scientific reports

Check for updates

OPEN

The influence of individual characteristics on children's learning with a social robot versus a human

Allison Langer[1✉], Jason R. Wilson[2], Lauren Howard[2] & Peter J. Marshall[1]

The field of child-robot interaction (CRI) is growing rapidly, in part due to demand to provide sustained, personalized support for children in educational contexts. The present study uses a within-subject design to compare how children between 5 and 8 years of age (n=32) interact with a robot and human instructor during a tangram learning task. To assess how the children's characteristics may influence their behaviors with the instructors, we correlated interaction metrics, such as eye gaze, social referencing, and vocalizations, with parent-reported scales of children's temperament, social skills, and prior technology exposure. We found that children gazed more at the robot instructor and had more instances of social referencing toward a research assistant in the room while interacting with the robot. Age was related to task time completion, but few other individual characteristics were related to behavioral characteristics with the human and robot instructors. When asked about preferences and perceptions of the instructors after completing the tangram tasks, children showed a strong preference for interacting with the robot. These findings have implications for the integration of social technologies into educational contexts and suggest individual differences play a key role in understanding how children will uniquely respond to robots.

Socially assistive robots (SARs) are being designed for educational contexts as tutors and learning aids[1]. As robots and other artificial agents become utilized in classrooms and other learning environments, it is important to understand their effectiveness, and for whom they may work best. Several studies have shown that robots can motivate and encourage learning[2–4]. Many of these prior studies have considered how specific robot features (e.g., physical appearance, behaviors, and personalization abilities) influence interactions with human learners. However, few studies have compared how and whether learners show different patterns of interactions with robots compared to humans[5–7], and there is a particular paucity of this work in young children[8]. Here we examine how children interact with a human and robot instructor during a tangram puzzle task. We focus on differences in children's behaviors towards a robot vs. human instructor, how children's individual characteristics influence their interactions, and their acceptance of both types of instruction. Our findings have implications for the real-world application of robots in classrooms, particularly for children of elementary school age.

Background
Robot vs. human instruction
To assess the potential benefits of including robots in educational contexts, it is important to understand how social robots perform relative to humans in their roles as instructors. Several studies have addressed this question by quantifying children's learning outcomes after interacting with humans, robots, or screens. These studies have broadly suggested that robots can be as effective as human tutors on certain cognitive and affective tasks[5] and in language learning[9]. While learning outcomes may be similar across robot and human instruction, the nature of these learning interactions may be vastly different.

Another approach to understanding the degree to which robots can be successful tutors is examining how children engage with and perceive robots as instructors. Taking a user-centric approach to understanding child-robot interactions (CRI), whereby children's individual differences and preferences are considered, may better elucidate how children are engaging with this technology and may inform how a robot can adapt its behaviors to different learners. In a study that examined young children's engagement, vocalization, and emotion expression with a social robot and human instructor during drawing tasks, researchers found that children engaged more

[1]Temple University, Philadelphia, USA. [2]Franklin & Marshall College, Lancaster, USA. ✉email: allison.langer@temple.edu

and produced more words with a human instructor compared to the robot instructor[9]. In another study, children found it easier to communicate with a human tutee than a robot that they were trying to teach a task to[10]. In the current study, we capture children's vocalizations towards both a human and robot tutor to further understand differences between children's responses to the different instructor types.

While children's vocal utterances are a commonly used and informative metric of children's engagement with a social robot, children also make use of non-verbal signals in learning tasks with robots[11]. One quantifiable behavioral cue researchers can use to study children's engagement with a tutor during a learning task is eye gaze. Eye gaze has been a well-studied topic in human-robot interaction and is linked to capturing attention, maintaining engagement, and increasing conversational fluidity between users and robots[12,13]. Children's gaze has been shown to be especially important for social engagement and effective robot tutoring[14] and is a frequently used metric in child-robot educational research[15]. While eye gaze towards a robot instructor may enhance children's experience with the robot, it is also possible that excessively gazing towards a robot can distract children from the task at hand. In line with this prior work, we capture children's eye gaze during their interactions with both the human and robot instructors to explore the relationship between eye gaze, engagement, and efficiency in the learning task.

Examining social referencing (i.e., children looking to others for information, particularly in uncertain situations), can provide useful insights into how children navigate learning tasks[16]. Studying social referencing may be particularly important in studies of CRI, as it reveals how children seek information from others when interacting with different kinds of interaction partners. Especially for young children, patterns of social referencing may indicate their level of comfort and willingness to engage with novel technological agents in learning environments. One study showed that children display different patterns of social referencing with robots compared to humans[17]. Specifically, children interacting with social robots engaged in more frequent referencing behaviors towards their caregivers than children interacting with human partners, with this pattern being consistent across multiple interaction sessions rather than decreasing as children become more familiar with the robot.

User-related characteristics

In the field of CRI, little is known about how individual differences in user characteristics impact how children interact with social robots in educational settings[18]. One characteristic that may shape interactions with robots is children's age; several studies have demonstrated age-related changes in how children perceive and respond to social robots. For instance, 3-year-olds showed a trend towards trusting a human more than a robot after playing with either agent, and 7-year-olds displayed the opposite trend[19]. Children's age also affects how well they can attend to robots. During an interaction with a peer-tutor robot, children who had just turned 3 years of age looked less at the robot during their interaction than children who were almost 4 years old[20]. In a subsequent review of CRI studies with an age range of 8–10-year-old's (N=276, M_{age} = 9.50), researchers found that younger children, compared to older children, may enjoy interacting with robots more, may be less sensitive to a robot's interaction style, and be more prone to anthropomorphize robots[21]. Given the nuance of age-based results in CRI throughout various stages of development, we seek to uncover age-related trends among 5–8-year-olds in the present study.

Although there is limited work in this area, CRI researchers have begun to investigate how characteristics of children's personalities may be salient predictors of their interactional style with robots. For example, shy children approach a social robot more distantly when compared with less shy peers and interact less expressively with social robots in general[22]. During a Simon Says activity with a social robot, children (M_{age}=4.5 years) who had higher levels of surgency, or a measure of extraversion, sociability, and high activity levels as indexed by the parent-report Child Behavior Questionnaire[23], spoke more with the robot than children with lower levels of surgency[24]. In language learning research in adults, dispositional negative attitudes toward robots impede the ability to learn new vocabulary from a robot tutor[25]. Children's temperament has been strongly associated with academic achievement (for a meta-analysis, see[26]), which suggests it is an important user factor when considering design parameters for child-robot interactions in learning contexts.

In addition to examining how children's temperament is associated with their interactions with social robots, it is also important to consider how differences in children's social development—including empathy, cooperation, and self-control—may impact child-robot interactions. For robot tutors to be able to best adapt to children's needs as human tutors do, it is important that they interact with children in developmentally appropriate ways. This may include identifying and responding to specific social communication needs during learning tasks. While child-robot interaction studies have used social skills measurements as outcome variables after robot-assisted interventions (see[27] for review), to the best of our knowledge, no study has explored whether children's social skills at baseline affect their interactions with robot in meaningful ways. To this end, we seek to understand how social developmental stage may influence interaction characteristics with a robot compared to a human instructor.

Children's prior experience with other social technologies like voice assistants (e.g., Alexa or Siri) may also shape how they will interact with robots. Children growing up in households with interactive technologies may view these technologies as social agents who have feelings and agency[28] and may overestimate their intelligence[29]. In this study, we explore how the amount of technology a child has at home and the degree to which a child engages with that technology may contribute to how they interact with a novel robot in a learning context.

Children's perception of robots

In line with advancements in the CRI community to include children as key stakeholders in the creation of robotic technologies for education contexts[30], we sought to understand children's preferences for instructor types between a human and robot instructor in the same task in this study. Some previous studies have used

hypothetical scenarios to explore children's fears and hopes regarding robots[31] or their intentions to use a robot[32]. A strength of the current study is we gather information on children's preferences for a robot or human instructor after they have completed a real-life interaction with both instructor types.

The current study

The current study examines children's behavior while completing two tangram tasks, one with a robot instructor and one with a human instructor. During the tasks, we gathered various metrics including children's looking behavior (eye gaze) and their speech towards the robot and human instructors. We also examine child characteristics, specifically age, parent-reported technology use, temperament, and social skills, to determine if and how these person-centered factors differentially influence how children interact with the human instructor compared to the robot instructor. While there is a literature examining the effects of a physically-present robot in assisting users to complete a puzzle task[33–36], the current work is novel through its inclusion of a young participant user group, its use of a within-subjects design to the compare a robot and human instructor in the same task, and the use of validated measures of child temperament and social skills and to understand how user-related traits affect CRI. Our three main research questions can be summarized as follows:

1. How do children's behaviors differ when interacting with a robot and a human instructor during a tangram puzzle task?
2. Which individual differences influence variation in child behaviors towards a robot and human instructor in a tangram task?
3. After interacting with a human and robot instructor in the same task, what are children's perceptions of and preferences for interacting with the different instructors?

Methods

Participants

The study involved one visit to a developmental psychology research laboratory on a university campus in a small city in the eastern United States. The study was advertised through emails sent to mailing lists of families who had expressed interest in participating in research at the university and through social media posts on the lead researcher's webpage.

The study visit lasted about 1 hour. Thirty-two children aged 5–7 years (M=6.81 years, SD=0. 94; 17 girls & 15 boys) completed the study. An additional 7 children began the laboratory visit but were excluded from analyses due to failing to complete the puzzle task with the robot (n=1) or with both the research assistant and the robot (n=5), or due to issues with the recording equipment (n=1). The Institutional Review Board (IRB) at Franklin and Marshall College reviewed and approved the IRB protocol for this study. Temple University agreed to cede IRB review and approval to Franklin and Marshall College. All research was performed in accordance with the regulations of the IRB at Franklin and Marshall College and with the Declaration of Helsinki.

Participants' parents/legal guardians signed an online informed consent form prior to visiting the laboratory, and children gave verbal assent before participating in the laboratory task. Thirty-one parents identified their child as being White, and 1 parent identified their child as being "other race, ethnicity, or origin," among the listed racial and ethnic categories (Hispanic, Latino/a/x, or of Spanish origin; American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or Other Pacific Islander; Prefer to self-describe). The median household income for participating families was between $100,000 and $124,999.

Procedure

During the laboratory visit, children completed two puzzle tangram tasks (one with a human instructor, one with a robot instructor). The order in which participants interacted with the robot and human instructor was counterbalanced across participants to minimize carry-over effects. Participating children were given a toy worth approximately $10 upon completion of the study. Participants' parents or guardians also completed a series of online questionnaires prior to the laboratory visit, including a demographic questionnaire, a technology usage questionnaire, and assessments of their child's temperament and social skills. Parents/guardians were compensated with a $10 gift card upon questionnaire completion.

Tangram task

For the tangram task, a research assistant accompanied the child into the testing room while their parent/guardian remained in a waiting room. Upon entering the room, the child was asked to sit across a table from the instructor. Various tangram pieces and a piece of paper were on the table in front of the participant (see Fig. 1). The research assistant sat in a chair in the corner of the room and remained occupied with paperwork throughout, so as not to influence the participant-instructor interactions. Three video cameras had been set up in the testing room: One was positioned on the table facing the participant, one was mounted on the wall behind the instructor, pointing towards the participant, and one was on a tripod in the back of the room, facing the instructor.

With each instructor (robot and human), participants first completed an "introduction phase" where the human or robot instructor asked the child a question about their favorite color or animal and then instructed them to form a simple shape (a house or a square) from two tangram pieces. This introduction phase allowed the participants to familiarize themselves with both the tangram pieces and the interactive nature of the human or robot instructor.

Following the introduction phase, the task phase began. The tangram task was modeled on previous human-robot interaction work with adult participants[23]. During the task, participants were asked to assemble colored tangram blocks into the shape of a fox (in their first task block) or a rabbit (in their second task block) using
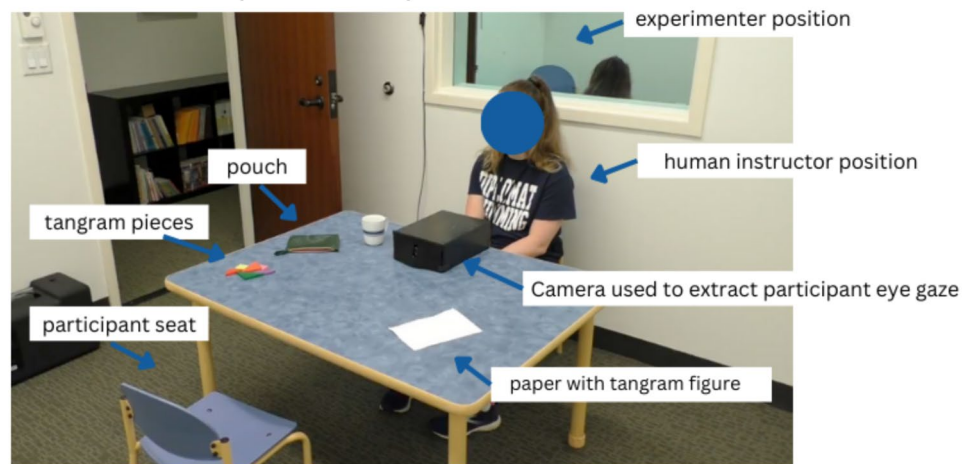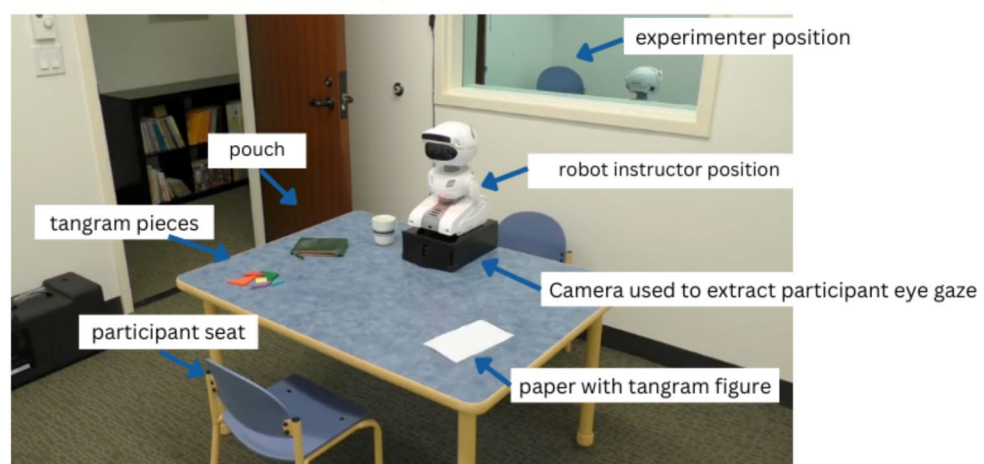
Fig. 1. (**A**) Experimental set-up with the human instructor and (**B**) with the robot instructor. Both conditions included tangram pieces, the sheet of paper with the illustration of the tangram puzzle shape, the pouch or mug containing the missing piece, and the participant's seat, the instructor's position at the table, and the camera used to extract participant eye gaze.

a picture as a guide. The picture was on the sheet of paper on the table in front of the child, with the picture originally facing down such that it was not initially visible to the child. The tangram task began with requests from the instructor: "Can you now flip the paper?" then "Now can you use all the pieces to build something like the picture?" The robot or human instructor provided feedback as the participants worked to complete the puzzle, either in the form of direct instructions (e.g., "rotate the blue piece"), or offering help (e.g., "do you need help?").

To elicit help-seeking behavior, one piece of the tangram puzzle was purposely hidden in a small pouch or cup on the testing table. The participant had to ask for help to locate this "missing piece," to which the instructor replied with directions on where to find the piece. The tangram task was completed when the child successfully completed the puzzle, at which point the instructor said, "Yay, you got it all! Hooray! I am so proud of you." For each child, the time to complete the task was obtained from the video recording, as was the duration of the introduction phase. The total interaction time was measured as the introduction phase plus the task completion phase.

To minimize differences between instructor conditions in the kinds of assistance offered and in affective expressivity, the human instructor was trained to use only the prompts and gestures that were programmed into the robot and to use a similar level of affect as the robot when interacting with the child. Both instructors were trained to only use prompts from a specific script (see supplementary materials for the full script) that provided both instructor types with sufficient responses for a variety of situations. To ensure consistency from the human across an interaction and between multiple interactions, the two research assistants who acted as the human instructors trained for over 30 hours by practicing cues and various responses in interactions with other lab members and with ten pilot participants.

The robot set-up

A Misty II robot (Misty Robotics) was used as the robot instructor for the tangram task due to its child-friendly appearance and interactive features. Through a "Wizard of Oz" setup, Misty was meant to appear as if it was acting autonomously, but the robot was being controlled by a human operator in a separate room. The operator used a touchscreen interface on a Chromebook to select from 70 predefined robot behaviors that provided motivation, offered guidance, engaged with the child, and acknowledged the child's achievements (see Fig. 2). All behaviors were composed of a combination of six fundamental robot actions: To speak, change facial expression, look in a direction, point in a direction, tilt its head, and pause for a short amount of time.

To ensure consistent robot behavior, the operator trained for over 30 hours by practicing cues and various responses in interactions with other lab members and with ten pilot participants. Additionally, the user interface used by the wizard underwent three months of iterative development to ensure that it included a sufficient breadth of behaviors to cover all expected scenarios, while also limiting the number of controls to allow the operator to quickly find and select the appropriate one. To further facilitate rapid responses by the operator, the layout was designed for two-handed operation, with the most frequently used controls on the right side of the touchscreen (the operator was right-handed). For further description of the Misty interface, see[37].

Eye gaze

To examine child looking behavior, we determined the proportion of time during the entire interaction (introduction phase and task phase) that each participant spent looking at the robot or human instructor. Eye gaze data was extracted from the video feed of the camera positioned on the table which provided a clear view of the child's face during the interaction (see Fig. 1). Coordinates corresponding to the participant's eye gaze were produced by processing the video with OpenFace (Baltrušaitis). For each frame of the video feed, the output of the video processing indicated whether the participant was looking at the agent (robot or human instructor) or at other areas in the room.

Social referencing and vocalizations

Three behavioral aspects of children's interactions with the instructor were manually coded: (1) social referencing, defined as a count of the number of times the participant looked to the research assistant in the room (who was sitting in the corner and perceived to be working on something else) during the task, (2) a count of the child's vocalizations towards the instructor, and (3) the count of the instructor's vocalizations towards to the child. Four coders viewed the videos of interactions and reliably coded a subset of the videos with 97% reliability. Social referencing and vocalizations are included in our analyses as a count per minute; specifically, "count" was divided by the total time of the interaction to control for participants' differing interaction times.

Parent-report social and behavioral measures

Participants' parents or guardians completed two questionnaires about their child: the Child Behavior Questionnaire (CBQ)[38] and the Social Skills Improvement Rating Scale (SSIS)[37]. The CBQ is a 97-item caregiver report measure designed to provide a detailed assessment of temperament in children aged between 3 and 7 years[38]. Temperament refers to consistent individual differences in children's styles of engagement with their surroundings, including how they respond to various stimuli and situations[38]. In the CBQ, parents are asked to describe their children's characteristic reactions to various situations based on descriptions that range from "extremely untrue" of their child to "extremely true" of their child. The CBQ indexes fifteen aspects of child
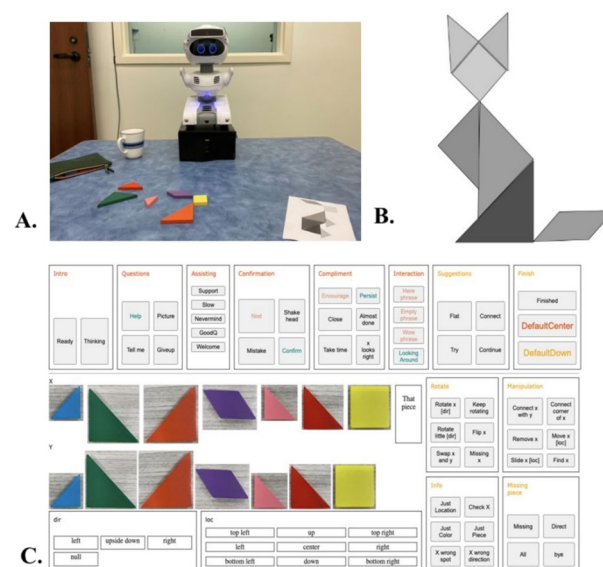

Fig. 2. (**A**). Experimental set-up of the Misty robot from the participant's viewpoint, (**B**) the fox tangram shape, and (**C**) the operator's touchscreen interface for triggering robot behaviors.

temperament, with subscales for Positive Anticipation, Smiling/Laughter, High Intensity Pleasure, Activity Level, Impulsivity, Shyness, Discomfort, Fear, Anger/Frustration, Sadness, Soothability, Inhibitory Control, Attentional Focusing, Low Intensity Pleasure, and Perceptual Sensitivity. The CBQ and SSIS are well-validated, commonly used, and statistically interpretable measures for assessing children's temperament and social skills, respectively. Further, prior work has demonstrated how similar developmental scales can be useful in examining psychological correlates of technology exposure and use[39,40] and are predictive of interactions with robots specifically[18,22].

The SSIS is a 73-item measure for caregivers of children aged 3–18 years and is designed to assess children's social skills and behavioral problems. Parents are asked to rate how often their child displays a particular behavior and how important the parent thinks the behavior is for their child's development[39]. The SSIS has two primary subscales, one that covers child behaviors in various social domains (including communication, cooperation, assertion, responsibility, empathy, engagement, self-control) and the other that taps into domains of problem behavior (including externalizing and internalizing behaviors, bullying, hyperactivity/inattention, and autism spectrum traits).

Participants' parents or guardians completed a survey of their child's technology usage at home. The technology use survey asked parents to indicate which technologies, including televisions, tablets/iPads, desktop or laptop computers, smartphones, video game devices, voice assistants (e.g., Alexa or Google Voice), robot toys, and other types of robots (like robotic vacuum cleaners), families had in their home. They were then asked how many of each type of technology families had in their home ("total technology"), as well as how frequently their child used each technology in the past week ("frequency of technology use").

Post-interaction interview

After completing the tangram task with the instructor, the child went with the experimenter to a second testing room where they were asked a series of post-interaction questions. A camera mounted on a tripod in the corner of the room videorecorded the child throughout. Moving to a second testing room allowed the child to answer questions about their interaction without the instructor present. It also allowed the original testing room to be reset for the second tangram task with the other instructor.

After each tangram task, participants were asked the following questions on a 5-point visual Likert scale from "Not at all" (1) to "Very much/A lot" (5): "How much did you trust [the instructor] to help you complete the puzzle? Would you like to play with [the instructor] again? How friendly do you think [the instructor] is? How smart do you think [the instructor] is?" Participants were told their responses would not be shared with either instructor.

After interacting with both instructors, children were asked which instructor they would prefer to play the tangram task with again if they had the chance (human/robot?), if they would rather play again with the robot instructor or a friend from school, and if they would rather play again with the human instructor or a friend from school. Children were also asked the following two questions specifically about the robot: "Do you think the robot has feelings?" (yes/no?) and "Do you think the robot is alive?" (yes/no?).

Results

Robot vs. human instruction

To address our first research question regarding robot vs. human instruction, we used paired t-tests to compare the means of various descriptive and behavioral metrics across the human and robot instructor conditions (see Table 1.). These metrics included introduction time (measured in minutes), task time (measured in minutes), gaze towards the instructor (proportion of time spent gazing at the instructor during the task, relative to the tangram pieces or elsewhere in the testing room), social referencing (count per minute), child vocalizations (count per minute), and instructor vocalizations (count per minute).

We first used Levene's Test to test for homogeneity of variance and found equal variances for all variables of interest except for social referencing, which had unequal variances between conditions $F_{(1, 30)} = 26.89$, $p < .001$. For the full results from Levene's Test, see the Supplementary Materials. We used Welch's Two Sample T-Test to compare means for social referencing across conditions and Student's T-Tests to compare condition means across all other variables of interest.

When completing the task with the robot instructor, children spent more time in the introduction phase ($t = -11.40$, $p < .001$), spent a greater proportion of time looking at the instructor ($t = -3.74$, $p < .001$), and referenced the research assistant in the room more often ($t = -3.79$, $p < .001$), compared to when they were completing the task with the human instructor. Task completion times did not differ between instructor types ($p > .05$). The number

	Human mean	Human SD	Robot mean	Robot SD	t	p	95% CI lower	95% CI upper
Introduction time (minutes)	0.97	0.35	1.90	0.30	−11.4	<.001	−1.09	−0.77
Task time (minutes)	3.15	2.37	3.47	1.91	−0.59	0.55	−1.39	0.77
Gaze at instructor (proportion)	0.34	0.15	0.48	0.15	−3.74	<.001	−0.21	−0.07
Social referencing (count/minute)	0.06	0.25	0.81	1.09	−3.79	<.001	−1.15	−0.35
Child vocalizations (count/minute)	3.12	3.48	4.59	5.37	−1.29	0.20	−3.73	0.79
Instructor vocalizations (count/minute)	16.34	11.93	15.00	8.47	0.52	0.61	−3.84	6.52

Table 1. Means, standard deviations, t-tests results for the human instructor and robot instructor conditions.

	Main effect of instructor		Main effect of order		Interaction: instructor: Order	
	F (30,1)	p	F (30,1)	p	F (30,1)	p
Introduction time (minutes)	137.7	<.001	0.01	0.94	3.59	0.07
Task completion time (minutes)	0.74	0.39	1.51	0.23	15.83	<.001
Gaze at instructor (proportion)	24.93	<.001	0.03	0.85	3.34	0.08
Social referencing (count/minute)	14.64	<.001	0.9278	0.343	2.54	0.121
Child vocalizations (count/minute)	4.57	0.04	7.63	0.01	7.70	0.009
Instructor vocalizations (count/minute)	0.54	0.47	1.12	0.30	11.40	0.002

Table 2. Univariate type III repeated-measures ANOVA for instructor order effects.

	Instructor order	Contrast	Estimate	p	Interpretation
Introduction	Human–robot	Robot - Human	0.78	0.00	Robot significantly higher
Introduction	Robot–Human	Robot - Human	1.08	0.00	Robot significantly higher
Task	Human–robot	Robot - Human	−1.16	0.04	Robot significantly lower
Task	Robot–Human	Robot - Human	1.80	0.00	Robot significantly higher
Gaze	Human–robot	Robot - Human	0.19	0.00	Robot significantly higher
Gaze	Robot–Human	Robot - Human	0.09	0.03	Robot significantly higher
Social referencing	Robot–Human	Robot - Human	1.06	0.00	Robot significantly higher
Child vocalizations	Robot–Human	Robot - Human	3.38	0.00	Robot significantly higher
Instructor vocalizations	Human–Robot	Robot - Human	−7.50	0.01	Robot significantly lower

Table 3. Summary of significant pairwise contrasts between instructors by instructor order.

of child vocalizations was greater towards the robot compared to the human instructor, but this difference did not reach statistical significance (p>.05). The human instructor had more vocalizations per minute than the robot instructor, but this was not a statistically significant difference (p>.05). As shown in Table 1., the variability around the various means (as shown by the standard deviations) supports further investigation into individual difference measures.

Order effects

Table 2. displays order effects which were tested using 2 x 2 mixed ANOVAs with Bonferroni corrections to examine the effects of Instructor (human vs. robot) and Order (Human-Robot or Robot-Human) and on all dependent behavioral outcomes: introduction time, task completion time, gaze at instructor, social referencing, child vocalizations, and instructor vocalizations. Pairwise contrasts were computed for each behavioral outcome across instructor type and instructor order (e.g. "emmeans(introduction, pairwise ~ Instructor | Order, adjust = "bonferroni"). Bonferroni adjusted p-values for significant pairwise contrasts are reported in Table 3.. Boxplots depicting the main and interaction effects and tables with marginal means for all contrasts are displayed in the Supplementary Materials.

There was a main effect of Instructor (F (1,30) = 137.72, p <.001) on the introduction phase length—participants spent significantly longer with the robot instructor than the human instructor during the introduction phase. There was no significant effect of order on introduction time (F (1,30) = 0.01, p = 0.94), suggesting that the order in which the participants interacted with either instructor had no impact on introduction time. There was no significant interaction effect of instructor and order on introduction time.

There was no significant main effect of instructor type on task time, (F (1, 30) = 0.74, p =.40), indicating that, overall, task time did not differ significantly between the human and robot instructors. The results indicated that there was no significant main effect of order on task time, (F (1, 30) = 1.51, p =.23), suggesting that the sequence in which participants experienced the instructors did not significantly impact task time. However, there was a significant instructor x order interaction (F (1, 30) = 15.83, p <.001), suggesting that the effect of instructor type (human vs. robot) on task time depended on the order in which the participants completed the task with each instructor. Children in the human-robot order completed the task with the human in 4.26 minutes and the robot in 3.10 minutes on average, whereas children in the robot-human order completed the task with the robot in 3.83 minutes and the human in 2.04 seconds on average. In other words, children in the Human-Robot order completed the task only 1.16 minutes faster on average during their second interaction (with the robot in this case), whereas children in the Robot-Human condition completed the task 1.79 minutes faster on average during their second interaction (the human, in this case), and this difference is significant.

There was a main effect of instructor on gaze (F (1,30) = 24.93, p <.001), whereby children spent longer proportions of time gazing at the instructor in the robot condition compared to the human condition. There was no significant effect of order on gaze, (F (1,30) = 0.034, p = 0.85), meaning the order in which the instructor

presented did not influence gaze behavior. There was no interaction of instructor x order (F (1,30) = 3.34, *p* = 0.08).

There was a main effect of instructor on social referencing (F(1,30) = 14.64, *p* <.001), showing that children engaged in social referencing more frequently when interacting with the robot compared to the human instructor. There was no significant effect of order on social referencing (F(1,30) = 0.93, *p* = 0.34), meaning the sequence in which the instructor presented did not impact social referencing behavior. Additionally, there was no significant interaction between instructor x order (F(1,30) = 2.54, *p* = 0.12), suggesting that the effect of the instructor on social referencing did not differ based on the order of the interactions with the instructors.

There was a significant main effect of instructor (F(1,30) =4.56, *p*=.04) and order (F(1,30) =7.63, *p*=.01) on child vocalizations where in the robot-human order, children spoke significantly more with the robot than the human instructor. There was no significant difference in child vocalizations in the human-robot order. Additionally, there was a significant order × instructor interaction (F(1,30) = 7.69, *p* =.01). Children spoke more with the robot when they interacted with the robot first, compared to children who encountered the robot in their second interaction.

There was no significant main effect of order on instructor vocalizations (F(1,30)=1.12, *p*=.30). Additionally, there was no significant main effect of instructor (F(1,30) =0.54, *p*=.47), suggesting that the type of instructor (human vs. robot) did not significantly influence instructor vocalizations. However, there was a significant instructor × order interaction (F(1,30) =11.40, *p*=.002), demonstrating that the robot instructor spoke significantly less when children interacted with the human first and robot second.

User-related characteristics

Means and standard deviations for all user-related predictors are listed in Table 2.. Cronbach's alpha for the three factors of the Child Behavior Questionnaire were 0.82 for Surgency, 0.80 for Negative Affect, and 0.81 for Effortful Control. The types of technology families endorsed having in their home and the amount their child used each technology is shown in Fig. 3.

To investigate whether individual differences were associated with variation in child behavior towards the robot and human instructor, we ran Pearson correlations to test the relations between the child's age, technology use, temperament, and social skills measures with the children's introduction time, eye gaze, and social referencing. For task time and child vocalizations, a partial correlation was used to control for the effect of Instructor Order (see Table 4.).

Age was significantly negatively associated with children's task time with both the human and the robot instructors after applying the Benjamin-Hochberg correction. Age negatively related, but not significantly, to social referencing and child vocalizations with the human but not the robot instructor. We found that perceptual sensitivity was significantly negatively related to social referencing with the robot instructor. After applying the Benjamin-Hochberg correction to all user-related predictors, no other correlations were significant.

A Pearson's product-moment correlation was conducted to examine the relationship between age and belief that the robot is alive. The results indicated a negative correlation that approached significance, *r*(34) = −.32, *p* =.06. The 95% confidence interval for the correlation ranged from −.59 to.01. A Pearson product-moment correlation was conducted to assess the relationship between age and belief that the robot has feelings. The correlation was positive but not statistically significant, *r*(34) =.25, *p* =.14. The 95% confidence interval ranged from −.09 to.53, indicating considerable uncertainty around the estimate.

Children's perception of the instructor

Paired t-tests showed that participants rated the robot significantly higher than the human instructor when asked about friendliness (t=−2.11, df=30, *p*=0.04) and about how much the child would like to play with the instructor again (t=−2.26, df=30, *p*=0.03). There were no significant differences between how children rated the trustworthiness of the robot and the human instructor (t=−1.49, df=30, *p*=0.15) or in ratings of how smart either instructor was (t=−0.83, df=30, *p*=0.41).

After completing the tangram task with both the robot and the human instructor, 93% of participants answered that they would prefer to play the game again with the robot, with the remaining 7% indicating a preference for the human instructor (Fig. 4). A Chi-Square Goodness of Fit Test with a null hypothesis assuming an equal preference for the human and robot instructor showed that responses significantly deviated from an even distribution, X2=22.53 (df=1, *p*<.001).

Discussion

This study utilized a within-subjects design to elucidate patterns and predictors of children's behavior during a tangram task with a robot or human instructor, to better understand in which contexts, and for whom, robot tutors may be a good fit. Specifically, we used children's eye gaze, social referencing, and vocalizations to compare the interactions across instructors. We also investigated how child-level characteristics, including the child's age, parental report of child behavioral and social tendencies (on the Child Behavior Questionnaire; CBQ and the Social Skills Improvement Rating Scale; SSIS) and a parental report measure of their child's technology usage at home, correlated with interaction elements. We assessed participants' technology usage to test the hypothesis that prior technology exposure would affect how children perceived the robot, as it has in previous studies with children and adults[41]. To highlight that children are key stakeholders in work that studies their interactions with robots, we also examined children's preferences for interacting with either instructor after they completed the study.

Overall, we found that children's age and temperament influenced only some aspects of their interactions. Children rated the robot instructor highly on friendliness and indicated that they would like to play a tangram task again with the robot. We did not find a relation between children's prior technology use with interaction
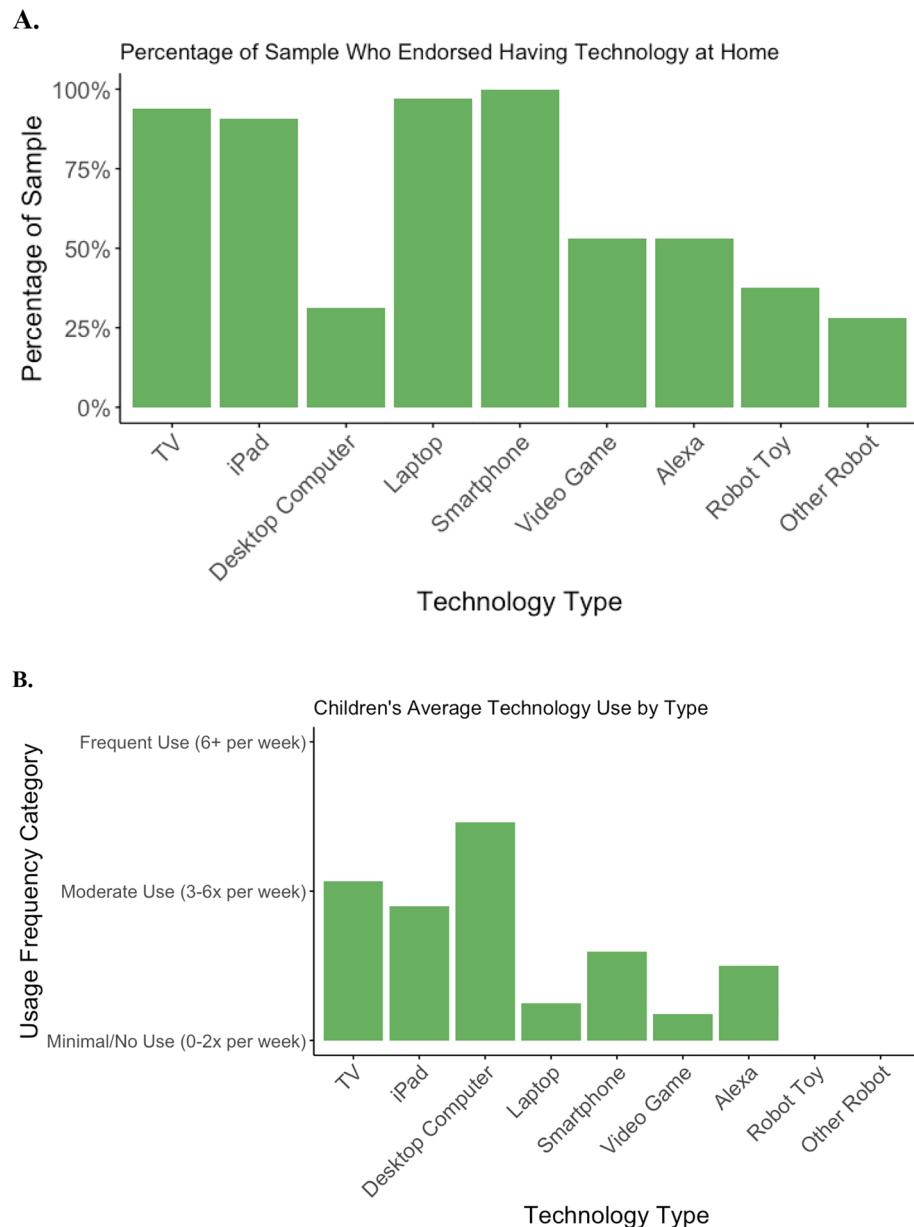
A.

B.

Fig. 3. Results from the Technology Use questionnaire showing (**A**) the types of technology parents reporting having in their home, and (**B**) the frequency at which their child uses each type of technology per week.

elements. We now review our main findings and suggest broader implications of these results for CRI in real-world contexts.

Robot vs. human instruction

Compared to the human instructor, children spent more time looking at the robot and referenced the research assistant in the room more often when interacting with the robot. One factor driving this finding may be the novelty of the robot. The children's increased tendency to socially reference the research assistant in the room while interacting with the robot suggests they may have been trying to use information from the research assistant's facial expression to inform their own perception of the robot in this less familiar learning context. Our finding replicates that of[17] who demonstrated that preschool-aged children (ages 4–5, n=20) utilized social referencing behaviors with their caregivers more often while interacting with a robot compared to a human. Tolksdorf and colleagues' results, as well as our findings, suggest that it is important to consider where and when a child may need additional support when interacting with novel digital technologies[17].

Although there was no difference in task completion times with the robot or human instructor, children spent more time in the introduction phase of the experiment with the robot compared to the human instructor. This may be due, in part, to slight lags in the operator's response during the introduction phase, which included more individualized responses than the task phase. Alternatively, children may have needed or wanted more

	M	SD	Human Instructor					Robot Instructor				
			Intro Time	Task Time	Gaze	Social Referencing	Child Vocalization	Intro Time	Task Time	Gaze	Social Referencing	Child Vocalization
Age	6.80	0.94	-0.21	-0.44*	0.25	-0.35°	-0.49°	-0.05	-0.54*	0.10	-0.07	-0.26
Child Behavior Questionnaire Subscales												
Surgency	4.58	0.92	0.03	-0.08	-0.10	-0.22	-0.05	0.08	0.18	-0.09	0.10	0.17
Negative affect	4.06	0.69	-0.12	-0.24	-0.06	0.00	-0.10	0.07	-0.28	-0.06	0.08	0.02
Effortful Control	5.34	0.63	-0.09	-0.09	0.19	0.00	0.02	0.08	0.12	0.03	-0.21	-0.15
Child Behavior Questionnaire Dimensions												
Activity Level	4.14	0.61	-0.30	-0.49°	0.07	-0.28	-0.06	-0.17	-0.19	-0.05	-0.14	-0.10
Anger/Frustration	4.37	0.83	-0.06	-0.38°	-0.04	-0.06	0.02	0.08	-0.30	-0.09	0.13	0.01
Approach	4.87	0.69	-0.25	-0.21	0.21	0.11	-0.08	-0.38°	-0.28	-0.29	-0.41°	-0.27
Attention	4.33	0.71	-0.39°	-0.26	-0.07	-0.12	-0.01	0.12	-0.13	-0.16	0.21	0.04
Discomfort	3.75	0.68	-0.08	-0.19	0.13	-0.16	-0.04	-0.05	-0.34	-0.06	-0.07	-0.03
Soothability	4.40	0.60	-0.23	-0.13	0.16	0.04	-0.02	0.00	0.09	-0.22	-0.02	0.07
Fear	3.63	0.85	0.18	-0.05	-0.18	0.06	0.04	0.04	-0.08	-0.44°	-0.19	-0.12
High Intensity Pleasure	4.46	0.92	-0.07	-0.07	-0.03	-0.08	-0.37°	-0.15	-0.03	-0.08	0.11	-0.10
Impulsivity	3.64	0.70	-0.16	-0.06	0.14	0.01	-0.17	-0.32	-0.29	0.02	-0.11	-0.14
Inhibitory Control	4.77	0.80	-0.28	-0.13	0.36°	0.02	-0.16	0.04	0.04	-0.04	-0.20	-0.19
Low Intensity Pleasure	5.79	0.67	0.07	-0.08	-0.01	0.06	0.23	0.12	0.12	-0.01	0.34	0.28
Perceptual Sensitivity	4.89	0.71	0.07	-0.14	0.17	-0.08	-0.14	-0.23	-0.33	-0.05	-0.53*	-0.38°
Sadness	3.98	0.78	-0.28	-0.50	-0.02	-0.11	-0.09	-0.08	-0.25	-0.18	0.04	-0.09
Shyness	4.09	0.58	-0.20	-0.02	0.25	0.22	-0.36°	-0.07	-0.01	-0.03	0.12	-0.21
Smiling and Laughter	3.98	0.45	-0.10	-0.12	-0.08	-0.09	0.26	-0.22	-0.35°	-0.02	-0.19	-0.12
Social Skills Improvement System-Rating Scales												
Social Skills	97.19	15.93	-0.08	0.21	-0.08	0.15	0.12	-0.01	0.21	0.03	-0.27	-0.03
Problem Behaviors	22.88	8.70	-0.12	0.04	-0.28	-0.04	-0.08	0.10	-0.11	-0.23	0.06	-0.02
Technology Use												
Total technology at home	11.38	7.00	-0.03	-0.23	0.01	-0.13	0.23	0.12	-0.05	0.10	-0.05	-0.05
Frequency of use	2.84	2.59	-0.09	0.06	0.13	-0.14	0.15	0.25	0.11	0.13	0.03	0.33

Table 4. Pearson partial and full correlations of child characteristics and interaction measures. Note: M=mean, SD=standard deviation. *p<.05 after Benjamin-Hochberg correction; °p<.05 before correction was applied. For "Task Duration" and "Child Vocalizations," partial correlation was performed to control for order effects of when the child interacted with either instructor type (Human-Robot or Robot-Human).
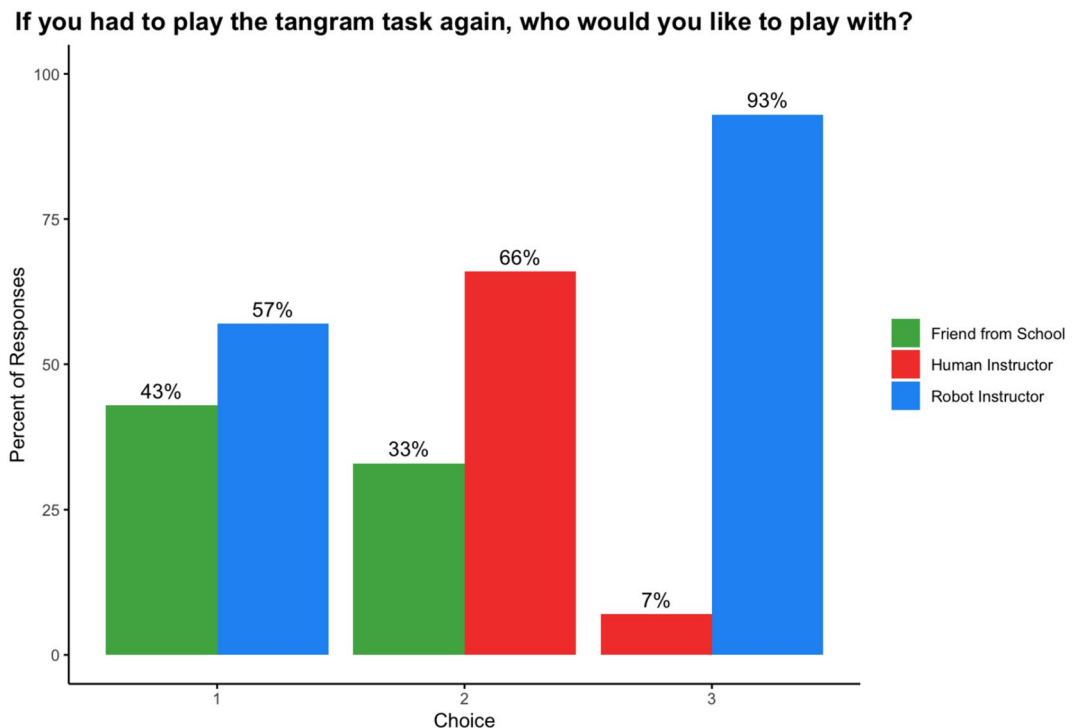
Fig. 4. Participants' preferences for playing the tangram task again with the robot instructor or a friend, the human instructor or a friend, and either the human or robot instructor.

time to get acquainted with the robot, again due to its novelty. Prior work has shown that novelty of a robot can be a distraction to children's learning[17], so it will be important to test repeated interactions with our paradigm to see if children still complete the task with the robot and human at similar rates after the novelty effect is likely to wear off.

Order effects

Though we tried to control for order effects by alternating the order in which participants interacted with the human and robot, we did find effects for task time, child vocalizations, and instructor vocalizations. While children in both orders got faster at completing the tangram task in their second interaction, children who interacted with the robot first and human second had a more significant decrease in task time than children who interacted with the human first and robot second. This could suggest that children learned more effectively from the robot instructor than the human instructor. This learning effect was present even though children gazed at the robot instructor for longer proportions of time than the human instructor more overall. Despite studies showing that the novelty effect of robots in learning tasks—as indexed by children's heightened gaze—may impede learning[17], we did not find this to be the case in our sample.

There was also an order effect on the amount children spoke to either instructor. Children who interacted with the robot first and human second spoke significantly more to the robot than children in the human-robot order. Interestingly, this is the same group of children who completed the task more quickly the second time around (with the human instructor), lending more evidence to the interpretation that children engaged with and learned from the robot effectively in this order. However, some of the effects of the differences in child vocalizations may be driven by the fact that children in the human-robot order spoke less in *both* conditions relative to the robot-human order, across both interactions (though not significantly so). It may then be a random effect of our relatively small sample that quieter children were randomly grouped together, which we then controlled for in our correlations between child vocalizations and individual predictors. Future studies may wish to establish a baseline level of vocalization among the participants to consider this when assigning instructor order to further understand differences in vocal engagement with human and robot instructors, and how this engagement may contribute to children's learning from either instructor type.

We also found an interaction involving instructor vocalizations, whereby the robot spoke significantly less in the human-robot order than in the robot-human order. This may have compounded with the other effects we observed for task time and child vocalizations—it is possible that if the robot spoke less, children may have been less engaged and thus spoke less, and less engagement may have led to longer task completion times on the second trial than expected. However, it's important to note that there are only sixteen participants in each instructor order group (robot-human; n=16, human-robot; n=16), which may not be sufficiently powered to fully understand the interaction effects we present here[42]. We hope larger samples may be better equipped to fully understand the mechanisms of learning with both instructor types, and why the order of interactions with robot instructors may matter when introducing robots to classroom settings.

User-related characteristics

Given the large body of literature on how development and age shape children's interactions with and perceptions of technologies, we examined the associations between age and interaction features across instructor types. We found that age was negatively correlated with task duration time across instructor types, suggesting older children completed the tangram faster. For the application of robots in educational settings, particularly with children across elementary school ages, it will be important to understand children's interest and motivation to sustain attention when interacting with social robots.

Though not a significant trend, older children were more likely than younger children to see the robot as alive, while younger children more often saw the robot as having feelings, compared to older children. This is in line with previous research suggesting children tend to anthropomorphize robots, even after brief interactions with them[3]. Children also may have been more inclined to anthropomorphize the robot given the Wizard-of-Oz setup used in the experiment; in studies where children are told that the robot is being controlled by a human, they are less likely to anthropomorphize the robot[43,44]. The non-significant finding that fewer older children thought the robot had feelings than younger children is congruent with developmental trends of younger children being more likely to attribute affective and mental states to inanimate objects, including robots[28,44,45]. Our results related to children's age and their perceptions should be interpreted with caution as they are not significant, however it is possible with a larger sample size or additional trials our results may continue to reflect what has been shown in prior literature relating age and perceptions of social robots.

Child temperament, as indexed by the three broad dimensions and fifteen subscales of the CBQ, was largely unassociated with introduction and task phase durations, eye gaze, social referencing, and vocalizations towards the robot and human instructors, with one exception. We found that perceptual sensitivity—which measures how children detect low-intensity stimuli from the external environment—was negatively correlated with the amount the child engaged in social referencing while interacting with the robot instructor. This may suggest that children who are more attuned to subtle facial expressions and emotional cues from the robot did not need to look towards the experimenter in the room for guidance as often. Alternatively, children who are more sensitive to small visual details may have been focusing heavily on the puzzle in front of them and were not as inclined to turn that focus elsewhere to the experimenter in the room.

We did not replicate the findings of[22], who showed that surgency was positively related to children's verbal engagement with a robot, or those of[18], who found that shy children interacted less expressively with a robot. We saw correlations that were nearly significant but did not withstand corrections for multiple comparisons. As such, these relationships warrant further investigation in future work. We saw that fear was negatively related (but not significantly so) to looking at the robot instructor, while shyness was related (but not significantly so), to speaking to the human instructor but not the robot instructor. As these were not significant correlations, they should not be over-interpreted. However, the differing patterns of relationships between children's temperament with humans compared to robots supports methodologies that directly compare humans and robots in the same task. When designing robot tutors, it may not be sufficient to rely on children's interaction characteristics that are known form human-human interactions. Instead, children's behavior may be specific to context—a shy child may only be shy with a robot or a human; their behavior in one scenario may not predict the other.

While we did not find any associations between social skills and engagement with the human and robot instructors, children's social developmental stage may still be a relevant user-characteristic to consider when designing and developing robotic tutors for young children. Similarly, we did not see any relationship between technology use at home and interaction characteristics. While some studies suggest prior exposure to robots may increase positive perceptions of future robot encounters[46], we found that our sample had very low exposure to robots or interactive voice agents (Fig. 3B) before participating. As robots become more ubiquitous in museums, libraries, or schools, or if children begin to increase their interactions with interactive agents like voice assistants at home, it is possible the level of exposure to such technologies may influence how a child learns from a robot tutor.

Despite the lack of associations between children's characteristics and their interaction behaviors with human and robot tutors shown in our study, we posit that the study of child characteristics in CRI is still crucial to building effective and ethical devices for children. Other CRI studies have shown that robots that personalize their tutoring according to children's academic knowledge, attention, working memory[47], or facial expressions[48] positively impact children's learning gains. As artificial intelligence and social robot technologies become more integrated, other characteristics that impact learning–such as age, temperament, social development, prior technology use—may be able to inform personalization architectures that contribute to better learning outcomes for children.

Children's perception of the instructors

Based on the pattern of responses to the post-interaction interviews, children strongly preferred the robot compared to the human instructor. However, since the human instructor was trained to only exhibit gestures and behaviors that the robot could portray, these preferences should be considered in the context of this experimental set-up. The human may have seemed unfriendly relative to how a young child may expect a human instructor to act. Indeed, children ranked the human instructor lower than the robot when asked about friendliness. Further work with repeated exposures to a robot and/or enhancement of instructor affective responses may be helpful for understanding these effects. While children's perceptions of the instructors may not be directly related to how well they can learn from either instructor, it is important that children's perspectives are actively sought and incorporated into the design process of social robots for educational contexts[49].

Limitations and future directions

Generalizability may be limited by the study sample, which was children with a homogenous racial and socioeconomic background from a small college town in the eastern United States. Further, children interacted with both instructor types for only a few minutes, and future research would benefit from studying longer or repeated interactions, particularly to gauge how any novelty effects of the robot may wane over time. Larger samples sizes would also allow for a closer look at the order effects we observed when children completed the tangram task with the robot instructor first compared to the human instructor first. Additionally, the study relied on parent reports of child temperament and social skills. Future studies may wish to include teacher report and behavioral assessments of temperament in the laboratory that could be used to supplement the parent report measures.

Broader limitations concern the Wizard-of-Oz setup and the "matching" of the human instructor's affect to the Misty robot. Given that the human experimenter was trained to resemble Misty in their interactions, children's behaviors elicited during the task may be different than with an adult in a real classroom who does not have this constraint.

Conclusion

With the development and application of robot tutors likely to become more prevalent, there is a need to better understand patterns of children's unique responses to robots in learning settings. The current study adds to the literature on individual differences in CRI by showing how child temperamental style may influence how children interact with robots compared to humans. The results from this study can guide the design and implementation of social robots for use with young children and can encourage further study of how child characteristics influence children's behaviors and experiences during and after interactions with a robot instructor.

Our results on individual predictors of CRI differ from prior work showing that children's levels of extroversion and shyness predict how they behave with a social robot. Instead, we found that a different predictor—how sensitive children are to environmental stimuli—was related to how often they referenced an experimenter in the room while interacting with the robot, but not the human, instructor. Given our focus on individual predictors of CRI, we felt it imperative to include a human condition to serve as a baseline relative to the robot condition. Our results suggest there may be differences in how children respond uniquely to robots compared to human instructors based on the child's behavioral traits.

We also found that children learned effectively from both instructor types, and gazing towards the robot was not a distraction to learning as previously studies have suggested. We also find potentially important order effects, where children who interacted with the robot first seemed to have been more engaged and were able to improve their tangram task time significantly more than the children who interacted with the human instructor first. However, as we explain in the discussion, these results should be interpreted with caution based on our sample size.

The implications of our findings, and the main contribution of this study, center on the potential for user-based traits to inform the development of adaptive agents, and metrics that can inform best practices for introducing robots to young users. Future applications may be able to use information derived from measures such as the CBQ and the SSIS to better adapt to a child user's needs. Alternatively, such psychological measures could be used to validate unsupervised learning models that generate interaction styles for a child user based on robot-generated temperament profiles.

While child behavior profiles can be used to guide adaptive interactions, this also raises ethical concerns related to the long-term implications of children's exposure to robots[46,50]. Studying children's individual differences regarding how they respond to robots draws attention to considerations around the adverse effects of CRI, which could shape best practices and regulations regarding how robots are introduced to children and at what ages.

Data availability

Please contact the corresponding author, Allison Langer, at Allison.langer@temple.edu for access to de-identified data files.

References

1. Belpaeme, T. et al. Guidelines for designing social robots as second language tutors. *Int. J. Soc. Robot.* **10**, 325–341 (2018).
2. Johal, W. Research trends in social robots for learning. *Curr. Robot. Rep.* **1**(3), 75–83 (2020).
3. Johal, W., Castellano, G., Tanaka, F. & Okita, S. Robots for learning. *Int. J. Soc. Robot.* **10**, 293–294 (2018).
4. Van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., Van der Ven, S. & Leseman, P. Social robots for language learning: A review. *Rev. Educ. Res.* **89**(2), 259–295 (2019).
5. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. & Tanaka, F. Social robots for education: A review. *Sci. Robot.* **3**(21), 5954 (2018).
6. Conti, Daniela, Cirasa, C., Di Nuovo, S. & Di Nuovo, A. Robot tell me a tale! A social robot as tool for teachers in kindergarten. *Interact. Stud.* **21**(2), 220–242 (2020).
7. Fong, Frankie TK., Sommer, Kristyn, Redshaw, Jonathan, Kang, Jemima & Nielsen, Mark. The man and the machine: Do children learn from and transmit tool-use knowledge acquired from a robot in ways that are comparable to a human model?. *J. Exp. Child Psychol.* **208**, 105148 (2021).
8. Papadopoulos, I. et al. A systematic review of the literature regarding socially assistive robots in pre-tertiary education. *Comput. Educ.* **155**, 103924 (2020).

9. Zhexenova, Z. et al. A comparison of social robot to tablet and teacher in a new script learning context. *Front. Robot. AI.* **7**, 99 (2020).

10. Serholt, S., Ekström, S., Küster, D., Ljungblad, S. & Pareto, L. Comparing a robot tutee to a human tutee in a learning-by-teaching scenario with children. *Front. Robot. AI.* **9**, 836462 (2022).

11. Tolksdorf, N.F., Mertens, U. "Beyond words: Children's multimodal responses during word learning with a social robot". In *International Perspectives on Digital Media and Early Literacy.* Routledge 90-102 (2020).

12. Admoni, H. & Scassellati, B. Social eye gaze in human-robot interaction: A review. *J. Human-Robot Interact.* **6**(1), 25–63 (2017).

13. Serholt, Sofia, Barendregt, W. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proc. 9th Nordic Conference on Human-Computer Interaction.* (2016).

14. Mwangi, Eunice, N. Gaze-based interaction for effective tutoring with social robots. (2020).

15. Sim, Gavin & Bond, R. Eye tracking in child computer interaction: Challenges and opportunities. *Int. J. Child-Comput. Interact.* **30**, 100345 (2021).

16. Walle, Eric A., Reschke, P. J. & Knothe, J. M. Social referencing: Defining and delineating a basic process of emotion. *Emot. Rev.* **9**(3), 245–252 (2017).

17. Tolksdorf, N.F., Crawshaw, C.E., Rohlfing, K.J. Comparing the effects of a different social partner (social robot vs. human) on children's social referencing in interaction. *Front. Educ.* Front. Med. SA 5 (2021).

18. Tolksdorf, Nils F., Viertel, F. E. & Rohlfing, K. J. Do shy preschoolers interact differently when learning language with a social robot? An analysis of interactional behavior and word learning. *Front. Robot. AI* **8**, 676123 (2021).

19. Di Dio, C. et al. Shall i trust you? From child–robot interaction to trusting relationships. *Front. Psychol.* **11**, 469 (2020).

20. Baxter, P., De Jong, C., Aarts, R., de Haas, M., Vogt, P. The effect of age on engagement in preschoolers' child-robot interactions. In *Proc. Companion of the 2017 ACM/IEEE Int. Conf. Human-Robot Interact.* (2017).

21. Van Straten, C. L., Peter, J. & Kühne, R. Child–robot relationship formation: A narrative review of empirical research. *Int. J. Soc. Robot.* **12**(2), 325–344 (2020).

22. Neumann, M. M., Koch, L. C., Zagami, J., Reilly, D. & Neumann, D. L. Preschool children's engagement with a social robot compared to a human instructor. *Early Child. Res. Quart.* **65**, 332–341 (2023).

23. Wilson J. R, Aung P.T, Boucher I. "When to help? A multimodal architecture for recognizing when a user needs help from a social robot". In *International Conference on Social Robotics. Cham* (*Springer Nature*. Switzerland, 2022)

24. Barry, Ruby-Jane., Neumann, M. M. & Neumann, D. L. Individual differences and young children's engagement with a social robot. *Comput. Hum. Behav.: Artif. Hum.* **4**, 100139 (2025).

25. Kanero, J. et al. Are tutor robots for everyone? The influence of attitudes, anxiety, and personality on robot-led language learning. *Int. J. Soc. Robot.* **14**(2), 297–312 (2022).

26. Nasvytienė, Dalia & Lazdauskas, T. Temperament and academic achievement in children: A meta-analysis. *Eur. J. Investig. Health Psychol. Educ.* **11**(3), 736–757 (2021).

27. Raptopoulou, Anastasia, Komnidis, A., Bamidis, P. D. & Astaras, A. Human–robot interaction for social skill development in children with ASD: A literature review. *Healthc. Technol. Lett.* **8**(4), 90–96 (2021).

28. Flanagan, T., Wong, G. & Kushnir, T. The minds of machines: Children's beliefs about the experiences, thoughts, and morals of familiar interactive technologies. *Dev. Psychol.* **59**(6), 1017 (2023).

29. Andries, V. & Robertson, J. Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. *Comput. Educ.: Artif. Intell.* **5**, 100176 (2023).

30. Mott, T., Bejarano, A., Williams, T. Robot co-design can help us engage child stakeholders in ethical reflection. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* (IEEE, 2022).

31. Rubegni, E., Malinverni, L., Yip, J. Don't let the robots walk our dogs, but it's ok for them to do our homework: Children's perceptions, fears, and hopes in social robots. In *Proc. 21st Annual ACM Interaction Design and Children Conference* (2022).

32. de Jong, C., Peter, J., Kühne, R. & Barco, A. Children's intention to adopt social robots: A model of its distal and proximal predictors. *Int. J. Soc. Robot.* **14**(4), 875–891 (2022).

33. Leyzberg, D., Spaulding, S., Toneva, M., Scassellati, B. The physical presence of a robot tutor increases cognitive learning gains. In *Proc. Annual Meeting of the Cognitive Science Society* 34 34 (2012).

34. Kirschner, D., Velik, R., Yahyanejad, S., Brandstötter, M., Hofbaur, M. YuMi, come and play with Me! A collaborative robot for piecing together a tangram puzzle." Interactive Collaborative Robotics. In *1st International Conference, ICR 2016, Budapest, Hungary, Proceedings 1. Springer International Publishing* 24-26 (2016).

35. Resing, W. C., Vogelaar, B. & Elliott, J. G. Children's solving of 'Tower of Hanoi'tasks: Dynamic testing with the help of a robot. *Educ. Psychol.* **40**(9), 1136–1163 (2020).

36. Zinina, A., Zaidelman, L., Kotov, A., Arinkin, N. The perception of robot's emotional gestures and speech by children solving a spatial puzzle. *Computational Linguistics and Intellectual Technologies.* In *Proc. International Conference.* Dialogue **19** 26 (2020).

37. Yang, Yuqi, Langer, A., Lauren Howard, L., Peter J. Marshall P.J, Wilson J.R. Towards an ontology for generating behaviors for socially assistive robots helping young children. In *Proc. AAAI Symposium Series* **2**(1) 213-218 (2023).

38. Rothbart, M. K., Ahadi, S. A., Hershey, K. L. & Fisher, P. Investigations of temperament at three to seven years: The children's behavior questionnaire. *Child. Dev.* **72**(5), 1394–1408 (2001).

39. Gresham, F., Elliott, S. N. Social skills improvement system (SSIS) rating scales. Bloomington, MN: *Pearson Assessments* (2008).

40. Gülay Ogelman, H., Güngör, H., Körükçü, Ö. & Erten Sarkaya, H. Examination of the relationship between technology use of 5–6 year-old children and their social skills and social status. *Early Child Dev. Care* **188**(2), 168–182 (2018).

41. O'Reilly, Z., Roselli, C., Wykowska, A. Does exposure to technological knowledge modulate the adoption of the intentional stance towards humanoid robots in children? (2023).

42. Kennedy, James, Baxter, Paul & Belpaeme, Tony. The impact of robot tutor nonverbal social behavior on child learning. *Front. ICT.* **4**, 6 (2017).

43. Nicolas, Sommet, Weissman, D. L., Cheutin, N. & Elliot, A. J. How many participants do I need to test an interaction? Conducting an appropriate power analysis and achieving sufficient power to detect an interaction. *Adv. Method. Pract. Psychol. Sci.* **6**(3), 25152459231178730 (2023).

44. van Straten, C. L., Peter, J. & Kühne, R. Transparent robots: How children perceive and relate to a social robot that acknowledges its lack of human psychological capacities and machine status. *Int. J. Hum.-Comput. Stud.* **177**, 103063 (2023).

45. Chernyak, N., & Gary, H. E. Children's cognitive and behavioral reactions to an autonomous versus controlled social robot dog. In *Young Children's Developing Understanding of the Biological World* 73-90 (Routledge, 2019).

46. Kahn, P. H. Jr. et al. Robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot. *Dev. Psychol.* **48**(2), 303 (2012).

47. Chen, Y. C., Yeh, S. L., Lin, W., Yueh, H. P. & Fu, L. C. The effects of social presence and familiarity on children-robot interactions. *Sens. (Basel).* **23**(9), 4231 (2023).

48. Almousa, Ohoud & Alghowinem, Sharifa. Conceptualization and development of an autonomous and personalized early literacy content and robot tutor behavior for preschool children. *User Model. User-Adapt. Interact.* **33**(2), 261–291 (2023).

49. Maaz, Nafisa, Mounsef, J. & Maalouf, N. CARE: Towards customized assistive robot-based education. *Front. Robot. AI* **12**, 1474741 (2025).

50. Langer, A., Marshall, P. J. & Levy-Tzedek, S. Ethical considerations in child-robot interactions. *Neurosci. Biobehav. Rev.* **151**, 105230 (2023).

Author contributions

Allison Langer wrote the manuscript and prepared all the figures. All authors contributed substantially to reviewing and editing the manuscript and were involved in the study design. Lauren Howard led data collection and behavioral coding efforts, and Jason Wilson instructed all robot-related programming, extracted gaze data, and supplied the team with the Misty Robot. Allison performed data analyses in R. Peter Marshall provided compensation for the families in the study from research funds.

Declarations

Conflict of interest

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-27476-x.

Correspondence and requests for materials should be addressed to A.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.