Measuring Children's Long-Term Relationships with Social Robots

Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal

Personal Robots Group, MIT Media Lab

20 Ames St., Cambridge, MA 02139

Email: {jakory,haewon,randiw12,cynthiab}@media.mit.edu

Abstract—Social robots are being increasingly developed for long-term interactions with children. However, there are few validated assessments for measuring young children's relationships with social robots. In this paper, we discuss a variety of relational assessments that could be used in this context. We present a pilot study of two assessments, the Inclusion of Other in Self task and a Social-Relational Interview, that we have adapted for use with children aged 3–7. We show that children can appropriately respond to these assessments and that both have high internal reliability.

I. INTRODUCTION

Social robots are increasingly being developed for use with children in application domains such as education, entertainment, healthcare, and therapy [9, 15, 18, 23]. 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, for example, ascribing them mental states, psychological attributes, and moral standing [13, 15, 20]. Furthermore, in long-term interactions, social robots are taking on a *relational* role—that is, they are situated as agents that actively attempt to build and maintain long-term social-emotional relationships [2]. They are introduced as peers, tutors, and learning companions [15, 23]. 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 measure these relationships.

Measuring children's relationships with robots will not only give us insight into how children think about robots through time, but also will 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 trust before it can effectively administer an intervention, or, alternatively, whether the child has become too attached, and thus, that the robot needs to recommend that the child seek out a person for help instead. Prior work has accomplished this with adults [14], 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.

In this paper, we specifically focus on measuring relationships with young children aged 3–7 years. Assessments for children this young can be especially difficult to craft because, e.g., the children may be pre-reading, may have short attention spans, and cannot fill out standard Likert-style questionnaires [4]. We explain several assessments that we have adapted for this age group below. The full instructions for each task are available on figshare: 10.6084/m9.figshare.5047657. We also briefly review additional assessments that could prove useful that we have not yet tested with this population.

II. ASSESSMENTS

A. 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]. 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.

B. 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—e.g., questions about whether children attribute certain properties to robots, such as the questions from [12] used in [15]—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 [8, 17]. Two questions asked about whether the robot was genuine, i.e., whether what it felt was real or whether it was just pretending. Each question offered three responses: yes, the robot would feel something (e.g., sad or happy), maybe / don't know, and 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 would have some context for understanding children's responses.

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

C. Narrative Description

In this task, a puppet asks the child to help it learn about people and robots. 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 best 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., after all the sessions versus after one session).

D. Targeted Self-disclosure

Because self-disclosure is one of the features of children's friendships [3, 8, 17, 21], we had the robot disclose information and prompt for information disclosure in return. The protocol was adapted from [21]. We expected that children would disclose more when the relationship was closer (e.g., more during a posttest than a pretest). As per [21], the amount of disclosure can be measured by counting the number of utterances made. A more detailed analysis might additionally code the kind of information disclosed.

E. Additional Assessments

We have begun investigating numerous additional assessments. For example, we could code children's speech transcripts for phrase matching [16], language style matching [11] or other kinds of linguistic markers of relationships and rapport. The Comfortable Interpersonal Distance scale has already been adapted for preschool children, and can give a measure of children's preferences for social distances [5].

Other behavioral measures such as short "scenarios" may be useful. For example, children may resolve conflicts using different strategies with friends versus other peers (disengaging, negotiating or bargaining, and reaching an equal solution versus standing firm, and reaching a winner/loser outcome) [10]. One scenario could involve the robot instigating

a conflict (such as a disagreement over the next activity, who should go first, or who gets which sticker), and we could see how children respond. Prior child-robot interaction work has successfully used scenarios, such as placing a robot in a closet, to investigate children's moral conceptions about robots [13].

If one desires a test in which children self-report their social competence, perhaps to get a baseline of children's abilities so one can control for their differing social competence when evaluating their relationship with the robot, one could use or adapt the Berkeley Puppet Inventory [19]. In this inventory, the experimenter has two puppets and tells the child that each puppet will say something about themselves. The puppets each anchor one end of a scale, such as "I'm shy when I meet new people" versus "I'm not shy when I meet new people". Then they say, "we want to learn about you." The child can describe themselves in relation to the two puppets.

Another self-report that may be useful is the Social Acceptance Scale [7]. In this scale, children are asked yes/no/maybe questions (with a visual scale of 3 smiley faces for children to point at) about their acceptance of peers with disabilities. Since robots generally have numerous limitations, which could be viewed as disabilities, it may be useful to adapt this scale to ask about children's acceptance of robots or technologies that are "disabled" (e.g., a robot that has trouble hearing, given the fact that automatic speech recognition is often subpar for young children).

Other tasks that have been used with young children include drawing activities, such as asking children to draw two pictures about two points in time (i.e., a differential), such as "When I first started kindergarten I..." versus "Now, I... [6]. In this kind of task, one looks not only at what children draw, but at what children say while they are drawing—i.e., looking at children's meaning-making as a process involving both drawing and narrating their drawing. However, this kind of task tends to be time-consuming, with children sometimes taking as long as 10–15 minutes to produce their drawings.

III. PILOT STUDY

A. Methodology

We are performing a pilot test of several of these assessments during a long-term child-robot interaction study at three Boston-area schools. Forty-four children aged 4–7 (M = 5.4, SD = 0.66) are interacting with a fully autonomous social robot, Tega, approximately 1–2 times a week, for a total of 10 sessions (Figure 1). The robot tells stories and children are asked to retell the stories. There are 16 children (8 F, 8 M) from school A, 13 children (9 F, 4 M) from school B, and 15 children (7 F, 8 M) from school C.

We administered the IOS task, Narrative Description, and SRI after children's first session with the robot. Due to its length, the Targeted Self-disclosure was implemented as part of a conversation at the start of the second session. All of these assessments will be administered a second time after the final session. The IOS task will also be administered at the midway point after half the sessions have been completed.

Below, we present preliminary results for the IOS task and the SRI.

B. Results

1) 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 to chance levels of responding (i.e., mean of 1). The results are shown in Table I. Children's responses differed from chance in the expected ways: overall, children said that the robot had friend-like qualities, in that it would be sad if another child was mean to it or if it had no friends, help another child who needed help, and cheer up another 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 asking whether the robot would prefer to share a secret with a friend did not differ from chance levels.

TABLE I
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.

Question	Mean (SD)	df	t-value	p-value
Sad if child is mean	1.62 (0.78)	38	4.92	< 0.001
Sad if no friends	1.74 (0.64)	38	7.29	< 0.001
Help another child	1.59 (0.82)	38	4.50	< 0.001
Cheer up another child	1.58 (0.83)	37	4.32	< 0.001
Really does want friends	1.53 (0.86)	37	3.77	< 0.001
Really does like you	1.84 (0.55)	36	9.21	< 0.001
Want to share secret	0.95 (1.00)	36	-0.33	0.744

One-way analyses of variance over children's age (5- and 6-year-olds only, because there were not enough 4- or 7-year-olds to constitute their own groups) revealed one main effect of age. Six-year-olds ($M = 2.00$, $SD = 0.00$) were more likely to say that the robot would be sad if it had no friends than five-year-olds ($M = 1.50$, $SD = 0.82$), $F(1,33) = 7.17$, $p = 0.011$, $\eta^2 = 0.178$.

Separate analyses with gender x school that included all children revealed several significant main effects of both gender and school. Post-hoc tests with Tukey's HSD showed that in particular, girls were more likely to say the robot would be sad if another kid was mean to it ($M = 1.90$, $SD = 0.44$) than boys were ($M = 1.28$, $SD = 0.96$), $F(1,33) = 6.64$, $p = 0.015$, $\eta^2 = 0.151$. Girls were more likely to say the robot liked them ($M = 2.00$, $SD = 0.00$) than boys were ($M = 1.65$, $SD = 0.79$), $F(1,31) = 6.17$, $p = 0.019$, $\eta^2 = 0.136$. Girls were also more likely to say the robot would help another child ($M = 1.81$, $SD = 0.60$), more than boys ($M = 1.33$, $SD = 0.97$), $F(1,33) = 5.30$, $p = 0.028$, $\eta^2 = 0.100$. However, there was also a significant interaction of gender and school, $F(2,33) = 4.11$, $p = 0.025$, $\eta^2 = 0.156$. Boys at school C ($M = 0.67$, $SD = 1.03$) were far less likely than both boys at school A ($M = 2.00$, $SD = 0.00$) and girls at school B ($M = 2.00$, $SD = 0.00$) to say the robot would help. The others were in between.

Regarding whether children thought the robot really wanted to be their friend, there were main effects of both gender, $F(1,32) = 12.78$, $p = 0.001$, $\eta^2 = 0.137$; and school, $F(2,32) = 8.09$, $p = 0.001$, $\eta^2 = 0.174$; as well as an interaction, $F(2,32) = 16.0$, p < 0.001, $\eta^2 = 0.344$. Post-hoc tests showed that girls were more likely to say the robot really wanted to be their friend ($M = 1.80$, $SD = 0.62$) than boys were ($M = 1.22$, $SD = 1.00$). Children at school A ($M = 1.85$, $SD = 0.55$) were also more likely to say the robot really wanted to be their friend than children at school C ($M = 1.08$, $SD = 1.04$), with school B ($M = 1.67$, $SD = 0.78$) in between. The interaction revealed that boys at school C ($M = 0.00$, $SD = 0.00$) were less likely to the robot wanted to be their friend than boys at school A ($M = 2.00$, $SD = 0.00$) or girls at any school.

There was an interaction of school and gender with regards to whether children thought the robot would help cheer up a sad child, $F(2,32) = 5.42$, $p = 0.009$, $\eta^2 = 0.223$. Boys at school C ($M = 0.67$, $SD = 1.03$) were less likely to think the robot would help than boys at school A ($M = 2.00$, $SD = 0.00$) or girls at school B ($M = 2.00$, $SD = 0.00$). Finally, children at school C ($M = 1.46$, $SD = 0.87$) were also less likely to say the robot would be sad if it had no friends than children at school B ($M = 2.00$, $SD = 0.00$), while School A was in between ($M = 1.57$, $SD = 0.852$) $F(2,33) = 3.66$, $p = 0.037$, $\eta^2 = 0.180$.

The reliability of the SRI was determined by measuring the internal consistency of the seven core questions using Cronbach's alpha. An alpha coefficient of 0.70 (95% CI: 0.48–0.88) was found. Item reliability was calculated through an item analysis, which revealed that all seven questions were correlated with the total score, with r values between 0.52–0.85 for all but one item. If we dropped the question about sharing secrets ($r = 0.30$), the reliability would improve to 0.75.

2) 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. The results are shown in Table II. 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.

One-way analyses of variance revealed no differences by age. Separate analyses of gender x school revealed a main effect of gender on children's ratings of the bad guy, $F(1,28) = 4.44$, $p = 0.044$, $\eta^2 = 0.101$. Boys' ratings ($M = 1.40$, $SD = 0.63$) were lower than girls ($M = 2.26$, $SD = 2.02$). There was also a main effect of school for children's ratings of their best friend, $F(2,29) = 4.51$, $p = 0.020$, $\eta^2 = 0.205$. Children at school B's ratings ($M = 3.55$, $SD = 1.97$) were significantly lower than both school A ($M = 5.60$, $SD = 1.78$) and school C ($M = 5.00$, $SD = 1.52$).

The reliability of the IOS task was determined by measuring the internal consistency of the seven core questions using Cronbach's alpha. An alpha coefficient of 0.70 (95% CI: 0.48–0.92) was found (the "bad guy" item was reverse-scored).

TABLE II
CHILDREN'S OVERALL IOS RESPONSES. ALL DIFFERED SIGNIFICANTLY FROM CHANCE (MEAN = 3.5), AS SHOWN BY ONE-SAMPLE T-TESTS.

Question	Median	Mode	Range	Inter-quartile Range	Mean (SD)	df	t-value	p-value
Best Friend	5	7	1-7	3.5	4.71 (1.89)	34	3.81	< 0.001
Parent	5	7	1-7	3.5	4.63 (1.97)	34	3.39	0.002
Pet or toy	5	5	1-7	2.5	4.71 (1.66)	30	4.06	< 0.001
Bad guy	1	1	1-7	1.00	1.88 (1.61)	33	-5.86	< 0.001
Robot	4	7	1-7	3.00	4.54 (1.80)	34	3.42	0.002

Item reliability was calculated through an item analysis, which revealed that all five items were correlated with the total score, with *r* values between 0.61–0.77 for all items.

IV. DISCUSSION

In this paper, we presented several assessments that we have adapted for measuring children's relationships with social robots. In our first pilot test, for the two assessments analyzed so far, we found that children could easily respond to both assessments in appropriate ways, and that both had 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 unreliable because some children may be taught at home or at school that it is not okay to keep secrets, and thus, sharing secrets is not a behavior they engage in with friends. Thus, we suggest replacing this question with a new item, "Let's pretend the robot is really happy or really upset about something. 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.

Both assessments indicated that even after just one session, children viewed the robot as a friend-like social, relational other. Their scores for the robot on the IOS task indicated that they felt the robot was as close as a friend or a pet. However, we have not yet analyzed the follow-up questions that asked children to explain why they chose the answers they did and whether they would feel the same way as the robot. It may be that children who said the robot was not their friend meant they had not spent sufficient time with it yet to consider it a friend, but it could also be that they meant the robot was incapable of being a friend due to its robotic nature. Analyzing children's explanations of their responses may illuminate this.

We saw few age differences, which could perhaps be due to the fact that we could only test differences between 5- and 6-year-olds, since there were insufficient children of other age groups. If more children were tested, we would expect to see developmental differences relating to children's developing social and friendship skills [8, 12, 21, 22]. However, we did see differences by gender and by school, suggesting that the assessments could capture some individual differences in friendships. The gender differences we saw, in which girls rated the robot's social nature more highly than boys, may reflect children's real friendships: prior research has found that girls' ratings of intimacy and alliance in their friendships tend to be higher than boys' [3, 8].

We saw several differences as result of children's schools. In particular, boys at school 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. Furthermore, children's ratings of their best friend in the IOS task were lower at school B than at schools A or C. These results indicate that the population of children was not homogenous across schools, however, without additional data we cannot be sure what caused the difference in children's perceptions of the robot. It may be that the children's socioeconomic backgrounds or the amount of technology generally used in each school influenced children's level of comfort with the robot. School policies discouraging children from having best friends may have influences children's ratings of best friends, in the same way that they may have affected the sharing secrets item.

V. FUTURE WORK

We are in the process of continuing pilot testing of these assessments. Administration of these assessments during posttests will allow us to examine test-retest differences and children's changing perceptions of the robot as a social other over time. We are also currently analyzing the initial Narrative Description and Targeted Self-disclosure task data.

The assessments developed so far 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. These assessments are also not automated. Some, such as the Targeted Self-disclosure questions, can be administered as part of a conversation that children have with a robot, and could potentially be automated given sufficiently good automatic speech recognition or by using automated transcription services, paired with analysis of speech content, or, following Rotenberg's [21] analysis, simple counting of the number of utterances children make.

ACKNOWLEDGEMENTS

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. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596–612, 1992. ISSN 1939-1315 0022-3514. doi: 10.1037/0022-3514.63.4.596.

[2] Timothy W. Bickmore and Rosalind W. Picard. Establishing and Maintaining Long-term Human-computer Relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2):293–327, June 2005. ISSN 1073-0516. doi: 10.1145/1067860.1067867.

[3] Duane Buhrmester and Wyndol Furman. The Development of Companionship and Intimacy. *Child Development*, 58(4):1101–1113, 1987. ISSN 0009-3920. doi: 10.2307/1130550.

[4] Christine T. Chambers and Charlotte Johnston. Developmental Differences in Children's Use of Rating Scales. *Journal of Pediatric Psychology*, 27(1):27–36, January 2002. ISSN 0146-8693. doi: 10.1093/jpepsy/27.1.27.

[5] Marshall P. Duke and Jan Wilson. A Note on the Measurement of Interpersonal Distance in Preschool Children. *The Journal of Genetic Psychology*, 123 (2):361–362, December 1973. ISSN 0022-1325. doi: 10.1080/00221325.1973.10532696.

[6] Johanna Einarsdottir, Sue Dockett, and Bob Perry. Making meaning: Children's perspectives expressed through drawings. *Early Child Development and Care*, 179 (2):217–232, February 2009. ISSN 0300-4430. doi: 10.1080/03004430802666999.

[7] Paddy C. Favazza and Samuel L. Odom. Use of the Acceptance Scale to Measure Attitudes of Kindergarten-Age Children. *Journal of Early Intervention*, 20(3): 232–248, July 1996. ISSN 1053-8151. doi: 10.1177/105381519602000307.

[8] Tracy R. Gleason and Lisa M. Hohmann. Concepts of Real and Imaginary Friendships in Early Childhood. *Social Development*, 15(1):128, February 2006. ISSN 0961205X.

[9] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 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, 2016.

[10] Willard W. Hartup, Brett Laursen, Mark I. Stewart, and Amy Eastenson. Conflict and the Friendship Relations of Young Children. *Child Development*, 59(6):1590–1600, 1988. ISSN 0009-3920. doi: 10.2307/1130673.

[11] Molly E. Ireland, Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker. Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, 22(1):39–44, January 2011. ISSN 0956-7976. doi: 10.1177/0956797610392928.

[12] Jennifer L. Jipson and Susan A. Gelman. Robots and rodents: Children's inferences about living and nonliving kinds. *Child development*, 78(6):1675–1688, 2007.

[13] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303, 2012.

[14] Cory D. Kidd and Cynthia Breazeal. Robots at home: Understanding long-term human-robot interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference On*, pages 3230–3235. IEEE, 2008.

[15] Jacqueline Kory and Cynthia Breazeal. 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*, pages 643–648, 2014. doi: 10.1109/ROMAN.2014.6926325.

[16] Jacqueline Kory Westlund, Sooyeon Jeong, Hae Won Park, Samuel Ronfard, Aradhana Adhikari, Paul Lansley Harris, David DeSteno, and Cynthia Breazeal. Flat versus expressive storytelling: Young children's learning and retention of a social robot's narrative. *Frontiers in Human Neuroscience*, 11, 2017. ISSN 1662-5161. doi: 10.3389/fnhum.2017.00295.

[17] Gary W. Ladd, Becky J. Kochenderfer, and Cynthia C. Coleman. Friendship Quality as a Predictor of Young Children's Early School Adjustment. *Child Development*, 67(3):1103–1118, June 1996. ISSN 1467-8624. doi: 10. 1111/j.1467-8624.1996.tb01785.x.

[18] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, 5(2):291–308, April 2013. ISSN 1875-4791, 1875-4805. doi: 10.1007/s12369-013-0178-y.

[19] Jeffrey R. Measelle, Jennifer C. Ablow, Philip A. Cowan, and Carolyn P. Cowan. Assessing Young Children's Views of Their Academic, Social, and Emotional Lives: An Evaluation of the Self-Perception Scales of the Berkeley Puppet Interview. *Child Development*, 69(6): 1556–1576, December 1998. ISSN 1467-8624. doi: 10.1111/j.1467-8624.1998.tb06177.x.

[20] Gail F. Melson, Jr. Kahn, Peter H., Alan Beck, and Batya Friedman. Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues*, 65(3):545–567, September 2009. ISSN 1540-4560. doi: 10.1111/j.1540-4560.2009.01613.x.

[21] Ken J. Rotenberg. Development of Children's Restrictive Disclosure to Friends. *The Journal of Genetic Psychology*, 156(3):279–292, September 1995. ISSN 0022-1325. doi: 10.1080/00221325.1995.9914823.

[22] Kenneth H. Rubin, William M. Bukowski, and Jeffrey G. Parker. Peer Interactions, Relationships, and Groups. In *Handbook of Child Psychology*. John Wiley & Sons, Inc., 1998. ISBN 978-0-470-14765-8. doi: 10.1002/9780470147658.chpsy0310.

[23] Sofia Serholt and Wolmet Barendregt. 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, pages 64:1–64:10, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4763-1. doi: 10.1145/2971485.2971536.