The lowest item was the sharing secrets item, r = -0.052; the other items ranged from r = 0.220 to r = 0.535; the composite score was r = 0.050.

IOS Task

The reliability of the IOS task for the pretest and the posttest was determined by measuring the internal consistency of the five questions using Cronbach's alpha. An alpha coefficient of 0.74 (95% CI: 0.61–0.86) was found for the pretest scores

Table 3. Children's overall IOS responses. All differed significantly from chance (mean = 3.5), as shown by one-sample t-tests.

Question	Time	Median	Mode	Range	Inter-quartile Range	Mean *(SD)*	df	t-value	p-value
Best Friend	*pre*	5	7	1–7	4	4.82 *(1.90)*	43	4.61	<0.001
	post	6	7	1–7	4	5.03 *(1.97)*	39	12.9	<0.001
Parent	*pre*	5	7	1–7	4	4.84 *(1.90)*	43	4.67	<0.001
	post	6	7	1–7	3	5.24 *(1.89)*	40	14.3	<0.001
Pet or toy	*pre*	5	5	1–7	3	4.68 *(1.72)*	39	4.33	<0.001
	post	5	7	1–7	4	5.00 *(1.88)*	40	13.6	<0.001
Bad guy	*pre*	1	1	1–7	1	1.81 *(1.55)*	42	-7.15	<0.001
	post	1	1	1–7	1	1.77 *(1.42)*	38	3.21	0.003
Robot	*pre*	4	3	1–7	3	4.55 *(1.73)*	43	4.01	<0.001
	post	5	7	1–7	3	5.10 *(1.79)*	40	14.7	<0.001

Table 4. Frequency of description types children used to describe their friend and the robot at the pretest and posttest. Not all children provided descriptions. Many children used more than one type of description. Here, "Soc/Cog." = "Social/Cognitive Attributes"; "Phys." = "Physical Attributes"; "Act." = "Activity".

Agent	Time	Name	Soc/Cog.	Phys.	Fact	Act.
Friend	*pre*	17	13	3	15	29
Friend	*post*	16	10	4	15	29
Robot	*pre*	14	12	10	20	17
Robot	*post*	14	12	12	13	19

Table 5. Mean word and sentence counts for children's disclosures.

Time	Disclosure	Words	Sentences
pre	1	6.30 (5.60)	1.30 (0.73)
post	1	9.12 (8.36)	1.57 (1.17)
pre	2	9.64 (8.42)	1.36 (0.75)
post	2	10.7 (10.8)	1.45 (1.13)

(the "bad guy" item was reverse-scored). Item reliability was calculated through an item analysis, which revealed that all five items were correlated with the total score, with *r* values between 0.63–0.77 for all items. For the posttest, an alpha coefficient of 0.68 (95% CI: 0.53–0.83) was found (the "bad guy" item was reverse-scored). Item reliability showed that all five items were again correlated with the total score, with *r* values between 0.62–0.75 for all items except the item asking about the robot, which had an *r* value of 0.56.

Table 6. Frequency of disclosure types children made at the pretest and posttest. Not all children disclosed information, and some disclosed more than one piece of information. Here, "Discl." = "Disclosure", "F. Mot." = "Fine Motor Skills"; "Phys." = "Physical Skills"; "Social" = "Social Skills", "Cog." = "Cognitive Skills", "N.S." = "Not specified".

Time	Discl.	Phys.	F. Mot.	Social	Cog.	N.S.
pre	1	9	6	3	3	2
post	1	13	6	5	12	3
pre	2	10	4	2	10	0
post	2	14	5	0	8	2

We computed test-retest reliability for each IOS item, which was poor overall and ranged from $r = 0.217$ to $r = 0.389$. This is not necessarily a problem, as we expected change since children do change their opinons about others over time.

Differences over time and by demographics

Social-Relational Interview

Mixed analyses of variance with time (within: pre vs. post), school (between: A, B, or C), and gender (between: male or female) with age as a covariate revealed several significant differences in children's SRI responses. All significant test results are listed in Table 7. A summary of children's responses with respect to time, school, and gender is shown in Table 8. First, there was a significant main effect of gender on whether children said the robot would be sad if another child was mean to it. Post-hoc tests with Tukey's HSD showed that in particular, girls were more likely to say the robot would be sad than boys were.

There was a significant main effect of school on whether children said the robot would be sad if it had no friends. Children at school C were less likely to say the robot could be sad. We also saw a main effect age, with 5-year-olds less likely to say the robot could be sad than 6-year-olds. Regarding whether children thought the robot would help another child, there was a main effect of school and a significant interaction between school and gender. Boys at school C were less likely to say the robot would help compared to all other groups. There was also a main effect of school for the question about whether the robot would cheer up a child who was sad. Children at school C were least likely to say the robot would cheer up another child.

There were significant main effects of gender and age, as well as a significant interaction of school with gender, regarding whether the robot really wanted to be friends. Girls were significantly more likely than boys to say that the robot really wanted to be friends. However, this was only true at schools B and C; at school A, boys were equally as likely to say the robot really wanted to be friends. Finally, 6-year-olds were more likely to think the robot really wanted to be their friend than 5-year-olds.

Table 7. Significant SRI results by time, gender, school, and age.

Question	Effect	df	F	p
Sad if mean	Gender	1,34	10.5	0.003
Sad no friends	School	2,34	3.63	0.037
Sad no friends	Age	3,34	6.38	0.016
Help child	School	2,34	10.4	<0.001
Help child	School*Gender	2,34	3.86	0.031
Cheer up child	School	2,33	3.79	0.033
Wants friends	Gender	1,33	5.17	0.030
Wants friends	Age	3,33	7.82	0.009
Wants friends	School*Gender	2,33	4.79	0.015
Likes you	Time	1,33	6.89	0.013
Likes you	School*Gender	2,32	3.97	0.029
Overall	School	1,33	6.42	0.004
Overall	Gender	1,33	6.27	0.019
Overall	School*Gender	2,33	4.25	0.023

Regarding whether children thought the robot really liked them, there was a significant main effect of time, and a significant interaction of school with gender. At the posttest, children were less likely to think that the robot really liked them. Boys at schools B and C were least likely to say the robot really liked them, but girls at these schools said the robot really liked them. Boys at school A said the robot really liked them more than girls at school A. Finally, there were no significant differences by groups for the question about sharing secrets.

When looking at the composite SRI score, there were significant main effects of school and gender, as well as a significant interaction of school with gender. Overall, girls rated the robot more highly as a social relational agent than did boys. Children at school C rated the robot less highly than children at schools A and B. The interaction revealed that boys at schools B and C rated the robot less highly than other children; boys at school A rated the robot as highly as girls did at all three schools.

IOS Task

Mixed analyses of variance with time (within: pre vs. post), school (between: A, B, or C), and gender (between: male or female) with age as a covariate revealed several differences in children's ratings. Table 9 lists descriptive statistics for each item for each group. There was a significant main effect of school on children's ratings of their best friends, $F(2,33) = 6.65$, $p = 0.004$. Children at schools A and C rated their best friends higher than children at school B.

With regards to the bad guy, there was a main effect of school, $F(2,31) = 6.82$, $p = 0.004$. Children at school C rated the bad guy more favorably than children at school A and school B. There was a significant interaction of time with gender, $F(1,32) = 7.16$, $p = 0.011$. This interaction showed that girls tended to rate the bad guy more favorably than boys during the pretest, and that their ratings decreased from the pretest to the posttest such that they rated the bad guy the same as boys during the posttest. This appeared to be driven by girls at school C, as suggested by the significant interaction of time, school, and gender, $F(2,32) = 7.97$, $p = 0.002$. These girls rated the bad guy more highly than other children, though their ratings decreased over time.

We saw significant main effects of school, $F(2,34) = 6.40$, $p = 0.004$ and age, $F(1,34) = 4.56$, $p = 0.04$, on children's ratings of their parent. Children at school A rated their parent more highly than children at schools B and C. Six-year-olds rated their parent more highly than 5-year-olds.

There was a trend toward a main effect of time on children's ratings of the robot, $F(1,35) = 3.01$, $p = 0.092$. Children's ratings were marginally higher during the posttest.

Narrative Description

Mixed analyses of variance with time (within: pre vs. post), agent (within: robot vs. friend), school (between: A, B, or C), and gender (between: male or female) with age as a covariate revealed several significant differences. For word count, there was a significant interaction of agent with gender for both word count, $F(1,75) = 6.10$, $p = 0.016$, and sentence count, $F(1,75) = 6.39$, $p = 0.014$. Girls gave longer descriptions of their friend than of the robot, while boys' descriptions did not differ significantly in length. While non-significant, we observed a trend for children to use more words to describe their friend than to describe the robot.

Self-disclosure Task

We performed mixed analyses of variance on the lengths of children's disclosures with time (within: pre vs. post), school (between: A, B, or C), and gender (between: male or female) with age as a covariate. There was a significant main effect of time on the word count of children's responses after the robot's first prompt, $F(1,36) = 4.51$, $p = 0.040$. Children used more words at the posttest than at the pretest. There was a significant main effect of school on word count for both the first prompt, $F(2,35) = 3.48$, $p = 0.042$, and the second prompt, $F(2,35) = 5.47$, $p = 0.009$. Children gave longer responses at school B than at schools A or C. There were no differences in sentence count.

When looking at the total number of disclosures made by each child, we found significant main effects of time, F(1,38) = 4.37, p = 0.043, and school, F(2,37) = 3.45, p = 0.042. Children disclosed more pieces of information at the posttest ($M = 1.57$, $SD = 1.52$) than at the pretest ($M = 1.11$, $SD = 0.92$). Children at school C disclosed less information ($M = 0.80$, $SD = 0.96$) than at schools A ($M = 1.52$, $SD = 1.22$) or B ($M = 1.73$, $SD = 1.46$).

DISCUSSION

In this paper, we presented four assessments for measuring children's relationships with social robots. In the following, we first discuss the reliability and validity results, which indicate that the design of the assessments were age-appropriate for children aged 5–6 years. The assessments were able to capture some long-term relationship adjustments between children and the robot. Then, relating to the second goal of this paper, we discuss the differences in children's responses with regards to age, gender, and school, which show that the assessments can capture individual differences and can provide a point of comparision for future work.

Table 8. Descriptive statistics by school, gender, and time for the SRI.

Question	School	Girls		Boys	
		Pretest *Mean (SD)*	Posttest *Mean (SD)*	Pretest *Mean (SD)*	Posttest *Mean (SD)*
Sad if mean	A	1.75 (0.71)	2.00 (0.00)	1.50 (0.93)	2.00 (0.00)
	B	2.00 (0.00)	1.78 (0.67)	1.50 (1.00)	1.00 (1.15)
	C	2.00 (0.00)	2.00 (0.00)	0.88 (0.99)	1.14 (1.07)
Sad no friends	A	1.75 (0.71)	1.88 (0.35)	2.00 (0.00)	2.00 (0.00)
	B	2.00 (0.00)	1.67 (0.71)	2.00 (0.00)	1.50 (1.00)
	C	1.57 (0.79)	2.00 (0.00)	1.12 (0.99)	1.14 (1.07)
Help child	A	1.75 (0.71)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)
	B	2.00 (0.00)	1.78 (0.67)	1.00 (1.15)	2.00 (0.00)
	C	1.71 (0.76)	1.60 (0.89)	0.75 (1.04)	0.86 (0.90)
Share secret	A	1.14 (1.07)	0.50 (0.93)	0.88 (0.99)	0.62 (0.92)
	B	1.33 (1.00)	1.00 (1.00)	0.50 (1.00)	0.50 (1.00)
	C	0.86 (1.07)	1.60 (0.89)	1.00 (1.07)	1.71 (0.76)
Cheer up child	A	1.43 (0.98)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)
	B	1.78 (0.67)	2.00 (0.00)	1.50 (1.00)	1.50 (1.00)
	C	1.71 (0.76)	1.60 (0.89)	0.75 (1.04)	1.43 (0.98)
Wants friend	A	1.43 (0.98)	1.75 (0.71)	2.00 (0.00)	1.50 (0.93)
	B	1.56 (0.88)	1.78 (0.67)	1.50 (1.00)	1.50 (1.00)
	C	2.00 (0.00)	2.00 (0.00)	0.25 (0.71)	0.86 (1.07)
Likes you	A	1.71 (0.76)	1.25 (0.89)	2.00 (0.00)	1.75 (0.71)
	B	2.00 (0.00)	1.56 (0.88)	1.50 (1.00)	1.00 (1.15)
	C	2.00 (0.00)	1.60 (0.89)	1.14 (1.07)	0.86 (1.07)
Total	A	10.9 (3.02)	11.4 (1.06)	12.4 (1.51)	11.9 (2.03)
	B	12.7 (1.73)	11.6 (2.13)	9.50 (3.00)	9.00 (5.03)
	C	11.7 (1.46)	10.3 (5.57)	6.00 (3.87)	8.00 (4.12)

We found that children could easily respond to the SRI and IOS assessments in appropriate ways. They used all answer options and gave a range of responses, with the majority picking responses that made sense (i.e., giving the "bad guy" a low IOS score and giving parents and friends high scores). On the IOS, children's ratings were not completely stable from pretest to posttest, as indicated by the low test-retest reliability. However, this is not necessarily a problem, as children do change their opinons about others over time. In particular, we expected they might change their rating of the robot. Their scores may also have been different at each test instance as a result of not thinking of the same best friend or same bad guy each time. They may also have been influenced by local events regarding their parent, pet, or friend; some days they may feel very close, and other days may feel less close. Finally, the children may have understood the IOS scale better during the posttest due to its familiarity.

The SRI and IOS both had reasonably high internal reliability. However, due to the low number of participants, the reliability results should be interpreted cautiously. For the SRI, we recommend computing a composite SRI score consisting of the sum of all the item scores to indicate children's overall view of the robot as a social-relational other. Furthermore, the sharing secrets question should be revised to improve its reliability. This question may have been less reliable because some children may be taught at home or at school that it is not okay to keep secrets—as we saw from children's explanations—and thus, sharing secrets is not a behavior they engage in with friends. We suggest replacing this question with a new item, "Let's pretend something really good or really bad happened to the robot. Would the robot not care about telling anyone, or would the robot want to tell a friend?" This new item may achieve the same goal of targeting intimacy/self-disclosure, but will need to be tested for reliability. We also saw that for the item regarding helping, some children had said the robot could not help because it could not move. This item should specify that the robot need not move in order to help.

For the Narrative Description and Self-disclosure tasks, children generally provided descriptions and disclosures when prompted. However, the Narrative Description required more initial prompting to help children think of a friend to talk about than was initially included in the protocol, such as "What's your friend's name?" and "Is your friend a boy or a girl?". The experimenters also mirrored back children's phrases to encourage them to say more. We suggest adding these prompts to the protocol. For the Self-disclosure Task, when the robot asked "Can you tell me…" in its prompt, many children responded with either "yes" or "no," rather than disclosing information. We suggest revising the prompt

Table 9. Descriptive statistics by groups for the IOS task.

Question	School	Girls		Boys	
		Pretest *Mean (SD)*	Posttest *Mean (SD)*	Pretest *Mean (SD)*	Posttest *Mean (SD)*
Best Friend	A	5.12 (1.96)	5.38 (2.26)	6.38 (0.74)	6.12 (0.83)
	B	4.00 (1.94)	4.50 (1.93)	2.00 (1.15)	4.00 (2.16)
	C	5.14 (1.77)	5.00 (1.87)	5.00 (1.31)	4.57 (2.51)
Parent	A	5.62 (1.41)	5.50 (1.51)	6.25 (1.49)	6.50 (0.53)
	B	4.78 (1.39)	4.56 (2.07)	2.75 (0.96)	5.50 (1.91)
	C	4.57 (2.30)	5.20 (2.68)	4.00 (2.14)	4.29 (2.14)
Pet or toy	A	4.62 (2.07)	5.38 (1.69)	6.00 (1.41)	5.62 (1.30)
	B	3.50 (1.51)	4.89 (2.20)	4.50 (1.91)	4.50 (2.38)
	C	5.00 (0.00)	5.20 (1.92)	4.50 (1.60)	4.14 (2.19)
Bad guy	A	1.50 (1.41)	1.43 (1.62)	1.00 (0.00)	1.00 (0.00)
	B	1.89 (1.54)	1.50 (0.76)	1.75 (0.50)	1.25 (0.50)
	C	3.83 (2.48)	2.20 (1.30)	1.38 (0.74)	3.14 (2.27)
Robot	A	4.50 (2.33)	5.12 (1.96)	5.38 (1.51)	5.50 (1.41)
	B	4.56 (1.67)	4.56 (2.13)	3.25 (0.96)	5.25 (2.06)
	C	4.57 (2.07)	5.80 (1.64)	4.38 (1.30)	4.71 (1.80)

to begin with “What about you? What are things. . . ” to make the question more open-ended. We also recommend adding a secondary prompt, “Tell me more,” to encourage children to share more. Furthermore, the pause after each of the robot’s disclosures tended to lead children to ask if the robot was not working right rather than leading them to spontaneously disclose information. We also observed that children’s responses to the first prompt were shorter; it may be that children’s confusion led them to answer more briefly. Thus, we suggest either removing the pause to help children better understand the task or have the robot signal nonverbally that it is thinking, e.g., by sighing or saying "hmm".

Regarding the age appropriateness of the assessments, the majority of 5- and 6-year-olds rated the robot similarly. 6-year-olds rated the robot more socially than 5-year-olds on two SRI items; they also rated their parent more highly on the IOS task. These differences may relate to children’s developing social and friendship skills [10, 15, 39, 40]. We expect that if more children were tested in a wider age range, we would see a variety of developmental differences.

The assessments were able to capture several changes over the two-month interaction in children’s ratings and descriptions of their relationship with the robot. During the pretest, all assessments indicated that even after just one session, children viewed the robot as a friend-like social, relational agent. Their scores for the robot on the IOS task indicated that they felt the robot was as close as a friend or a pet. Children disclosed information to the robot, and some described it at as much length as they did a friend. The SRI showed that they thought the robot felt the same way they did about wanting friendship (such as being sad without friends) and about taking friendship actions (such as helping another child). Children’s responses to the SRI follow-up questions indicated that they generally thought the robot would feel the same way they would, and they frequently referenced the social and relational qualities of the robot when justifying their answers, such as its feelings, attributes, and moral obligations. The few children who said the robot did not really want to be friends indicated that they thought the robot was pretending; i.e., incapable of being a friend due to its robotic nature. This is similar to what has been found in prior work in which social robots were viewed as social agents, not like artifacts, but also not quite like people [17, 22, 30].

In the posttest, children felt they were more close to the robot (IOS), described the robot and their best friend at a more similar length (Narrative Description), and disclosed more information (Self-disclosure). In the SRI, children’s opinions on some items changed over time, suggesting that the interaction with the robot affected how they perceived it and that the SRI was capable of capturing that change. However, not all of these patterns were statistically significant. This could be for several reasons. The seven sessions with the robot may not have been sufficient for measureable change to occur, since relationships can develop slowly, and frequency of contact is one factor influencing children’s perception of friendship [33, 40]. A longer timeframe may lead to more change in children’s responses. The robot also did not take many actions explicitly toward building a relationship—e.g., continuity behaviors such as talking about what they did while apart, referencing shared experiences, or prosocial helping behaviors [3, 10, 27]. We expect that a robot that performs more relationship-building actions will lead to greater change in the perception of the relationship over time. Finally, the assessments may not be measuring relevant aspects of children’s relationships. Since these assessments only targeted some of children’s friendship behaviors, it may be that other behaviors will be more telling.

As expected, we saw numerous differences between genders and schools. This suggests that the assessments can capture some individual differences in friendships. The gender differences we observed, in which girls generally rated the robot's social nature more highly than boys, may reflect children's real friendships. This is in line with prior work that has found that girls' ratings of intimacy and alliance in their friendships tend to be higher than boys' [7, 10].

We observed several interactions between gender and school, with boys at the higher-SES school A responding more similarly to girls at school A than boys at the lower-SES schools B and C. Children's individual backgrounds likely influenced the types of gender roles and gendered opinions the children held; SES may also be related to the development of social-emotional skills and friendships [5, 32]. We observed that children at school A gave higher ratings for all their relationships than children at the other schools. Children at schools B and C were less likely to say that the robot would help another child, be sad if it had no friends, and that it did want to be their friend. Differences in children's ethnicity, socioeconomic backgrounds, technology use at home and at school may all have influenced children's level of comfort with the robot and their perception of it as a relational agent. However, given the small sample size, it is difficult to generalize from these results.

LIMITATIONS AND FUTURE WORK

We should note several limitations of this work. First, the pilot study included a small number of participants with an unequal number of children at each age and of each demographic group. This is due to the nature of long-term Human-Robot Interaction (HRI) studies, in which it can be challenging to recruit large numbers of participants. Thus, in future work it will be important to test these assessments with a larger population of children that is balanced across age and demographics. As a result of these imbalances, the analyses reported here may be under-powered.

In addition, an experimenter was present throughout the Self-disclosure Task. It may be that the experimenter's presence led children to disclose less information or engage less fully in the conversation with the robot. Because multiple experimenters administered the study, there may have been slight differences in their administration of the assessments, e.g., mirroring back the children's responses differently.

The assessments developed so far also have several limitations. First, they are not continuous. Future work should investigate measures that can be used every session with a robot, or even multiple times throughout a session. This would allow researchers to build better relationship models and create robots that personalize in real-time to children's developing relationships. Some, such as the Self-disclosure Task questions, were administered as part of a conversation that children had with a robot using automatic speech recognition, but the system was not able to provide real-time context understanding. As such, the robot was not able to determine whether the child's response was related to the question the robot asked. A natural language context model could be implemented for the robot in order to guide the child's response via an expected language structure.

However, despite these limitations, this work is an important contribution to HRI. The assessments we are developing and adapting will be useful tools for other researchers who wish to assess children's relationships with social robots—or potentially with other social technological agents—during either short or long-term studies. Our goal is to enable child-robot interaction research. Thus, it behooves us to share these assessments for others to use, test, modify, and improve.

SELECTION AND PARTICIPATION OF CHILDREN

We recruited children aged 5–6 years to participate in our pilot study from three Boston-area schools. We recruited from multiple schools because it was not possible to recruit sufficient children from a single school. This had the benefit of allowing us to recruit a diverse population of children (one higher-SES school and two lower-SES schools). We invited all children in the classrooms that contained predominantly 5- or 6-year-olds to participate— forty-four children in total. Children's parents gave written informed consent prior to the start of the study, and all children assented to participate. All children had the opportunity to play a storytelling game with a social robot seven times over the course of two months. The protocol was approved by the MIT Committee On the Use of Humans as Experimental Subjects.

Acknowledgements

This research was supported by an MIT Media Lab Learning Innovation Fellowship and by the National Science Foundation (NSF) under Grant IIS-1523118. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not represent the views of the NSF.

REFERENCES

1. Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596–612. DOI: http://dx.doi.org/10.1037/0022-3514.63.4.596
2. Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2008. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Nov. 2008), 71–81. DOI: http://dx.doi.org/10.1007/s12369-008-0001-3
3. Ellen Berscheid and Harry T. Reis. 1998. Attraction and Close Relationships. In *The Handbook of Social Psychology, Vols. 1-2, 4th Ed*, D. T. Gilbert, S. T. Fiske, and G. Lindzey (Eds.). McGraw-Hill, New York, NY, US, 193–281.
4. Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. DOI:http://dx.doi.org/10.1145/1067860.1067867
5. Robert H. Bradley and Robert F. Corwyn. 2002. Socioeconomic Status and Child Development. *Annual review of psychology* 53, 1 (2002), 371–399.

6. Cynthia Breazeal, Paul L. Harris, David DeSteno, Jacqueline M. Kory Westlund, Leah Dickens, and Sooyeon Jeong. 2016. Young Children Treat Robots as Informants. *Topics in Cognitive Science* (March 2016), 1–11. DOI:http://dx.doi.org/10.1111/tops.12192
7. Duane Buhrmester and Wyndol Furman. 1987. The Development of Companionship and Intimacy. *Child Development* 58, 4 (1987), 1101–1113. DOI: http://dx.doi.org/10.2307/1130550
8. Christine T. Chambers and Charlotte Johnston. 2002. Developmental Differences in Children's Use of Rating Scales. *Journal of Pediatric Psychology* 27, 1 (Jan. 2002), 27–36. DOI: http://dx.doi.org/10.1093/jpepsy/27.1.27
9. Simon Gächter, Chris Starmer, and Fabio Tufano. 2015. Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale. *PLOS ONE* 10, 6 (June 2015), e0129478. DOI: http://dx.doi.org/10.1371/journal.pone.0129478
10. Tracy R. Gleason and Lisa M. Hohmann. 2006. Concepts of Real and Imaginary Friendships in Early Childhood. *Social Development* 15, 1 (Feb. 2006), 128.
11. Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children's Second Language Skill. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Palo Alto, CA.
12. Willard W. Hartup, Brett Laursen, Mark I. Stewart, and Amy Eastenson. 1988. Conflict and the Friendship Relations of Young Children. *Child Development* 59, 6 (1988), 1590–1600. DOI: http://dx.doi.org/10.2307/1130673
13. Adam O. |Greenberg Horvath. 1989/00/00. Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology* 36, 2 (1989/00/00), 223–33.
14. Sooyeon Jeong, Deirdre E. Logan, Matthew S. Goodwin, Suzanne Graca, Brianna O'Connell, Honey Goodenough, Laurel Anderson, Nicole Stenquist, Katie Fitzpatrick, and Miriam Zisook. 2015. A Social Robot to Mitigate Stress, Anxiety, and Pain in Hospital Pediatric Care. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, 103–104.
15. Jennifer L. Jipson and Susan A. Gelman. 2007. Robots and Rodents: Children's Inferences about Living and Nonliving Kinds. *Child development* 78, 6 (2007), 1675–1688.
16. Peter H. Kahn, Heather E. Gary, and Solace Shen. 2013. Children's Social Relationships With Current and Near & Future Robots. *Child Development Perspectives* 7, 1 (2013), 32–37.
17. Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, You'll Have to Go into the Closet Now": Children's Social and Moral Relationships with a Humanoid Robot. *Developmental psychology* 48, 2 (2012), 303.
18. James Kennedy, Paul Baxter, and Tony Belpaeme. 2017. Nonverbal Immediacy as a Characterisation of Social Behaviour for Human–Robot Interaction. *International Journal of Social Robotics* 9, 1 (Jan. 2017), 109–128. DOI:http://dx.doi.org/10.1007/s12369-016-0378-3
19. Cory D. Kidd and Cynthia Breazeal. 2008. Robots at Home: Understanding Long-Term Human-Robot Interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference On*. IEEE, 3230–3235.
20. Jacqueline Kory. 2014. *Storytelling with Robots: Effects of Robot Language Level on Children's Language Learning*. Master's Thesis. Massachusetts Institute of Technology, Cambridge, MA.
21. Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with Robots: Learning Companions for Preschool Children's Language Development. In *2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. 643–648. DOI:http://dx.doi.org/10.1109/ROMAN.2014.6926325
22. Jacqueline M. Kory Westlund and Cynthia Breazeal. 2015. The Interplay of Robot Language Level with Children's Language Learning During Storytelling. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI'15 Extended Abstracts)*. ACM, New York, NY, USA, 65–66. DOI: http://dx.doi.org/10.1145/2701973.2701989
23. Jacqueline M. Kory Westlund, Leah Dickens, Sooyeon Jeong, Paul L. Harris, David DeSteno, and Cynthia L. Breazeal. 2017a. Children Use Non-Verbal Cues to Learn New Words from Robots as Well as People. *International Journal of Child-Computer Interaction* (2017). DOI: http://dx.doi.org/10.1016/j.ijcci.2017.04.001
24. Jacqueline M. Kory Westlund, Sooyeon Jeong, Hae Won Park, Samuel Ronfard, Aradhana Adhikari, Paul Lansley Harris, David DeSteno, and Cynthia Breazeal. 2017b. Flat versus Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative. *Frontiers in Human Neuroscience* 11 (2017). DOI: http://dx.doi.org/10.3389/fnhum.2017.00295
25. Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2017c. Measuring Children's Long-Term Relationships with Social Robots. *Workshop on Perception and Interaction dynamics in Child-Robot Interaction, held in conjunction with the Robotics: Science and Systems XIII.* (2017).

26. Ivana Kruijff-Korbayová, Elettra Oleari, Anahita Bagherzadhalimi, Francesca Sacchitelli, Bernd Kiefer, Stefania Racioppa, Clara Pozzi, and Alberto Sanna. 2015. Young Users' Perception of a Social Robot Displaying Familiarity and Eliciting Disclosure. In *Social Robotics*. Springer, Cham, 380–389. DOI: http://dx.doi.org/10.1007/978-3-319-25554-5_38

27. Gary W. Ladd, Becky J. Kochenderfer, and Cynthia C. Coleman. 1996. Friendship Quality as a Predictor of Young Children's Early School Adjustment. *Child Development* 67, 3 (June 1996), 1103–1118. DOI: http://dx.doi.org/10.1111/j.1467-8624.1996.tb01785.x

28. Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics* 5, 2 (April 2013), 291–308. DOI: http://dx.doi.org/10.1007/s12369-013-0178-y

29. I. Leite, C. Martinho, A. Pereira, and A. Paiva. 2009. As Time Goes by: Long-Term Evaluation of Social Presence in Robotic Companions. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*. 669–674. DOI: http://dx.doi.org/10.1109/ROMAN.2009.5326256

30. Gail F. Melson, Jr. Kahn, Peter H., Alan Beck, and Batya Friedman. 2009. Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues* 65, 3 (Sept. 2009), 545–567. DOI: http://dx.doi.org/10.1111/j.1540-4560.2009.01613.x

31. Morton J. Mendelson and Frances E. Aboud. 1999. Measuring Friendship Quality in Late Adolescents and Young Adults: McGill Friendship Questionnaires. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 31, 2 (1999), 130.

32. Charles Murray. 2013. *Coming Apart: The State of White America, 1960-2010*. Three Rivers Press.

33. Andrew F. Newcomb and Catherine L. Bagwell. 1995. Children's Friendship Relations: A Meta-Analytic Review. *Psychological Bulletin* 117, 2 (1995), 306–347. DOI:http://dx.doi.org/10.1037/0033-2909.117.2.306

34. Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Altered Attitudes of People toward Robots: Investigation through the Negative Attitudes toward Robots Scale. In *Proc. AAAI-06 Workshop on Human Implications of Human-Robot Interaction*. 29–35.

35. Stephen Nowicki and Marshall P. Duke. 1974. A Preschool and Primary Internal-External Control Scale. *Developmental Psychology* 10, 6 (1974), 874–880. DOI: http://dx.doi.org/10.1037/h0037253

36. Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017a. Telling Stories to Robots: The Effect of Backchanneling on a Child's Storytelling. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 100–108.

37. Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. 2017b. Growing Growth Mindset with a Social Robot Peer. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 137–145. DOI: http://dx.doi.org/10.1145/2909824.3020213

38. Janet C. Read. 2008. Validating the Fun Toolkit: An Instrument for Measuring Children's Opinions of Technology. *Cognition, Technology & Work* 10, 2 (April 2008), 119–128. DOI: http://dx.doi.org/10.1007/s10111-007-0069-9

39. Ken J. Rotenberg. 1995. Development of Children's Restrictive Disclosure to Friends. *The Journal of Genetic Psychology* 156, 3 (Sept. 1995), 279–292. DOI: http://dx.doi.org/10.1080/00221325.1995.9914823

40. Kenneth H. Rubin, William M. Bukowski, and Jeffrey G. Parker. 1998. Peer Interactions, Relationships, and Groups. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. DOI: http://dx.doi.org/10.1002/9780470147658.chpsy0310

41. Sofia Serholt and Wolmet Barendregt. 2016. Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, 64:1–64:10. DOI: http://dx.doi.org/10.1145/2971485.2971536

42. Elaine Short, Katelyn Swift-Spong, Jillian Greczek, Aditi Ramachandran, Alexandru Litoiu, Elena Corina Grigore, David Feil-Seifer, Samuel Shuster, Jin Joo Lee, Shaobo Huang, and others. 2014. How to Train Your Dragonbot: Socially Assistive Robots for Teaching Children about Nutrition through Play. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium On*. IEEE, 924–929.

43. Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. *Journal of Human-Robot Interaction* 1, 1 (2012), 78–95.

44. Lise M. Youngblade, Kathryn A. Park, and Jay Belsky. 1993. Measurement of Young Children's Close Friendship: A Comparison of Two Independent Assessment Systems and Their Associations with Attachment Security. *International Journal of Behavioral Development* 16, 4 (Dec. 1993), 563–587. DOI:http://dx.doi.org/10.1177/016502549301600403