



# Evaluating the Role of a Socially Assistive Robot in Children’s Mental Health Care

Elizabeth R. Kitt<sup>1</sup> · Molly K. Crossman<sup>1</sup> · Angela Matijczak<sup>1</sup> · Gillian B. Burns<sup>1</sup> · Alan E. Kazdin<sup>1</sup>

Accepted: 30 April 2021 / Published online: 19 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Socially assistive robots (SARs) present a promising intervention for addressing the increasing prevalence of childhood stress. This innovative technology has become increasingly common in practical implementation. However, empirical support has not kept pace with the robots’ growing popularity. The present study set out to provide an empirical test of the stress-buffering capability of this technology. We examined the effects of the presence of an SAR on self-reported measures of stress in a community sample of 70 children (40 girls, 30 boys) between the ages of 7 and 10 years old. Child participants were randomly assigned to either a robot-present or a robot-absent condition during the Trier Social Stress Test for Children. Contrary to our predictions, we did not detect a stress-buffering effect of the presence of the robot. Instead, the primary analysis showed that the robot’s presence led to a larger decrease in positive affect following the task. However, further analyses found no significant difference in positive affect when accounting for baseline group differences. Exploratory analyses of the nature of children’s interactions with the SAR during the stressful task found that higher levels of parent-reported social anxiety were associated with greater prosocial behaviors towards the SAR, thus signifying a potential target for future interventions. Further work on SARs is needed to determine the optimal timing and robotic specifications that would maximize the potential of this developing technology to improve children’s mental health care.

**Keywords** Socially assistive robot · Child · Stress · Intervention · Mental health care

## Highlights

- We examined the ability of a socially assistive robot (SAR) to buffer against stress experienced by children.
- We did not detect expected benefits of the SAR for alleviating children’s stress.
- We propose future research to examine the optimal use of SARs for children’s stress.

It is commonly accepted that technological advances have shifted family dynamics over the past few decades (Bavelier et al., 2010; Padilla-Walker et al., 2019). Researchers, politicians, and commentators alike have warned against the potential deleterious impacts of this increasing reliance on technology. The pervasive integration of technology into society has brought with it the risk of parents distracted by

their phones, children indoctrinated by the television, and family members growing distant from one another (Longman et al., 2019).

While the risks of technology have captured public attention for years, a growing literature has begun to explore the potential benefits of technology for children’s mental health. With regards to everyday use of screen-based technologies, such as mobile devices and televisions, recent findings have led to the development of the Goldilocks Hypothesis, which posits that there is an optimal level of technology use for children (Przybylski & Weinstein, 2017). By this theory, moderate use of technology (roughly 1 to 2 h per day of screen time) provides advantages for children’s mental health by connecting youth with their peers and providing opportunities for developmentally appropriate learning (Przybylski et al., 2021; UNICEF, 2017).

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s10826-021-01977-5>.

✉ Elizabeth R. Kitt  
elizabeth.kitt@yale.edu

<sup>1</sup> Yale University, New Haven, CT, USA

Beyond these potential advantages of everyday use, additional benefits may arise from targeted, therapeutic applications of technology. A wide range of traditional technologies, such as videoconferencing and mobile phone applications, have already been successfully integrated into children's mental health care to improve the scalability of services aimed at prevention, diagnosis, and treatment of mental illness (for a review, see Boydell et al., 2014). In addition to incorporating traditional technologies into the delivery of care, some researchers have developed new technologies altogether with the intent of improving children's mental health care. One particularly promising solution involves the rapidly growing field of socially assistive robotics (Kazdin, 2019; Rabbitt et al., 2015). As the name implies, socially assistive robots (SARs) are designed to provide aid or comfort to humans through social interactions (Feil-Seifer & Mataric, 2005). At the most basic level, these robots identify human social cues and respond with behaviors that are appropriate to the situation (Mataric & Scassellati, 2016; Fong et al., 2003). They are also able to portray simulated emotions to enhance social interactions. Some examples of SARs include Paro, a companion robot based on a baby harp seal; Nao, a child-sized anthropomorphic robot; and Pekee, a vehicular-shaped robot (Cabibihan et al., 2013).

Designed to fulfill a similar function to companion animals, SARs derive much of their theoretical grounding from research on the benefits of therapy animals for addressing childhood stress. Several studies have found that children's interactions with both pet dogs and unfamiliar dogs can buffer against the onset of self-reported and physiological metrics of stress in response to a stressful task (Beetz et al., 2011; Kerns et al., 2018; Kertes et al., 2017). In each of these studies, the stress-buffering effects of the presence of the dog were strongest when the children showed greater prosocial behaviors towards the dog, thus suggesting that these effects are driven by both the quality and quantity of the interactions.

Addressing children's stress is of the utmost importance considering its high prevalence and far-reaching consequences (American Psychological Association, 2017; Costello et al., 2003). Chronic exposure to stress—even normative stress—at a young age can have severe psychiatric and physical complications in both the short- and long-term when not properly addressed (Lau, 2002). Excessive childhood stress is considered a risk factor for a range of psychiatric illnesses, including alcoholism, post-traumatic stress disorder, anxiety, depression, and suicidality (Fryers & Brugha, 2013; Middlebrooks & Audage, 2008). High levels of early life stress are also associated with behavioral concerns, including poor academic performance and increased risk-taking behaviors (Shonkoff et al., 2012). Left untreated, high levels of childhood stress are

associated with a host of somatic illnesses, including chronic obstructive pulmonary disease, ischemic heart disease, liver disease, and metabolic disorders (Danese et al., 2009; Middlebrooks & Audage, 2008).

Although companion animals have been found to help children manage their stress, the reach of this particular model of stress-buffering can be limited by several key factors. For example, dogs introduce concerns relating to contamination, allergies, and children's fears (Crossman, 2016). Due to these and other concerns, live companion animals cannot always be easily introduced into some environments, such as schools and intensive care units in hospitals. SARs have thus been designed to provide the support and comfort associated with therapy animals while circumventing many of these barriers to care. Socially assistive robots could be used in locations where therapy animals are not permitted, as well as with a wide range of individuals, regardless of their allergies or fears of animals (Broadbent et al., 2018; Logan et al., 2019). These features, in addition to their transportable design, allow them to be easily disseminated and used in a vast array of settings (Rabbitt et al., 2015).

Past research has found that children and their parents typically view social robots as an acceptable intervention for children (Dawe et al., 2019; de Jong et al., 2019; Rabbitt et al., 2015). Moreover, the perception of SARs as non-judgmental makes them particularly well-suited as a tool in children's mental health care (Cabibihan et al., 2013). This nonjudgmental characterization could especially benefit children with increased sensitivity to negative evaluation, such as those with higher levels of social anxiety (Rapee & Heimberg, 1997). Such individuals might receive increased comfort from interacting with a nonjudgmental robot, as this would diminish the evaluative aspect of the interaction (Clark & Wells, 1995).

Socially assistive robots have already been introduced in several areas of mental health care, most commonly with geriatric patients with dementia or children with autism spectrum disorder (Rabbitt et al., 2015). A series of pilot studies and several key randomized controlled trials have indicated modest but promising benefits of SARs for improving patient outcomes in both populations (Marino et al., 2019; Moyle et al., 2017; Pennisi et al., 2016; Robinson et al., 2013). In addition, a small yet growing body of research supports the stress-buffering capacity of SARs in high-stress populations, such as hospitalized children (Logan et al., 2019).

Several preliminary studies have begun to explore this stress-buffering role in community samples. In line with research on therapy animals, most of this research has focused on the effects of SARs on children's mood and anxiety symptoms. Qualitative research in school settings found that teenagers were highly engaged in interactions

with SARs and retrospectively expressed that their interactions with the robot reduced their stress levels (Björling et al., 2019). Following these results, the same research team is currently running an interdisciplinary effort to design an SAR with the express purpose of reducing teens' stress (Rose et al., 2019). In younger children, one qualitative study found that the majority of children in several rural schools reported experiencing numerous positive emotions, such as "calm" and "happy," after interacting with an SAR (Broadbent et al., 2018). Furthermore, our research group recently evaluated the ability of an SAR to address mood and anxiety symptoms in a community sample of young children (Crossman et al., 2018). The study found that brief, unstructured interactions with an SAR after a stressful task increased children's positive affect more than interacting with the robot turned off or than waiting without an interaction partner for an equivalent period of time. However, negative affect, self-reported state anxiety, and salivary cortisol levels were not altered by interacting with a robot.

While these preliminary findings are encouraging, the empirical support for the use of SARs in children's mental health care has not kept pace with the rate of production and implementation of these robots (Dawe et al., 2019). Furthermore, some researchers have warned that these robots may be more harmful than beneficial if children use SARs as a crutch or attribute sentience to the robots (Pearson & Borenstein, 2014; Tolksdorf et al., 2021; Westlund et al., 2015). Additional research is necessary to better understand the nature of the effects of SARs and negate the potential for harm.

The purpose of the current study was to investigate whether the presence of an SAR during a stressful situation could buffer against stress experienced by children. SARs have been found to alleviate children's stress-related symptoms *after* a stressful task. Previous work on the stress-buffering capacities of dogs suggest that the robots might be effective *during* such a task, as well. We predicted that children would show diminished negative responses to a stressful task if the Paro robot was present during the procedure. Specifically, we hypothesized that, as compared to participants who completed the task alone, participants who completed the task in the presence of the robot would show (1) smaller declines in positive affect, (2) smaller increases in negative affect, and (3) smaller increases in negative emotional response following the stressful task. Particularly for children with higher levels of social anxiety, this non-human, nonjudgmental model may be ideally suited to buffer against stress; thus, we included baseline social anxiety as a possible adjustment variable. Finally, we explored how the degree of interaction between the participant and the robot related to the other measures examined in the study.

## Method

### Participants

Seventy-eight child participants between the ages of 7 and 10 were recruited from the local community. Of the 78 participants who began the procedure, 8 participants were excluded from the analyses. Participants were excluded from the final analyses if they omitted a significant proportion of the procedure by choice or if the experimenters substantially altered the procedure in order to prevent excessive stress. Participants were informed that they may stop the procedure at any time or skip a component of the procedure. Furthermore, experimenters demonstrated an abundance of caution at all stages of the protocol and modified and/or stopped the experiment in cases of any concern for the participant's wellbeing. In this study, it was only necessary to alter the protocol during aspects of the stress induction procedure (see description below). For the purposes of the analyses, participants were excluded if they completed less than 75% of the stress induction procedure. We included participants who had completed the majority of the stress-induction procedure but did not finish the full task in order to retain the maximum number of participants possible and to avoid excluding the children who may have experienced the greatest degree of normative stress during the protocol. Specifically, four participants requested not to participate in a component of the stressful task and thus were excluded from analyses. In three cases, experimenters decided to abbreviate the stressful task in response to observable signs of a participant's rising stress level. The final excluded participant needed to leave in the middle of the stressful task due to a family emergency.

We recruited our sample from the local community to address the pressing need for interventions that combat increasing rates of chronic childhood stress. In addition, because the present study is one of the first to examine the benefits of SARs for children's mental health, we hoped to investigate the impact of SARs on children's stress-related symptoms in general before narrowing in on a specific disorder. Thus, we did not impose any preliminary exclusionary criteria, and we did not select for any clinical criteria. Although we did not screen for task-independent levels of stress or anxiety in selecting our sample, these baseline features are linked with reactivity to stress induction procedures, such as the one included in our protocol (Krämer et al., 2012). We therefore accounted for these characteristics by collecting baseline levels of all measures of interest (see Table 1 for descriptive characteristics).

We chose the age range of 7 to 10 years old since transdiagnostic cognitive vulnerabilities have been shown to emerge in this age range and high stress at this age can have long-term impacts on a person's physical and mental health

**Table 1** Descriptive characteristics by study condition

	Experimental condition ( <i>n</i> = 36)	Non-robotic control ( <i>n</i> = 34)
Female [ <i>N</i> (%)]	20 (55.56)	20 (58.82)
Male [ <i>N</i> (%)]	16 (44.44)	14 (41.18)
Age [years, <i>M</i> ( <i>SD</i> )]	8.66 (1.19)	8.86 (1.28)
Race [ <i>N</i> (%)]		
Asian	6 (16.67)	2 (5.88)
Black/African American	0 (0.00)	1 (2.94)
Hispanic/Latino	1 (2.78)	2 (5.88)
White, Non-Hispanic	25 (69.44)	25 (73.53)
Other	1 (2.78)	2 (5.88)
Multiple Races	2 (5.56)	2 (5.88)
Unknown	1 (2.78)	0 (0.00)
Ethnicity [ <i>N</i> (%)]		
Hispanic or Latino	3 (8.33)	4 (11.76)
Not Hispanic or Latino	30 (83.33)	30 (88.24)
Unknown	3 (8.33)	0 (0.00)
PANAS-C-P positive affect [ <i>M</i> ( <i>SD</i> )]	19.64 (2.53)	20.15 (2.60)
PANAS-C-P negative affect [ <i>M</i> ( <i>SD</i> )]	8.22 (2.63)	8.33 (2.71)
SCAS-P total [ <i>M</i> ( <i>SD</i> )]	16.8 (8.16)	16.09 (7.53)
SCAS-P Social Phobia subscale [ <i>M</i> ( <i>SD</i> )]	4.17 (2.24)	4.33 (2.64)

PANAS-C-P (Positive and Negative Affect Schedule for Children, Parent Version) scores range from 5 to 25 for each subscale. SCAS-P (Spence Children's Anxiety Scale, Parent Version) total scores range from 0 to 114. Scores on the SCAS-P Social Phobia subscale range from 0 to 18. For all scales, higher scores reflect greater endorsement of the construct

(Hayden et al., 2013; Hong et al., 2017; Middlebrooks & Audage, 2008). Moreover, reliable and well-validated measures are available for assessing emotional responses in this age group. Thus, this age range presents an important target for interventions, and these new interventions can be examined in a systematic manner. Our sample was drawn from the region surrounding a university in the Northeastern United States. This county has a total population of approximately 850,000, and its residents are primarily White, Non-Hispanic (U.S. Census Bureau, 2019). The median household income in the region is approximately \$67,000 as of 2018, with a poverty rate of 11.6%.

Participants were recruited through a variety of methods. These included in-person recruiting at community events, online postings about the project, flyers posted around the community, and referrals from previous participants. After participating, child participants received a certificate of participation, a small prize, and a \$10 gift card to Amazon. This study was reviewed and approved by the Institutional Review Board of Yale University.

## Measures

### Positive and negative affect

We predicted that having the SAR present during a stressful task would buffer against reductions in positive affect and increases in negative affect in response to the stressor. To measure positive and negative affect, we used the Positive and Negative Affect Schedule for Children, Short Form (PANAS-C-S; Ebesutani et al., 2012). The PANAS-C-S is a 10-item self-report measure that asks children to rate their experience of 10 different feeling states “right now” on a five-point Likert scale ranging from “very slightly or not at all” to “extremely.” These 10 feeling states are equally divided between positive and negative affect into two separate subscales of five items each. Cronbach's alpha was 0.82 for the positive affect scale and 0.73 for the negative affect scale at baseline in the present study, demonstrating that these constructs were reliably assessed. Negative affect scores on the PANAS-C-S differentiate children with any anxiety and/or mood disorders from those without any anxiety or mood disorder. Furthermore, positive affect scores differentiate children with mood disorders from those with no mood disorder and from those with externalizing disorders (Ebesutani et al., 2012).

In addition to these self-report ratings of current positive and negative affect, we included the Positive and Negative Affect Schedule for Children-Parent Version (PANAS-C-P) in order to assess parent-reported positive and negative affect in the weeks leading up to the experiment as potential adjustment variables (Ebesutani et al., 2012). The 10 feeling states included in the PANAS-C-P and the associated Likert scale mirror those found in the PANAS-C-S, but parents were asked to rate the child's experience of each feeling state “during the past few weeks.” Cronbach's alpha was 0.80 for parent-reported positive affect and 0.71 for parent-reported negative affect in the present study, indicating acceptable internal consistency. Past research has also established the construct validity of the PANAS-C-P (Ebesutani et al., 2012).

### Emotional response

We hypothesized that the presence of the SAR during a stressful situation would buffer against increases in self-reported negative emotional responses. To assess children's emotional responses to the stressor, participants completed the Self-Assessment Manikin (SAM) at three time points (Bradley & Lang, 1994). The SAM is a nonverbal (i.e., graphic) self-assessment measure. Children reported their emotional response by selecting the image that best described their current feelings on a five-point pictorial Likert scale depicting non-gendered figures. The SAM

measures three principal dimensions of emotion: Pleasure/Valence, Arousal, and Dominance/Control (for a review, see Bynion & Feldner, 2017). This dimensional approach is commonly used as a model for assessing emotional states (Lonsdorf et al., 2017). The construct validity of the SAM has been well established across populations of various ages and cultures; it captures the expected arousal and valence of individuals' responses to images from the International Affective Picture System and correlates with verbal measures of affective response (Bynion & Feldner, 2017). Importantly, the SAM assesses emotional responses in a non-verbal manner. This aspect of the SAM is particularly useful in working with children since this group has a wide range of verbal abilities (Bynion & Feldner, 2017). Drawing on this particular benefit, the SAM has previously been used in the context of measuring children's emotional responses to stress induction tasks (Beetz et al., 2011; Gunnar et al., 2009; Kertes et al., 2017). The three components of the SAM (Pleasure, Arousal, and Dominance) were considered as separate variables in the present study.

### Trait anxiety

We included participants' trait anxiety levels as a possible covariate. To obtain a comprehensive view of participants' baseline anxiety levels, we used the Spence Children's Anxiety Scale, Parent Version (SCAS-P; Nauta et al., 2004). The SCAS-P is a parent-report measure that assesses anxiety levels in children and adolescents ages 6 to 18. It includes six subscales: Panic Attack and Agoraphobia, Separation Anxiety, Physical Injury Fears, Social Phobia, Obsessive-Compulsive Disorder, and Generalized Anxiety Disorder. These subscales are combined to provide an overall anxiety score. Cronbach's alpha was 0.83 for parent-reported trait anxiety, indicating this construct was measured reliably. The SCAS-P has demonstrated acceptable internal consistency and satisfactory test-retest reliability for children within our age range (DeSousa et al., 2014; Nauta et al., 2004; Wang et al., 2016). Scores on the SCAS-P are highly correlated with scores on the Revised Children's Manifest Anxiety Scale and are less highly, but still significantly, correlated with scores on the Children's Depression Index, thus demonstrating its convergent and divergent validity (Nauta et al., 2004).

Social anxiety was of particular interest due to the specific implications of the perception of SARs as nonjudgmental for children who typically struggle with interpersonal social interactions. The Social Phobia subscale of the SCAS-P was thus considered in isolation. Cronbach's alpha for parent-reported social anxiety was 0.72 in the present study. We also included the complementary child self-report version of this subscale, the six-item Social Phobia subscale of the Spence Children's Anxiety Scale,

Child Version (SCAS-C; Spence, 1998). The Social Phobia subscale of the SCAS-C has demonstrated high internal consistency and satisfactory 12-week test-retest reliability (Spence et al., 2003). Each individual subscale of the SCAS-C discriminates between a clinical and community sample, thus demonstrating its discriminant validity (Arendt et al., 2014). However, Cronbach's alpha was 0.57 for self-reported trait social anxiety in the present study, thus indicating that self-reported social anxiety was not assessed reliably in this study. We examined the intra-item correlations to determine whether the internal consistency of the scale could be improved; however, there were no individual items with low correlations to the rest of the scale that could be removed to improve the internal consistency (the correlations of individual items on the scale ranged from  $-0.08$  to  $0.49$ ). Consequently, self-reported social anxiety was excluded from further analysis.

### Nature of the Interaction with the SAR

We hypothesized that the effects of the robot's presence on the above signs of stress would be more pronounced in participants who demonstrated increased social behaviors towards the SAR during the stressful task. Children's level of interaction with the SAR was quantified with an observational coding scheme that was developed for this study based on previous studies of the benefits of therapy animals during stressful situations (Beetz et al., 2011; Kertes et al., 2017). A research assistant coded the duration of time that each participant in the experimental condition spent touching or gazing at the SAR from video recordings of the testing component of the stressful task. All behavioral coding in this project was completed using the coding program Behavioral Observation Research Interactive Software (BORIS; Friard & Gamba, 2016). Twenty-five percent of recordings (9 cases) were viewed by more than one coder. Inter-rater reliability was assessed using a two-way mixed, consistency, single measures intra-class correlation coefficient (ICC) for each variable of interest (both touch of and gaze at the SAR during each component of the stressful task) for a total of six variables (Hallgren, 2012). The reliability for all observational variables of interest was excellent, with an average ICC of 0.97 and range of 0.89 to 1.00 across these coded variables (Cicchetti, 1994).

### Stress Induction

In order to test our hypothesis that the SAR would minimize symptoms of stress experienced during a stressful situation, we used the Trier Social Stress Test for Children (TSST-C) to induce moderate psychosocial stress (Buske-Kirschbaum et al., 1997). The TSST-C draws on the three principal factors of psychosocial stress: social evaluation, uncontrollability,

and unpredictability (Dickerson & Kemeny, 2004). This task produces reliable increases in psychological and physiological indicators of stress in children ages 7 to 10 (Gunnar et al., 2009).

In this task, child participants are read the beginning of a story and then asked to develop an ending to that story as part of a competition against other children. Although the task is not truly competitive, the TSST-C is structured to enhance the stressful circumstances. Child participants are also told that they will be presenting in front of judges and that their story will be videotaped. They are then given three minutes alone to plan their ending to the story, with the experimenter reminding them of the remaining time after each minute. This constitutes the anticipatory phase of the task. After planning their story in anticipation of the competition, participants are then brought to a new room for the testing component of the task. A video camera and two judges wearing lab coats are waiting in the testing room when the participant arrives. Participants are asked to tell their ending to the story for five minutes, and then they are asked to complete a difficult mental arithmetic problem adapted for participant age. In order to prevent excessive stress, the judges smile and provide encouraging feedback throughout the task. The judges were trained to monitor child participants for signs of excessive stress (e.g., failure to respond for an extended period or signs of tearfulness). Whenever a participant showed signs of excessive stress or expressed a desire to stop, the procedure or component of the procedure was terminated immediately. We implemented these safeguards to ensure that no participant experienced an excessive level of stress. Additionally, the TSST-C is a well-established paradigm that has repeatedly been shown to typically induce only moderate levels of stress (Gunnar et al., 2009).

## Conditions

### Experimental condition

Participants in the experimental condition were accompanied by the robot during all aspects of the TSST-C. The SAR was presented to the participant immediately after the completion of baseline questionnaires. Participants were told that the SAR would be their “buddy” throughout their time at the lab, and they were informed of its robotic features, including its ability to respond to tactile and auditory stimulation and to recognize its name. Participants were then given three minutes to interact with the SAR in an unstructured manner. The SAR was present during the introduction to the TSST-C and the anticipation phase, and it was moved to the judges’ room for the testing component of the TSST-C. The SAR was handed back to participants immediately after they entered the judges’ room. Participants were permitted to place it on a small nearby

table if desired due to the robot’s weight; however, they were told to play with it as much as they wanted throughout the task. In the present study, the particular SAR used was the Paro robot. Based on a baby harp seal, the Paro robot is a soft companion robot weighing 2.7 kg (Intelligent Systems Co., Ltd., Nanto, Japan; [www.parorobots.com](http://www.parorobots.com)). The Paro robot responds to tactile and auditory stimulation with movements and automatized seal noises, and it adjusts its behavior based on its interactions in accordance with three simulated, easily recognizable moods (Japan’s National Institute of Advanced Industrial Science and Technology [AIST], 2004).

### No-robot control

In order to test our hypotheses about the potential stress-buffering capacities of the Paro robot, we included a no-robot control group to establish how children would respond to the procedure in the absence of the robot. Since this was the first test of the stress-buffering abilities of an SAR for this age range, we elected to include only a no-robot control group. This design allowed us to investigate the question of whether the SAR had any effect in this role, although the lack of a more targeted control group precluded us from examining whether this effect was specific to the robotic features of the SAR.

Instead of interacting with the robot, children in the no-robot condition waited for an equivalent amount of time before the stress induction task, as familiarization with the lab could influence the child’s stress. The protocol scripts for the two conditions were identical, beside the omission of any mention of the robot for the control condition. Separate from the data collection procedure, participants in this condition were given five minutes to interact with the robot after all outcome measures had been collected in order to ensure that all participants had a chance to interact with the SAR.

## Procedures

Upon a participant’s arrival, the experimenter randomly assigned the participant to either the experimental or no-robot control condition using an online random number generator (Research Randomizer; [www.randomizer.org](http://www.randomizer.org)). Randomization was divided based on the participant’s reported sex in order to ensure that each condition contained equivalent numbers of male and female participants. Parental consent and participant assent were obtained for all participants, and child participants were informed that they could end the procedure at any point if they wished. The parent of each child participant was given a background questionnaire to complete while the child began the procedure. This questionnaire included basic demographic information about the participant, the PANAS-C-P, and the SCAS-P.

**Table 2** Pre-, mid- and posttest scores for self-report outcome measures by study condition

	Experimental condition ( <i>n</i> = 36)			Non-robotic control ( <i>n</i> = 34)		
	Pretest <i>M</i> ( <i>SD</i> )	Mid-test <i>M</i> ( <i>SD</i> )	Posttest <i>M</i> ( <i>SD</i> )	Pretest <i>M</i> ( <i>SD</i> )	Mid-test <i>M</i> ( <i>SD</i> )	Posttest <i>M</i> ( <i>SD</i> )
PANAS-C-S positive affect	19.44 (4.78)	–	16.17 (6.23)	16.36 (5.65)	–	16.21 (5.88)
PANAS-C-S negative affect	7.03 (3.22)	–	7.58 (3.34)	6.71 (2.25)	–	7.14 (3.08)
SAM Pleasure	4.31 (0.82)	4.47 (0.81)	4.08 (1.11)	4.41 (0.70)	4.15 (0.99)	3.91 (1.08)
SAM Arousal	3.31 (1.47)	3.19 (1.37)	2.56 (1.34)	3.03 (1.47)	3.15 (1.42)	2.94 (1.54)
SAM Dominance	3.36 (1.22)	3.22 (1.35)	3.17 (1.42)	3.38 (1.13)	3.21 (1.31)	3.32 (1.30)

PANAS-C-S (Positive and Negative Affect Schedule for Children, Short Form) scores range from 5 to 25 for each subscale. SAM (Self-Assessment Manikin) scores range from 1 to 5 for each subscale. For all scales, higher scores reflect greater endorsement of the construct. Pretest scores were obtained before participants began the protocol, mid-test scores were obtained after the anticipatory period immediately prior to the testing component of the stressful task, and posttest scores were obtained immediately after completion of the stressful task

A total of three experimenters ran the child participants through the procedure, and a total of eight served as judges during the TSST-C, with one experimenter and two judges present for each participant. All experimenters were White, female, and students in either undergraduate or graduate programs. The ages of experimenters ranged from 19 to 24.

Participants began by completing the Social Phobia subscale of the SCAS-C before completing the baseline administration of the SAM and the PANAS-C-S. The participant was then either introduced to the SAR (experimental condition) or told that the rest of the procedure would continue shortly (control condition), and the experimenter left the room. After three minutes, the experimenter returned and introduced the story component of the TSST-C. The experimenter then left the room for another three minutes while the participant planned their story with the SAR either present or absent. The experimenter returned every minute to inform the participant of the remaining time. At the end of this anticipation period, the participant completed the mid-test administration of the SAM. The efficient nature of the SAM allowed for mid-task assessment without disrupting the timeline of the TSST-C. For the experimental condition, the SAR was removed while the participant completed the SAM, and the participant was informed that the SAR would be waiting in the judges' room. The participant was then led to a different room, where they completed the TSST-C.

After the TSST-C, the participant completed the posttest SAM and PANAS-C-S. Participants in the control condition were then given an opportunity to interact with the SAR after completing these final outcome measures. At the conclusion of the procedure, each participant was debriefed and told they did a wonderful job. Each participant received a small toy, a certificate of completion, and a gift certificate.

## Data analysis

We computed total scores for all multi-item self-report and parent-report measures. The mean of the completed items was

used to prorate missing data when individual items were missing from a particular scale. When a given measure was missing one quarter of the total items or greater, the measure was excluded for that participant. In order to test our prediction that participants completing the stressful task in the presence of the robot would show diminished negative responses to the task, we elected to use several Condition x Time mixed analysis of variance (ANOVAs) to investigate the changes in self-reported outcome measure throughout the procedure (Table 2). We examined changes in positive and negative affect across two time points (pretest to posttest), and we investigated changes in the three domains of emotional responsivity across three time points (pretest, posttest, and a mid-test score after the anticipatory phase of the TSST-C). We repeated the analysis with parent-reported positive affect as an adjustment variable to account for the fact that parent-reported positive affect correlated with change in affect across the experiment. Additionally, we performed a one-way analysis of covariance (ANCOVA) to account for baseline differences in self-reported positive affect. Furthermore, we conducted an unplanned Condition x Time mixed ANOVA assessing changes in positive affect using only participants who completed the full stress induction procedure in order to account for possible differences relating to whether the participant completed the full stress induction procedure. Finally, we were interested in exploring how participants' prosocial behaviors towards the robot might relate to changes in outcome measures or to trait anxiety levels. To address this aim, we conducted a series of supplementary correlations within participants in the experimental condition.

## Results

### Preliminary Analyses

Seventy child participants from the local community between the ages of 7 and 10 ( $M = 8.76$ ,  $SD = 1.23$ ) completed the procedure (Table 1). Forty (57.1%) were female and 30

(42.9%) were male. According to parental report, 50 (71.4%) were White, Non-Hispanic; 8 (11.4%) were Asian; 3 (4.3%) were Hispanic/Latino; 1 (1.4%) was Black/African American; 3 (4.3%) selected “Other;” 4 (5.7%) selected multiple categories; and 1 (1.4%) chose not to report race. The discrepancy in the total percentage is due to standard rounding. In terms of ethnicity, 7 (10%) were identified as Hispanic or Latino, 60 (85.7%) were identified as Not Hispanic or Latino, and 3 declined to provide a response (4.3%).

To examine whether participants in the two conditions differed on demographic or background variables, we conducted preliminary analyses with chi-squares and independent samples *t*-tests. The number of participants who were excluded from analysis did not significantly differ between the two conditions,  $p = ns$ . Participants in the two conditions did not differ in terms of age, race, ethnicity, or sex,  $ps = ns$ . Participants in the two conditions also did not differ in baseline negative affect or parent-reported positive or negative affect over “the past few weeks” leading up to the experiment,  $ps = ns$ . However, participant reports of baseline positive affect were significantly higher for participants in the experimental condition ( $M = 19.44$ ,  $SD = 4.78$ ) than those in the control condition ( $M = 16.36$ ,  $SD = 5.65$ ),  $t(67) = 2.45$ ,  $p = 0.017$ . There was a small, significant correlation between these self-reported baseline positive affect scores and parent-reported positive affect over the weeks before the experiment,  $r(68) = 0.28$ ,  $p = 0.020$ . In order to account for these baseline differences in the calculation of the results, we performed a one-way analysis of covariance (ANCOVA) adjusting for baseline self-reported positive affect along with the planned analysis to further probe the effects of the robot on changes in positive affect.

We used Pearson Product Moment Correlations to check for redundancy among the baseline self-report measures. We used a threshold of 0.71, which denotes a shared variance of 50%. As negative and positive affect are theoretically distinct constructs, we hypothesized that the two would not be significantly correlated (Watson & Tellegen, 1985; Zevon & Tellegen, 1982). As predicted, positive and negative affect were not significantly correlated,  $r(69) = -0.18$ ,  $p = 0.138$ . Positive affect was significantly correlated but not redundant with all dimensions of emotional response, as measured by the SAM: Pleasure,  $r(69) = 0.27$ ,  $p = 0.024$ ; Arousal,  $r(69) = 0.35$ ,  $p = 0.003$ ; and Dominance,  $r(69) = 0.35$ ,  $p = 0.003$ . Negative affect was significantly negatively correlated but not redundant with Pleasure,  $r(69) = -0.46$ ,  $p < 0.001$ . Negative affect was not significantly correlated with Arousal,  $r(69) = 0.07$ ,  $p = 0.527$ , or with Dominance,  $r(69) = -0.04$ ,  $p = 0.751$ . Finally, the Arousal and Dominance domains of children’s emotional response were significantly correlated but not redundant with one another,  $r(69) = 0.45$ ,  $p < 0.001$ .

As three different experimenters led participants through the procedure, we examined whether children’s baseline scores or change in self-reported measures of stress differed across experimenters. A one-way ANOVA revealed that there was no significant difference in baseline or change in self-reported positive affect, negative affect, or emotional response (Pleasure, Arousal, or Dominance) across the experimenters,  $ps = ns$ . Moreover, since it was not possible to keep experimenters masked to condition, the overall friendliness of the judges in the TSST-C was rated in a set of approximately 45% of the total participants (35 cases). A research assistant coded the judges’ friendliness in video clips that had been cropped to exclude any evidence of condition. A reliability coder rated 29% of this subset of videos (10 cases). Inter-rater reliability was assessed with a two-way mixed, consistency, single measures ICC. The reliability was good, with an ICC of 0.73 (Cicchetti, 1994). These results indicate that both coders rated the judges’ friendliness similarly. When these ratings of judge friendliness were compared across conditions, there was no significant difference in judge friendliness between the experimental ( $M = 3.26$ ,  $SD = 0.56$ ) and control groups ( $M = 3.22$ ,  $SD = 0.31$ ),  $t(33) = 0.28$ ,  $p = 0.782$ .

Finally, we predicted that children’s baseline levels of social anxiety would impact the degree to which their self-reported positive affect, negative affect, and emotional response changed during the procedure. As previously mentioned, we excluded self-reported social anxiety due to low internal consistency. Contrary to our predictions, a Pearson Product Moment Correlation revealed that neither parent-reported overall anxiety nor parent-reported social anxiety was significantly correlated with change in any of the self-reported measures,  $ps = ns$ . As a result, we did not include social anxiety in the main analyses.

### Effects of Interaction with the Robot

We predicted that participants who completed the stressful task in the presence of the Paro robot would show less of a decrease in positive affect than those who completed the task without the robot present. The 2 (condition) x 2 (time) mixed ANOVA revealed a medium, significant effect of condition on positive affect scores,  $F(1, 67) = 4.45$ ,  $p = 0.039$ ,  $\eta^2_p = 0.062$  (Cohen, 1988; Richardson, 2011). Contrary to our prediction, participants in the experimental condition ( $M = -3.28$ ,  $SD = 6.70$ ) showed a significantly greater decrease in positive affect scores after the stressful situation than those in the control condition ( $M = -0.06$ ,  $SD = 5.51$ ). Since parent-reported positive affect was significantly correlated with change in positive affect, we also performed a 2 (condition) x 2 (time) mixed ANOVA with parent-reported positive affect as an adjustment variable. The pattern of results remained constant even after adjusting



for differences in parent-reported positive affect in the weeks before the experiment,  $F(1, 65) = 6.23$ ,  $p = 0.015$ ,  $\eta^2_p = 0.087$ . Finally, in light of the pretest differences in self-reported positive affect, we conducted a one-way ANCOVA adjusting for baseline positive affect in order to account for the possibility that these baseline differences could influence the relative change in the measure over time. This ANCOVA revealed that, accounting for baseline differences in positive affect, there was not a significant effect of condition on positive affect scores,  $F(1, 66) = 1.12$ ,  $p = 0.294$ ,  $\eta^2_p = 0.017$  (see Supplementary Material for a table with these outcomes).

We also predicted that having the SAR present would lead to smaller increases in negative affect in response to the stressful task. In this case, the 2 (condition)  $\times$  2 (time) mixed ANOVA did not reveal a significant effect of condition on change in negative affect,  $F(1, 68) = 0.02$ ,  $p = 0.196$ ,  $\eta^2_p = 0.025$ . Participants in the two conditions did not differ in terms of change in negative affect in response to the stressful task.

Additionally, we predicted that participants in the experimental condition would show diminished changes in three domains of emotional response: Pleasure, Arousal, and Dominance. Contrary to our predictions, a 2 (condition)  $\times$  3 (time) mixed ANOVA did not reveal a significant effect of condition on change in Pleasure,  $F(1, 68) = 0.91$ ,  $p = 0.343$ ,  $\eta^2_p = 0.013$ . The two groups also did not significantly differ in terms of change in Arousal,  $F(1, 68) = 3.37$ ,  $p = 0.059$ ,  $\eta^2_p = 0.051$ . Finally, there was not a significant effect of condition on Dominance,  $F(1, 68) = 0.16$ ,  $p = 0.693$ ,  $\eta^2_p = 0.002$ . Participants in the two conditions did not differ in changes in state Pleasure, Arousal, or Dominance across the three time points examined.

### Unplanned analyses

As discussed above, in order to avoid unnecessarily excluding participants, we included all participants who completed at least 75% of the testing component of the TSST-C. Since all three participants who had completed greater than 75% but less than 100% of the task were in the experimental condition, including these participants might have contributed to the group differences we previously detected in change in positive affect. To explore this possibility, we conducted unplanned analyses to evaluate our main predictions after excluding all participants who did not complete the full stress-induction procedure. In the case of positive affect, the results of the 2 (condition)  $\times$  2 (time) mixed ANOVA did not reveal a significant effect of condition,  $F(1, 64) = 2.83$ ,  $p = 0.098$ ,  $\eta^2_p = 0.042$ . When only those participants who completed the full stress-induction procedure were included in the analysis, there was no difference in change in positive affect across the two conditions.

### Supplementary Analyses

In line with research on therapy animals, we were interested in characterizing the influence of the degree of interaction between the participants and the robot during the TSST-C on participants' responses to the procedure. We conducted exploratory analyses to investigate the potential factors involved in variation across participants in prosocial behaviors towards the robot. We explored whether the duration of participants' gaze towards or touch of the SAR correlated with changes in participants' self-reported positive affect, negative affect, Pleasure, Arousal, or Dominance. However, these correlations were not significant,  $ps = ns$ .

Since the perception of these robots as nonjudgmental is believed to be one of their strengths, we also explored whether participants' anxiety levels were related to the participants' prosocial behaviors towards the robot (see Supplementary Material for a table with these correlations). Parent-reported overall anxiety scores were significantly correlated with the total duration of prosocial behaviors (gaze towards and physical contact with the robot),  $r(30) = 0.37$ ,  $p = 0.043$ , such that participants with higher total anxiety scores engaged in more prosocial behaviors towards the robot. Parent-reported social anxiety was also significantly correlated with total duration of prosocial behaviors,  $r(30) = 0.44$ ,  $p = 0.014$ .

To better understand this connection, we looked at possible correlations between parent-reported anxiety and subcategories of this overall duration of prosocial behavior. Parent reports of the participants' overall anxiety level were significantly correlated with the overall duration of the participants' gaze towards the robot,  $r(30) = 0.54$ ,  $p = 0.002$ . We also investigated potential differences during the three segments of the testing component of the TSST-C. Parent-reported overall anxiety was significantly correlated with participant gaze during all three subparts of the TSST-C: the story,  $r(30) = 0.52$ ,  $p = 0.004$ ; the instructions in between the two parts,  $r(30) = 0.39$ ,  $p = 0.035$ ; and the math,  $r(30) = 0.54$ ,  $p = 0.002$ . Parent-reported social anxiety was also significantly correlated with the overall duration of participants' gaze towards the robot,  $r(30) = 0.43$ ,  $p = 0.017$ . Furthermore, parent-reported social anxiety was significantly correlated with gaze at the robot during two components of the TSST-C: the story,  $r(30) = 0.45$ ,  $p = 0.013$ , and the math,  $r(30) = 0.41$ ,  $p = 0.026$ . Parent-reported social anxiety was not significantly correlated with participant gaze at the robot during the instructions in between the two tasks of the TSST-C,  $r(30) = 0.23$ ,  $p = 0.223$ .

Parent-reported anxiety levels were not significantly correlated with the overall duration of the participant's physical contact with the robot, overall anxiety:  $r(30) = 0.22$ ,  $p = 0.25$ , social anxiety:  $r(30) = 0.36$ ,  $p = 0.051$ .

Parent-reported overall anxiety was also not significantly correlated with participants' physical contact with the robot during the individual subparts of the TSST-C,  $ps = ns$ . However, parent-reported social anxiety was significantly correlated with the duration of physical contact with the robot during the instructions in between the story and math segments of the TSST-C,  $r(30) = 0.41$ ,  $p = 0.024$ . Parent-reported social anxiety was not significantly correlated with duration of physical contact during the story,  $r(30) = 0.31$ ,  $p = 0.096$ , or during the math,  $r(30) = 0.35$ ,  $p = 0.059$ .

## Discussion

We did not find evidence that having the Paro robot present during a stressful task buffered against the onset of stress-related symptoms in children ages seven to ten. In fact, before accounting for differences in baseline positive affect, those who completed the stressful task in the presence of the Paro robot showed a greater reduction in positive affect after the task than those who did not have the robot present. This finding contrasts with the results of our group's previous study, which found that interacting with the Paro robot after a stressful task led to increased positive affect relative to waiting for the same amount of time or interacting with the robot while its robotic features were turned off (Crossman et al., 2018). One possible explanation for this seemingly deleterious effect on positive affect is that the presence of the Paro robot during this period of intense concentration and stress interrupted the participants' own coping mechanisms. The TSST-C requires children to focus on the task at hand. Perhaps the presence of the robot distracted the child and thus served as a hindrance to their feelings of accomplishment. Furthermore, although participants were explicitly encouraged to use the SAR as much as they wanted during the TSST-C, it is possible that children felt a social expectation to avoid playing with the SAR (or stuffed animals in a general context) in front of authority figures. In fact, perhaps this social pressure to refrain from interacting with the SAR served as an additional stressor during the task. It is conceivable that the SAR functions best after the stressor, as indicated by our group's prior work, but this study raises questions about the consistency of the effects of SARs on children's stress. On the whole, when examining the two existing studies on the effects of SARs on children's stress as a pair, we do not see consistent effects.

While it is plausible that the robot's presence might have negatively impacted participants' positive affect during the stressful situation, a methodological explanation could alternatively account for this unexpected pattern of results. Baseline levels of positive affect differed between conditions; on average, participants in the experimental condition began the procedure with higher levels of positive affect

than their counterparts in the control condition. Thus, apparent reductions in positive affect among participants in the experimental condition may have resulted from this failure of randomization rather than from the effects of the presence of the robot. Support for this possibility comes from the fact that the effect of condition on positive affect was no longer significant when we accounted for the difference in baseline scores.

In addition to the possibility that these results were an artifact of differences in baseline positive affect, our inclusion criteria might have contributed to this finding. In order to avoid being too liberal in the exclusion of participants, we elected to include all who had completed the majority of the task. We reasoned that all participants who experienced at least three-quarters of the task had been exposed to the same key aspects of the task as participants who completed the full task. However, it could be argued that the need to stop the procedure reflected the fact that these participants experienced the task as more stressful than other participants. One might expect that those who felt the stressor to the greatest degree would report the greatest drop in positive affect after the task. Alternatively, because all participants who fell into this category were in the experimental condition, it is possible that the presence of the robot contributed to their ability to complete the majority, but not all, of the stressful task. When these participants were removed from the analyses, the effect of condition on positive affect was no longer significant. Thus, on the basis of these methodological considerations, our results do not provide evidence for a relation between interacting with the robot during a stressful situation and changes in positive affect.

We did not detect a significant relation between parent-reported social anxiety and changes in self-reported stress measures. These results might reflect the restricted range of the sample. We chose to use a community sample in order to address the pressing need to address childhood stress on a broad scale. Because we did not specifically select a sample with high levels of social anxiety, the majority of participants tended to fall into the lower range of social anxiety scores.

Although we did not detect a significant correlation between social anxiety and measures of distress, we did find that social anxiety was significantly correlated with children's prosocial behaviors towards the robot during the testing component of the TSST-C. Children with higher levels of social anxiety demonstrated increased prosocial behaviors towards the robot during times of psychosocial stress; thus, this robot might be a promising target for providing social support to these children who might struggle to receive social support from other sources (Condren et al., 2002). However, the increased duration of prosocial behaviors towards the robot did not lead to buffering against the onset of stress-related symptoms. In

addition, it is possible that the robot provided an opportunity for distraction from the task, such as a way to avoid making eye-contact with the experimenters, rather than a form of social support. Further research would be necessary to disentangle whether this result marks an adaptive form of seeking support or rather a manifestation of symptoms related to higher levels of social anxiety.

Several specific limitations might have precluded us from detecting an effect of the presence of the SAR on children's stress in the current study. First, this study used only a single robot. A wide variety of SARs exist, each with its own unique set of capabilities and functions (Pennisi et al., 2016). It is entirely possible that a different SAR would have had a different effect in this situation. Future research might explore the particular qualities that help a robot optimally function in this role. For instance, while we used a robot based on a baby seal, a more familiar animal might prove more beneficial. As SARs are continuously being developed, this line of research could ensure that new technologies are maximally effective in addressing the pressing need for interventions that reduce children's stress. The current study provides a useful paradigm for future research evaluating the potential benefits of other SARs for this population. Future research could employ a similar methodology to investigate the stress-buffering potential of robots with other features in order to investigate which robotic capacities might be particularly well-suited to such a role.

Second, considering the discrepancy between our group's findings when the SAR was offered to children during versus after the stressor, future research might also continue to examine the ideal timing for interventions with SARs. Although our group measured emotional response in anticipation of the stressful task, more frequent sampling of key outcome variables could test the nuances of the effects of the SAR on children's stress responses in anticipation of or during the stressor. Additionally, the timing of intervention before, during, or after a stressor could be manipulated in a randomized controlled trial to more directly examine the optimal timing for interventions with SARs for children's stress.

Third, it is essential to consider the potentially limited generalizability of our results in the context of the relatively homogeneous demographic composition of our sample and limited breadth of demographic data collected. It is of particular importance for research addressing childhood stress to consider the representativeness of the sample given that variations in children's exposure to stress in their daily lives may have implications for how they respond to novel stressful situations in a lab setting. Previous research has demonstrated that a range of factors, including socioeconomic status and temperament, have the capacity to influence stress responses in the laboratory (Evans et al., 2013). Keeping this in mind, it is possible that we would see

different results with a more diverse sample of participants. In order to enhance the generalizability of findings, it is necessary to include historically underrepresented populations in future research on interventions intended to alleviate stress.

Finally, all indicators of participant stress were self-report measures. Excessive reliance on self-report data raises the concern of introducing bias through the participants' knowledge that they are being assessed (Kazdin, 2016). The validity of exclusive reliance on self-report measures is particularly debated in child samples due to questions of whether children are able to accurately report the nuances of their emotions (e.g., Blair, 2000). Thus, these measures of stress might not have fully captured the extent of the participants' experiences over the course of the procedure. However, children have been found to successfully report on their own internalizing problems as early as 5 years of age (e.g., Jalongo et al., 2001). Furthermore, to address potential concerns about the effect of variance in verbal and cognitive abilities on children's ability to self-report on these metrics of distress, we included the SAM in the analysis. This pictorial measure is designed to circumvent possible impediments in verbal and cognitive capacities. Although we accounted for potential complications involved with child self-report measures, future research could explore such self-report measures in conjunction with behavioral or physiological data.

Overall, these findings suggest that the benefits of incorporating SARs into children's mental health care might be more limited than theoretical arguments and existing applications in practice would suggest. While we found increases in positive affect when interacting with the robot after a stressful task in our previous study, we did not see beneficial effects on positive affect when interacting with the robot during the task. These robots are being produced and implemented at a rapid rate; it is of the utmost importance that future research continues to evaluate the role of this technology in children's mental health care in order to negate the potential risk for harm and ensure that the robots are truly effective in alleviating stress. We currently do not have a clear understanding of when or if SAR interventions effectively reduce children's stress, but this study provides a useful starting point for informing future investigation into the use of these robots with this population by examining their efficacy in a controlled design. These findings indicate the importance of using caution when implementing such interventions until future research addresses these lingering questions.

**Acknowledgements** This research was supported in part by the Jack Parker Foundation, Inc. The Paro robots were provided for use in this research by Takanori Shibata, PhD of the Intelligent System Research Institute, who designed and developed the robot. For further information on the Paro robot, see <http://www.parorobots.com/index.asp>.

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

## References

- American Psychological Association. (2017). Stress in America: coping with change [Press release]. Retrieved from <http://www.apa.org/news/press/releases/stress/2016/coping-with-change.pdf>
- Arendt, K., Hougaard, E., & Thastum, M. (2014). Psychometric properties of the child and parent versions of Spence Children's Anxiety Scale in a Danish community and clinical sample. *Journal of Anxiety Disorders, 28*, 947–956.
- Bavelier, D., Green, C. S., & Dye, M. W. (2010). Children, wired: for better and for worse. *Neuron, 67*, 692–701.
- Betz, A., Kotschal, K., Turner, D. C., Hediger, K., Uvnäs-Moberg, K., & Julius, H. (2011). The effect of a real dog, toy dog and friendly person on insecurely attached children during a stressful task: an exploratory study. *Anthrozoös, 24*, 349–368.
- Björling, E. A., Rose, E., Davidson, A., Ren, R., & Wong, D. (2019). Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot. *International Journal of Social Robotics, 12*, 65–77.
- Blair, J. (2000). Assessing Protocols for Child Interviews. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman & V. S. Cain (Eds.), *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Boydell, K. M., Hodgins, M., Pignatiello, A., Teshima, J., Edwards, H., & Willis, D. (2014). Using technology to deliver mental health services to children and youth: a scoping review. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 23*, 87–99.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.
- Broadbent, E., Feerst, D. A., Lee, S. H., Robinson, H., Albo-Canals, J., Ahn, H. S., & MacDonald, B. A. (2018). How could companion robots be useful in rural schools? *International Journal of Social Robotics, 10*, 295–307.
- Buske-Kirschbaum, A., Jobst, S., Wustmans, A., Kirschbaum, C., Rauh, W., & Hellhammer, D. (1997). Attenuated free cortisol response to psychosocial stress in children with atopic dermatitis. *Psychosomatic Medicine, 59*, 419–426.
- Bynion, T. M., & Feldner, M. T. (2017). Self-assessment manikin. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. Cham: Springer.
- Cabibihan, J. J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics, 5*, 593–618.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284.
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. Heimberg, M. Liebowitz, D. A. Hope & F. R. Schneier (Eds.), *Social Phobia: Diagnosis, Assessment and Treatment* (pp. 69–93). New York, NY: Guilford.
- Cohen, D. (1988). *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- Condren, R. M., O'Neill, A., Ryan, M. C. M., Barrett, P., & Thakore, J. H. (2002). HPA axis response to a psychological stressor in generalised social phobia. *Psychoneuroendocrinology, 27*, 693–703.
- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*, 837–844.
- Crossman, M. K. (2016). Effects of interactions with animals on human psychological distress. *Journal of Clinical Psychology, 73*, 761–784.
- Crossman, M. K., Kazdin, A. E., & Kitt, E. R. (2018). The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Professional Psychology: Research and Practice, 49*, 48–56.
- Danese, A., Moffitt, T. E., Harrington, H., Milne, B. J., Polanczyk, G., Pariante, C. M., & Caspi, A. (2009). Adverse childhood experiences and adult risk factors for age-related disease: Depression, inflammation, and clustering of metabolic risk markers. *Archives of Pediatrics & Adolescent Medicine, 163*, 1135–1143.
- Dawe, J., Sutherland, C., Barco, A., & Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatrics Open, 3*, 1–16.
- de Jong, C., Peter, J., Kühne, R., & Barco, A. (2019). Children's acceptance of social robots: a narrative review of the research 2000–2017. *Interaction Studies, 20*, 393–425.
- DeSousa, D. A., Pereira, A. S., Petersen, C. S., Manfro, G. G., Salum, G. A., & Koller, S. H. (2014). Psychometric properties of the Brazilian-Portuguese version of the Spence Children's Anxiety Scale (SCAS): Self-and parent-report versions. *Journal of Anxiety Disorders, 28*, 427–436.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355–391.
- Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-item Positive and Negative Affect Schedule for Children, Child and Parent Shortened Versions: Application of item response theory for more efficient assessment. *Journal of Psychopathology and Behavioral Assessment, 34*, 191–203.
- Evans, B. E., Greaves-Lord, K., Euser, A. S., Tulen, J. H., Franken, I. H., & Huizink, A. C. (2013). Determinants of physiological and perceived physiological stress reactivity in children and adolescents. *PLoS ONE, 8*, e61724.
- Feil-Seifer, D., & Mataric, M. (2005). "Defining socially assistive robotics." 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*, 143–166.
- Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution, 7*, 1325–1330.
- Fryers, T., & Brugha, T. (2013). Childhood determinants of adult psychiatric disorder. *Clinical Practice and Epidemiology in Mental Health, 9*, 1–50.
- Gunnar, M. R., Frenn, K., Wewerka, S. S., & Van Ryzin, M. J. (2009). Moderate versus severe early life stress: associations with stress reactivity and regulation in 10–12-year-old children. *Psychoneuroendocrinology, 34*, 62–75.
- Gunnar, M. R., Talge, N. M., & Herrera, A. (2009). Stressor paradigms in developmental studies: What does and does not work to produce mean increases in salivary cortisol. *Psychoneuroendocrinology, 34*, 953–967.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23–34.
- Hayden, E. P., Olino, T. M., Mackrell, S. V., Jordan, P. L., Desjardins, J., & Katsiroumbas, P. (2013). Cognitive vulnerability to depression during middle childhood: Stability and associations with maternal affective styles and parental depression. *Personality and Individual Differences, 55*, 892–897.

- Hong, R. Y., Lee, S. S., Tsai, F. F., & Tan, S. H. (2017). Developmental trajectories and origins of a core cognitive vulnerability to internalizing symptoms in middle childhood. *Clinical Psychological Science*, 5, 299–315.
- Ialongo, N. S., Edelsohn, G., & Kellam, S. G. (2001). A further look at the prognostic power of young children's reports of depressed mood and feelings. *Child Development*, 72, 736–747.
- Japan's National Institute of Advanced Industrial Science and Technology [AIST] (2004). *Seal-type robot "PARO" to be marketed with best healing effect in the world* [Press release]. Retrieved from <http://www.parorobots.com/pdf/pressreleases/PARO%20to%20be%20marketed%202004-9.pdf>
- Kazdin, A. E. (2016). *Research Design in Clinical Psychology*. 5th ed. Boston, MA: Pearson Education Limited.
- Kazdin, A. E. (2019). Annual research review: expanding mental health services through novel models of intervention delivery. *Journal of Child Psychology and Psychiatry*, 60, 455–472.
- Kerns, K. A., Stuart-Parrigon, K. L., Coifman, K. G., van Dulmen, M. H., & Koehn, A. (2018). Pet dogs: Does their presence influence preadolescents' emotional responses to a social stressor? *Social Development*, 27, 34–44.
- Kertes, D. A., Liu, J., Hall, N. J., Hadad, N. A., Wynne, C. D., & Bhatt, S. S. (2017). Effect of pet dogs on children's perceived stress and cortisol stress response. *Social Development*, 26, 382–401.
- Krämer, M., Seefeldt, W. L., Heinrichs, N., Tuschen-Caffier, B., Schmitz, J., Wolf, O. T., & Blechert, J. (2012). Subjective, autonomic, and endocrine reactivity during social stress in children with social phobia. *Journal of Abnormal Child Psychology*, 40, 95–104.
- Lau, B. W. K. (2002). Does the stress in childhood and adolescence matter? A psychological perspective. *The Journal of the Royal Society for the Promotion of Health*, 122, 238–244.
- Logan, D. E., Breazeal, C., Goodwin, M. S., Jeong, S., O'Connell, B., Smith-Freedman, D., & Weinstock, P. (2019). Social robots for hospitalized children. *Pediatrics*, 144, 1–13.
- Longman, P., Mundy, L., Black, R., Bornfreund, L., Byrum, G., Cramer, R., McCarthy, M. A. (2019). *Strengthening Ties: The Case for Building a Social Policy Centered on Families*. Retrieved from <http://files.eric.ed.gov/fulltext/ED558749.pdf>
- Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., & Drexler, S. M. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, 77, 247–285.
- Marino, F., Chilà, P., Sfrassetto, S. T., Carrozza, C., Crimi, I., Failla, C., & Ruta, L. (2019). Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 49, 1–15.
- Matarić, M. J., & Scassellati, B. (2016). Socially assistive robotics. In *Springer Handbook of Robotics* (pp. 1973–1994). Springer, Cham.
- Middlebrooks, J. S., & Audage, N. C. (2008). *The Effects of Childhood Stress on Health Across the Lifespan*. Atlanta: National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention.
- Moyle, W., Jones, C. J., Murfield, J. E., Thalib, L., Beattie, E. R., Shum, D. K., & Draper, B. M. (2017). Use of a robotic seal as a therapeutic tool to improve dementia symptoms: A cluster-randomized controlled trial. *Journal of the American Medical Directors Association*, 18, 766–773.
- Nauta, M. H., Scholing, A., Rapee, R. M., Abbott, M., Spence, S. H., & Waters, A. (2004). A parent-report measure of children's anxiety: Psychometric properties and comparison with child-report in a clinic and normal sample. *Behaviour Research and Therapy*, 42, 813–839.
- Padilla-Walker, L. M., Coyne, S. M., & Memmott-Elison, M. K. (2019). Media and the family. In B. H. Fiese, M. Celano, K. Deater-Deckard, E. N. Jouriles, & M. A. Whisman (Eds), *APA handbooks in psychology®. APA handbook of contemporary family psychology: Applications and broad impact of family psychology* (p. 365–378). American Psychological Association.
- Pearson, Y., & Borenstein, J. (2014). Creating "companions" for children: the ethics of designing esthetic features for robots. *AI & Society*, 29, 23–31.
- Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: a systematic review. *Autism Research*, 9, 165–183.
- Przybylski, A. K., Orben, A., & Weinstein, N. (2021). How much is too much? Examining the relationship between digital screen engagement and psychosocial functioning in a confirmatory cohort study. *Journal of the American Academy of Child & Adolescent Psychiatry*. in press
- Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the Goldilocks Hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science*, 28, 204–215.
- Rabbitt, S. M., Kazdin, A. E., & Hong, J. H. (2015). Acceptability of robot-assisted therapy for disruptive behavior problems in children. *Archives of Scientific Psychology*, 3, 101–110.
- Rabbitt, S. M., Kazdin, A. E., & Scassellati, B. (2015). Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. *Clinical Psychology Review*, 35, 35–46.
- Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, 35, 741–756.
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6, 135–147.
- Robinson, H., MacDonald, B., Kerse, N., & Broadbent, E. (2013). The psychosocial effects of a companion robot: a randomized controlled trial. *Journal of the American Medical Directors Association*, 14, 661–667.
- Rose, E. J., Björling, E., & Cakmak, M. (2019). Participatory design with teens: a social robot design challenge. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 604–609). ACM.
- Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., & McGuinn, L., Committee on Early Childhood, Adoption, and Dependent Care. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, 129, e232–e246.
- Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, 36, 545–566.
- Spence, S. H., Barrett, P. M., & Turner, C. M. (2003). Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. *Journal of Anxiety Disorders*, 17, 605–625.
- Tolksdorf, N. F., Siebert, S., Zorn, I., Horwath, I., & Rohlfing, K. J. (2021). Ethical considerations of applying robots in kindergarten settings: Towards an approach from a macroperspective. *International Journal of Social Robotics*, 13, 129–140.
- Wang, M., Meng, Q., Liu, L., & Liu, J. (2016). Reliability and validity of the spence children's anxiety scale for parents in mainland Chinese children and adolescents. *Child Psychiatry & Human Development*, 47, 830–839.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–235.
- Westlund, J. K., Breazeal, C., & Story, A. (2015). Deception, secrets, children, and robots: What's acceptable. Paper presented at the

- 10th ACM/IEEE International Conference on Human-Robot Interaction Portland, OR.
- UNICEF. (2017). *How does the time children spend using digital technology impact their mental well-being, social relationships and physical activity? An evidence-focused literature review*. Retrieved from <https://www.unicef-irc.org/publications/pdf/Children-digital-technology-wellbeing.pdf>
- U.S. Census Bureau. (2019). *New Haven County, Connecticut, Connecticut*. Retrieved from <https://www.census.gov/quickfacts/table/newhavencountyconnecticut,newhavencityconnecticut/PST045219>
- Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: an idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, *43*, 111–122.