2018

Effectiveness and Acceptability of a Robot-Based Social Skills Intervention for Children with Autism Spectrum Disorder

Sarah Alberta Koch
University of Alabama at Birmingham

EFFECTIVENESS AND ACCEPTABILITY OF A ROBOT-BASED SOCIAL SKILLS INTERVENTION FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

by

SARAH A. KOCH

MARIA I. HOPKINS, COMMITTEE CHAIR
FRANKLIN R. AMTHOR
FRED J. BIASINI
SCOTT W. SNYDER
LAURA A. STOPPELBEIN

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2018

EFFECTIVENESS AND ACCEPTABILITY OF A ROBOT-BASED SOCIAL SKILLS INTERVENTION FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

SARAH A. KOCH

MEDICAL/CLINICAL PSYCHOLOGY

ABSTRACT

This study examined the effectiveness, feasibility, and acceptability of a robot-based intervention program designed to improve social-emotional skills in school-age children with Autism Spectrum Disorder (ASD). Twenty-two children with ASD and mild-to-no cognitive impairment were randomized to intervention (n = 11) or waitlist control groups (n = 11) for eight weeks. Participants who completed the robot-based intervention displayed decreased overall engagement from baseline to post-intervention, based on an eye-tracking measure. Nonetheless, they reported high favorability ratings at post-intervention, including consistently high ratings of happiness, increased comfort ratings, and only slightly decreased ratings of desire for future interactions across time. Group comparisons indicated significant improvement in overall accuracy for identifying face drawings and photos corresponding with robotic emotional facial expressions for individuals in the intervention group. There were no group differences for amount of socially directed gaze with the robot during baseline and post-intervention sessions. Similarly, there were no group differences over time for generalized affect recognition and theory of mind skills. Taken together, results support the use of the robot-based intervention within this population as a tool for promoting an enjoyable learning environment conducive to skill development. Improved accuracy within the intervention group for matching robotic facial expressions, along with decreased visual engagement at post-intervention, suggests a shift from effortful processing to more automatic responding as a result of training. However, it is unclear whether this skill improvement resulted from learning of specific facts or the development of more generalized emotion decoding and understanding. Given strong baseline scores on robot-specific and generalized measures of emotion knowledge, results suggest that the information presented in the intervention may have been too simplistic for the sample included in the study, and future research will examine the efficacy and ultimate benefit of this tool within other subsets of children with ASD.

Keywords: autism spectrum disorder, social robot, robot-based intervention, social-emotional skills, emotion recognition

ACKNOWLEDGMENTS

This dissertation research represents the collective effort of many invaluable individuals. I would like to thank the Social Technology for Autism Research (STAR) lab, particularly Dr. Hopkins and Dr. Biasini, for their continued guidance and support throughout my graduate training. I would like to thank Carl Stevens for his vital role in the construction and programming of the robot. I would also like to thank each member of my committee for their helpfulness and feedback throughout the many phases of this project, including initial conceptualization of the robot, study design and implementation, data analyses, and finally preparation and editing of this written document. Additionally, I would like to thank Mitchell’s Place and the Autism Society of Alabama for aiding in recruitment, as well as the children and families who participated in the study and made this research possible. This study was funded in part by a grant from Civitan International Research Center.

TABLE OF CONTENTS

LIST OF TABLES

Table *Page*

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADI	Autism Diagnostic Interview
ADOS	Autism Diagnostic Observation Schedule
ANOVA	analysis of variance
APA	American Psychiatric Association
AR	Affect Recognition
ASD	Autism Spectrum Disorder
CDC	Centers for Disease Control and Prevention
CHANGE	Recognizing Change in Expression
CI	confidence interval
DSM	Diagnostic and Statistical Manual of Mental Disorders
FACE	Facial Automaton for Conveying Emotions
INFER	Inferring Emotion
KBIT	Kaufman Brief Intelligence Test
LL	lower limit
MATCH-D	Expression Matching: Schematic Drawings
MATCH-F	Expression Matching: Face Photos
MODEL	Modeling Emotional Expressions
NEPSY	Developmental Neuropsychological Assessment
R-BESK	Robot-Based Emotion Skills Knowledge
ROIs	regions of interest

RRB	Restricted and Repetitive Behavior
SAM	Socially Animated Machine
SCQ	Social Communication Questionnaire
STAR	Social Technology for Autism Research
T1	baseline assessment
T2	post-intervention assessment
TAII	technology aided instruction and intervention
TOM	Theory of Mind
UAB	University of Alabama at Birmingham
UL	upper limit

INTRODUCTION

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disability that affects approximately 1 out of 68 children in the United States (Centers for Disease Control and Prevention, 2016). ASD is a lifelong disorder that occurs across all racial, ethnic, and socio-economic groups (Anderson, Liang, & Lord, 2014; CDC, 2016). The disorder is characterized by two broad symptom domains, including significant impairments in social communication and interaction as well as restricted and repetitive patterns of behavior, interests, or activities (American Psychiatric Association, 2013). Within the domain of social communication and interaction, symptoms include deficits in social-emotional reciprocity, nonverbal social communication, and relationship development. Deficits within the domain of restricted and repetitive behaviors include stereotyped or repetitive motor movements, object use, and speech; ritualized behaviors and adherence to routines; restricted interests; and abnormal sensory responses. Although type and severity of symptoms manifest differently across individuals with ASD, observed symptoms present in early childhood and cause significant impairment in functioning.

Social Skills Impairments

Impairments across both domains of symptomology can have a negative impact on the social functioning of an individual with ASD. Impairments in social-emotional reciprocity include difficulty maintaining interactions and conversation. Individuals with

ASD have trouble understanding the pragmatics of social interaction, such as turn-taking, perspective-taking, and using context appropriately when talking with others (Hobson, 1986; Paul, 2008). Their conversations often lack a sense of social reciprocity, as they tend to ask fewer questions, provide limited responses, and dominate conversations by focusing on personal interests (Tager-Flusberg, Paul, & Lord, 2005). Other common problem areas include reduced sharing of interests and emotions as well as difficulty interpreting nonliteral language and humor (Shaked & Yirmiya, 2003; Tager-Flusberg & Joseph, 2003).

Individuals with ASD also have marked difficulties in producing and perceiving nonverbal social behaviors when interacting with others. This often includes unusual eye contact, such as reduced gaze, excessive staring, and failure to integrate eye contact with other social behaviors (Buitelaar, van Engeland, de Kogel, de Vries, & van Hooff, 1991; Volkmar, Paul, Klin, & Cohen, 2005). Individuals with ASD tend to use fewer facial expressions and gestures and often fail to read and accurately interpret these nonverbal behaviors when used by others (Hobson, 1986). They also have difficulty understanding social boundaries such as maintaining personal space during conversation (Kennedy & Adolphs, 2014). As result of such social difficulties, many individuals with ASD struggle to develop and maintain fulfilling relationships. They often display less interest in peers, have difficulty finding appropriate friends, and experience poorer quality friendships (Bauminger & Kasari, 2000; Church, Alisanski, & Amanullah, 2000).

Although not classified within the domain of social communication and interaction, restricted and repetitive interests and behaviors can also have negative social implications for individuals with ASD. Symptoms within this domain can influence how these individuals approach social situations and how they are perceived and accepted by others. For example, individuals with ASD may spend extended amounts of time fixated on restricted interests and therefore have fewer opportunities to connect with peers. Restricted interests can also negatively impact the appropriateness of social interactions, as individuals with ASD may focus solely on personal interests in conversation and have difficulty discussing and relating to other topics (Tager-Flusberg et al., 2005). Additionally, stereotyped and repetitive behaviors and movements, such as spinning objects, hand flapping, and echoing speech, may seem unusual to peers and can be socially stigmatizing (Esbensen, Seltzer, Lam, & Bodfish, 2009).

Long-Term Outcomes

As social interactions are pervasive in daily life, positive social skills are crucial for facilitating successful communication with others (Bellack & Hersen, 1979). Individuals with strong social skills are able to express themselves verbally through spoken language, tone of voice, and volume of speech. They also frequently utilize nonverbal social behaviors, such as gestures, facial expressions, and body language, while reading and adapting to these cues from others. The ability to engage others in conversation and display empathy helps to create social bonds. Forming close and supportive friendships is particularly important in childhood, as it leads to enhanced self-awareness and self-esteem (Bukowski, Newcomb, & Hartup, 1988).

Alternatively, children who lack proficient social skills, such as those with ASD, may have difficulty building a network of supportive friends. Failure to establish strong positive peer relationships can lead to seclusion and low self-worth (Bukowski, Newcomb, & Hartup, 1988). The long-term implications of social skills deficits can be significant. Social skills impairments do not typically remit with development, but are often exacerbated in adolescence and adulthood as the social world becomes more complex and demanding (Tantam, 2003). When integrated with typically developing peers, adolescents with ASD are at an increased risk for peer rejection and social isolation (Chamberlain, 2001). Nearly half of adolescents and most adults with ASD lack even one close friend (Eaves & Ho, 2008; Howlin, Goode, Hutton, & Rutter, 2004). Furthermore, social skills deficits may contribute to academic and occupational underachievement, as well as mood and anxiety problems later in development (Howlin & Goode, 1998; Tantam, 2003). Adolescents and adults with ASD report increased rates of social anxiety, loneliness, and depression as well as lower levels of overall life satisfaction (Bellini, 2004; Mazurek, 2014; White & Roberson-Nay, 2009).

Social Skills Interventions

The increasing prevalence of ASD, coupled with the devastating long-term implications, necessitates an immediate effort to improve social skills among children with this disorder. Several strategies have been developed to address social skills difficulties in children with ASD, which differ in terms of underlying theoretical principles, method of administration, and tools utilized for implementation. Effective social skills interventions documented in previous literature have been categorized to fit within five domains, including child-specific interventions, environmental modifications, peer-mediated interventions, collateral skills interventions, and comprehensive interventions (Bellini, Peters, Benner, & Hopf, 2007; McConnell, 2002; McEvoy, Odom, & McConnell, 1992).

Child-specific interventions involve direct instruction and reinforcement to improve specific social behaviors. This domain captures a variety of techniques, including general instructional interventions, high-frequency prompting and reinforcement to prime social responding, and social skills training. Environmental interventions are those that promote the use and development of social interaction through modifications to the physical and social environment, such as structured activities and developmentally integrated peer groups. Peer-mediated interventions involve working with typically developing peers to promote interactive behaviors, such as prompting and praise, that can improve the social functioning of children with ASD. Collateral skills interventions are those where improvements in social behaviors result from direct training in related skills, such as play and language. Finally, comprehensive interventions are those that incorporate techniques from two or more domains (Bellini, Peters, Benner, & Hopf, 2007; McConnell, 2002; McEvoy, Odom, & McConnell, 1992).

Evidence-based practices identified by the Autism Evidence-Based Practice Review Group (Wong et al., 2014) specifically addressing social skills in school-age children with ASD include intervention approaches from various domains. Within child-specific interventions, the group identified several fundamental applied behavioral techniques, such as discrete trial training, pivotal response training, functional communication training, prompting, time delay, differential reinforcement, redirection, and extinction. Applied-behavioral techniques rely on basic behavioral principles, such as positive reinforcement, to bring about a meaningful change in behavior. This type of intervention typically involves one-on-one sessions to address treatment goals that are developed and specifically modified for the unique needs of the individual child (White, Keonig, & Scahill, 2007).

Social skills training was also identified by the Autism Evidence-Based Practice Review Group (Wong et al., 2014) as an effective child-specific intervention. Social skills training involves individual and/or group instruction on general communication skills. This technique typically incorporates manualized treatment goals and allows individuals to practice learned skills through in-vivo role-plays. Positive behaviors learned through this technique include eye contact, smiling, sharing, sustaining conversation, listening, and displaying empathy (Bellack & Hersen, 1979). Other effective child-specific strategies recognized by the group include cognitive behavioral interventions, modeling and video modeling, parent training, self-management, social scripts, and social narratives. Environmental interventions such as structured playgroups; peer-mediated instruction; and collateral skills interventions such as visual supports and picture exchange communication systems were also identified as effective for school-age children with ASD (Wong et al., 2014).

Although the intervention strategies listed and described above differ across many factors, they share a common trait in that they must be administered through direct person-to-person contact. This typically involves the presence of a therapist who works directly with the child to develop particular skills. However, it is important to consider that even in a therapeutic setting, social interaction can be confusing and stressful for a child with ASD. Children with ASD have difficulty attending to multiple cues during social interactions, making these interactions complex and difficult to understand

(Koegel, Koegel, Harrower & Carter, 1999). As such, the child may have difficulty learning presented skills due to stress regarding the unpredictability of the situation (Dautenhahn & Werry, 2004; Ricks & Colton, 2010). Furthermore, the limited availability of professional resources coupled with mounting therapy costs necessitates a more accessible form of intervention.

Technology-Based Interventions

Literature suggests that people with ASD feel comfortable in predictable environments and enjoy interacting with technology (Moore, McGrath, & Thorpe, 2000). As an alternative to traditional therapist-administered interventions, researchers have explored the use of technology-based interventions to teach social skills to individuals with ASD. As technology-based tools follow a predictable format, they can reduce the stress and pressure that children with ASD often experience during direct human contact, thereby creating a more enjoyable and effective learning environment (Dautenhahn & Werry, 2004; Ricks & Colton, 2010). In addition to the interventions described above, technology aided instruction and intervention (TAII) has been identified as an evidence-based strategy for improving social skills in school-age children with ASD (Wong et al., 2014). Unlike other therapist-administered interventions that occur in one-on-one or small group sessions, technology-based interventions can also be more easily replicated and distributed to meet the needs of a widespread number of individuals.

One technology-based strategy for improving social skills in children with ASD is computer-administered avatar-based interventions. Avatars are computer embodied virtual people that have a knowledge base and the ability to converse with humans in natural language (Hopkins et al., 2011). Avatars offer an advantage over traditional techniques in that they can stimulate social interaction in a predictable and non-threatening environment. Researchers investigating the use of avatars for children with ASD have reported improvements in facial recognition, emotion recognition, and social interactions following intervention (Baron-Cohen, Golan, Chapman, & Granader, 2007; Hopkins et al., 2011). Past studies have also demonstrated generalization of skills to social interactions with peers in a naturalistic playground setting following intervention (Hopkins et al., 2011).

Recent research also supports the value of using social robots to improve social skills in children with ASD. Similar to avatar-based methods, the use of robots allows for a simplified, safe, predictable, and reliable environment where the complexity of interaction can be controlled (Robins, Dickerson, Stribling, & Dautenhahn, 2004). However, as robots exist and interact in three-dimensional space, they have the ability to generate an interpersonal environment that cannot be achieved with other technology-based strategies. Social robots offer a promising method for social skills intervention, as they seem to "occupy a special niche between inanimate toys and animate social beings" (Scassellati, Admoni, & Mataric, 2012, p. 276). As such, social robots can actively stimulate social communication without eliciting the confusion and distress that can accompany direct human interaction. Social robots may therefore have the ability to elicit certain desirable social behaviors in children with ASD that are not typically seen in other interaction settings (Ricks & Colton, 2010).

Numerous robotic approaches and applications have been developed to improve a variety of social communication skills in individuals with ASD. Several studies have examined whether the presence of a robot can be used to elicit social behaviors during an interaction session. Duquette, Michaud, & Mercier (2008) investigated the use of a mobile robot to facilitate social interaction behaviors in four children with ASD. Children were assigned to the robot group or a group paired with a human mediator. Children paired with the robot displayed increased shared attention and engaged in less repetitive behaviors compared to children paired with a human mediator. Kim et al. (2012) examined how the presence of Pleo, a socially expressive robotic dinosaur, affected social behaviors during a triadic interaction between a child with ASD, an interaction partner, and an adult confederate. The interaction partner varied across conditions and included the presence of Pleo, a touchscreen computer game, or an additional human mediator. In a sample of 24 children with ASD, children spoke more in general and directed more speech to the adult confederate when Pleo served as the interaction partner.

Some studies have also explored the use of robots as tools to teach particular skills that can be learned and eventually transferred to interactions with humans. Huskens, Verschuur, Gillesen, Didden, and Barakova (2012) utilized the NAO robot from Aldebaran Robotics, a humanoid robot with a sleek design and simple face, to implement an applied behavior analysis-based intervention in a group of six children with ASD. Children were randomly assigned to receive the robot intervention or a similar intervention administered by a human trainer. For both groups, the number of self-initiated social questions increased significantly following the intervention, yet there was no difference between groups in skills improvement. Pioggia et al. (2005) developed Facial Automaton for Conveying Emotions (FACE), a life-like android that can recognize and produce basic emotions. A group of four children with ASD participated in an intervention with FACE teaching emotion recognition skills. Following the intervention, each child showed improvement in the categories of Emotional Responses and Relating to People on the Childhood Autism Rating Scale (Pioggia et al., 2008).

Overall, researchers investigating robots as tools for social interaction in ASD have reported increased engagement and positive social behaviors during these interactions (Diehl, Schmitt, Villano, & Crowell, 2012; Feil-Seifer & Mataric, 2009; Ricks & Colton, 2010; Robins et al., 2005; Scassellati, 2007; Stanton et al., 2008). Robots have been shown to help children with ASD express appropriate social skills such as initiating communication, turn-taking, imitation, emotion recognition, eye gaze, and joint attention (Dautenhahn & Werry, 2004; Kozima & Nakagawa, 2006; Pioggia et al., 2005; Pioggia et al., 2008; Robins, Dautenhahn, Te Boekhorst, & Billard, 2005). The results of previous research efforts are promising; however, it can be difficult to draw firm conclusions based on these studies, as most studies are only exploratory, rely on single case examples or small sample sizes, and have methodological limitations (Cho & Ahn, 2016; Diehl, Schmitt, Villano, & Crowell, 2012). While these studies offer valuable insight, many questions still remain regarding the role of robots in social skills interventions for children with ASD.

THE ROBOT SAM

Design Rationale

Acknowledging successes and limitations with varying robotic designs and techniques employed in previous studies, researchers in the Social Technology for Autism Research (STAR) lab at the University of Alabama at Birmingham (UAB) set out to develop a novel social robot and intervention program geared toward improving social-emotional skills in school-age children with ASD. A contending theory of social skills deficits in ASD is that difficulties in recognizing and responding to emotions may underlie core social difficulties associated with the disorder (Baron-Cohen, Golan, & Ashwin, 2009). Emotion recognition skills are thought to be essential to the development of more complex social perception skills, such as identifying one's own mental states (e.g., thoughts, desires, and intentions) and attributing these mental states to others, a concept called "theory of mind" (Ashwin, Chapman, Colle, & Baron-Cohen, 2006; Baron-Cohen, 2000). The goal of addressing such social-emotional difficulties presented specific challenges when considering robotic design, specifically regarding the appearance of the robot and level of anthropomorphism, or "human likeness."

To target skills such as emotion recognition and understanding others' perspectives, a robot must have the ability to demonstrate different emotional states. Though affective states can be displayed through multiple modalities (e.g., vocalizations, body posture, gestures), the face plays a particularly important role in the expression of emotion. Past research suggests that 55% of affective information is displayed via facial expression, whereas spoken language and paralanguage account for only 7% and 38% of this information, respectively (Mehrabian, 1968). In order to accurately convey emotional facial expressions, the robot must maintain a certain level of facial detail and adherence to features observed in the human face. Humanoid robots can offer a better approximation of human-to-human interactions and may provide the greatest potential for skill generalization (Ricks & Colton, 2010).

Additionally, a robot developed for use with children should appear friendly, playful, and approachable and should be interesting and engaging during human-to-robot interactions. As opposed to humanoid robots that more closely imitate the human form, non-humanoid robots can take on many fun and attractive shapes and appearances (e.g., animals, cars, toys) and can be built to efficiently complete specific tasks. This allows for increased flexibility in accentuating key features or social cues necessary for developing specific social skills (Scassellati, Admoni, & Mataric, 2012). These robots also avoid evoking an “uncanny valley” experience (Mori, 1970), by which a decrease in familiarity and acceptance occurs when an artificial agent becomes too human-like (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2011). By evading the reaction of withdrawal and avoidance that can accompany interactions with human-like beings, non-humanoid robots can be particularly engaging to children with ASD (Ricks & Colton, 2010).

Appearance

The STAR lab attempted to bridge this gap in robotic design by creating a humanoid robot with an approachable, animal-like appearance, while preserving the essential features of a human face. The intent was to design a robot that would be interesting and engaging to children with ASD, yet maintain the capability to model facial expressions that convey emotional subtlety. Results of research efforts led to the creation of the humanoid robotic monkey, “Socially Animated Machine” (SAM; Figure 1). SAM was developed by the STAR lab and constructed and programmed by Carl Stevens, a UAB graduate student (Stevens et al., 2015). SAM is designed for desktop use and sits at 50 cm tall. It has a head, torso, two arms, and two non-moving legs. SAM possesses a total of 10 degrees of freedom, all of which are controlled through servo motors. Six servos move SAM’s neck and arms, offering a good range of mobility throughout the upper torso. These motors allow SAM to carry out basic conventional and instrumental gestures, such as clapping, pointing, head nodding, and head shaking.

Figure 1. Socially Animated Machine

The head of SAM, measured at 16 x 16 cm, contains eyes, eyebrows, ears, a nose, and two lips forming a mouth. SAM is equipped with four servos that control different parts of the face, including the eyebrows and mouth. By changing the angle of the eyebrows as well as the curvature of the lips and the gap between them, SAM can produce a variety of facial expressions (Figure 2). The lips also move in synchronization with verbal output. SAM's eyes are displayed on a 7 x 4 cm digital screen. Using a digital display provides a way to capture the more detailed, subtle changes to the eyes and surrounding area (e.g., eye lids, skin folds) that accompany different facial expressions. SAM's eyes are rounded and exaggerated, similar to those seen in a cartoon character. With this setup, SAM has been programmed to model the six universally recognized emotional expressions: happiness, sadness, anger, fear, surprise, and disgust (Ekman & Friesen, 1975; Figure 3).

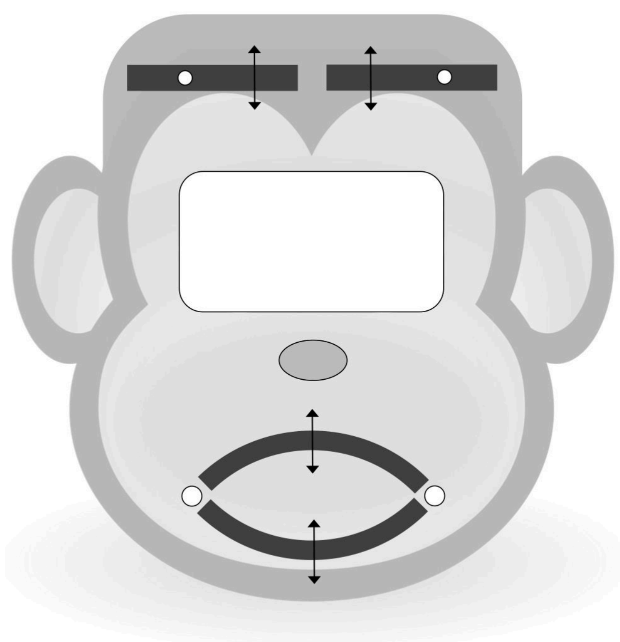

Figure 2. SAM's Facial Degrees of Freedom

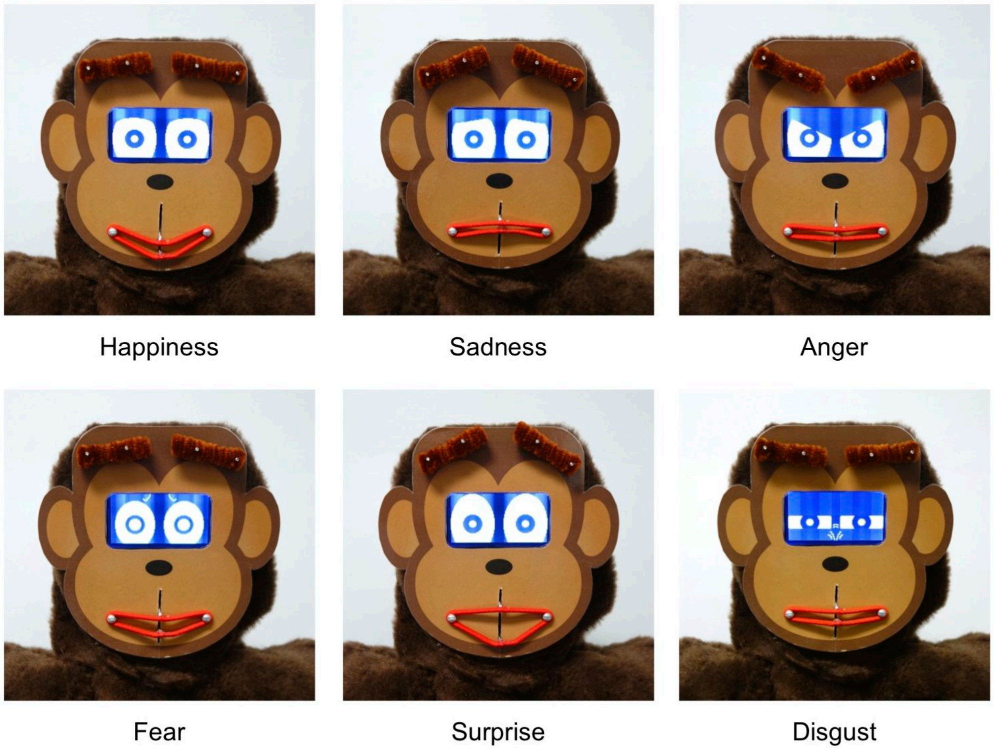

Figure 3. SAM's Emotional Facial Expressions

A recent usability study evaluating SAM's robotic design suggests that it operates as intended. The robotic facial design allows for the display of detailed emotional expressions that are highly identifiable and true to expressions observed in the human face. Typically developing children were able to label SAM's expressions at a rate of 83.3% accuracy with substantial agreement ($\kappa = 0.681$) and match SAM's expressions to human expressions at a rate of 68.1% accuracy with moderate agreement ($\kappa = 0.434$). SAM's robotic design and interaction style also create an environment that is socially engaging and enjoyable. School-age children with ASD engaged in more socially directed gaze (e.g., spent a higher percentage of time looking at the face while engaged) during an interaction with SAM ($M = 28.52\%$) compared to a similar human interaction

($M = 7.69\%$). On 10-point Likert-type rating scales, school-age children with ASD also reported feeling happy ($M = 9.54$) and comfortable ($M = 9.23$), and were eager to have a future interaction with SAM ($M = 9.00$; Koch et al., 2015).

Programming and Operation

SAM operates in automatic mode. The intervention sessions are programmed through Presentation® software (Version 17.2; Neurobehavioral Systems, 2004), a system that provides stimulus delivery and communication with external devices. SAM interacts with human users by displaying different social behaviors following responses made on a touchscreen tablet. Using Presentation's programming language, coded signals are sent from the tablet to SAM via USB. SAM then displays a response, which has been programmed in the Arduino integrative development environment and uploaded to the embedded Arduino Mega 2560 microcontroller. Responses consist of preprogrammed motor movements synchronized with vocal recordings. For example, when the correct answer is chosen on the tablet, SAM responds by exclaiming, "That's right!" while clapping and displaying a happy expression. By asking questions and responding with various preprogrammed social behaviors, SAM is able to maintain the social interaction without requiring outside control by experimenters. This design also allows for a high level of consistency across interactions with different users.

Intervention Program

SAM is programmed to autonomously lead a social skills intervention aimed at improving children's identification and understanding of emotions. The robot-based intervention involves a series of eight sessions comprised of different games designed to introduce varying aspects of emotion knowledge. See Table 1 for session content including presentation and order of games. Length of the intervention protocol and specific timing of individual sessions was modeled after the FaceSay computer-based social skills training program previously developed by the STAR lab (Hopkins et al., 2011). For the FaceSay study, the intervention protocol was limited to six weeks in addition to baseline and post-intervention sessions to allow for completion of the study during one school term. The protocol length was determined to maximize the number of sessions within a single term thus minimizing attrition and lapses in study protocol.

Table 1

Intervention Program Session Content

Session	Content	Estimated Length
Baseline	MODEL (6 trials) MATCH-D (12 trials) MATCH-F (12 trials)	15 min
Session 1	MODEL (6 trials) MATCH-D (24 trials)	15 min
Session 2	MODEL (6 trials) MATCH-F (24 trials)	15 min
Session 3	MODEL (6 trials) CHANGE (12 trials) INFER (12 trials)	25 min
Session 4	MODEL (6 trials) CHANGE (24 trials)	25 min
Session 5	MODEL (6 trials) INFER (24 trials)	25 min

(continued)

Session	Content	Estimated Length
Session 6	MODEL (6 trials) MATCH-D (6 trials) MATCH-F (6 trials) CHANGE (6 trials) INFER (6 trials)	20 min
Post-Intervention	MODEL (6 trials) MATCH-D (12 trials) MATCH-F (12 trials)	15 min

The social skills intervention designed for use with SAM falls within the domain of child-specific interventions (McEvoy, Odom, & McConnell, 1992). Games within the intervention focus on improving understanding of Ekman & Friesen's (1975) six universal emotions: happiness, sadness, anger, fear, surprise, and disgust. During intervention sessions, SAM speaks directly to the child by telling stories, showing pictures on the tablet, asking questions, and giving helpful feedback on performance. The child is able to interact with SAM via responses made on the touchscreen tablet. Techniques such as positive reinforcement, modeling, and social narratives are incorporated within the intervention to promote skill development. Games played during the intervention sessions include: *Modeling Emotional Expressions (MODEL), Expression Matching: Schematic Drawings (MATCH-D), Expression Matching: Face Photos (MATCH-F), Recognizing Change in Expression (CHANGE),* and *Inferring Emotion (INFER).*

Regarding specific content, the first three games focus on basic emotion identification and matching. During *MODEL*, SAM introduces the child to the topic of emotions by modeling different facial expressions and labeling the accompanying emotions. During *MATCH-D*, the child is presented with an emotion that is modeled by SAM and asked to identify the schematic drawing depicting the same emotional facial expression. The basic, black-and-white drawings utilize facial features essential for emotion recognition (MacDonald, Kirkpatrick, and Sullivan, 1996). Similarly, during *MATCH-F*, the child is presented with an emotion that is modeled by SAM and asked to identify the full-face human photo depicting the same emotional facial expression. The model used in these black-and-white photos was trained to display emotions with particular muscles and features emphasized for different facial expressions (Ekman and Friesen, 1975).

The final two games involve a more complex understanding of emotions in others and across different situations. During *CHANGE*, the robot recites a brief social scenario depicting a personal story with emotional content. While reading the story, SAM displays the corresponding emotional facial expression. After a 5-second delay, SAM models an additional emotion, and the child is asked to identify whether SAM's expression has changed. During *INFER*, SAM recites a brief social scenario while maintaining a flat, emotionless facial expression. After finishing the story, the child is asked to identify the emotion corresponding to the story. Social scenarios used in the intervention were developed by the STAR lab and validated within a group of typically developing adults and children, who were able to identify emotional content from the scenarios with 95.4% and 82.7% accuracy, respectively (Koch, Lebersfeld, Clesi, Hopkins, & Biasini, 2015).

OBJECTIVES

The overarching goal of this study was to assess the effectiveness, feasibility, and acceptability of a robot-based intervention program designed to improve social-emotional skills in school-age children with ASD. Though a wide range of previous studies explore the application of robots in social skills interventions for children with ASD, most studies are only exploratory and rely on case examples or small sample sizes of three or four children (Diehl, Schmitt, Villano, & Crowell, 2012). Many studies also have methodological limitations, such as failing to implement a control condition, not performing group analyses, and utilizing subjective and qualitative outcome measures. Researchers in the field have called for more rigorous research efforts involving large-scale longitudinal studies that provide quantitative information about how individuals with ASD interact with social robots (Cho & Ahn, 2016; Diehl, Schmitt, Villano, & Crowell, 2012; Scassellati, Admoni, & Mataric, 2012).

Additionally, the bulk of previous research explores whether the presence of a robot elicits social behaviors during an interaction session. A limited number of studies evaluate the use of robotic interventions to teach and help children develop specific social skills that can generalize to human interactions in the natural environment. Even fewer examine the potential to help children with ASD improve poorly developed social-emotional skills, such as emotion recognition, that are believed to underlie more complex social perception abilities and contribute to global social deficits in ASD (Baron-Cohen, Golan, & Ashwin, 2009). Past research suggests that the robot designed for this study, SAM, may be capable of teaching these skills through a comfortable and engaging interaction environment (Koch et al., 2015). The current study evaluated the efficacy of the robot-based intervention program in a group of children with ASD utilizing a randomized controlled trial experimental design with objective, standardized, and norm-based outcome measures. Specific aims and hypotheses are outlined below.

Specific Aims

Aim 1: To examine whether the robot-based intervention provides an engaging and enjoyable therapeutic environment throughout the course of the intervention program. Within the intervention group, it was predicted that participants would display consistently high levels of engagement with the robot, as measured by time spent viewing relevant stimuli, across baseline and post-intervention sessions (Hypothesis 1.1). It was further hypothesized that participants in the intervention group would endorse consistently high levels of happiness, comfort, and desire for future interactions, as measured by self-report questionnaire ratings, across baseline and post-intervention sessions (Hypothesis 1.2).

Aim 2: To analyze whether the robot-based intervention improves robot-based social-emotional skills in children with ASD immediately following the intervention program. It was expected that the intervention group would show increased accuracy for matching robotic emotional facial expressions compared to the control group following the intervention (Hypothesis 2.1). It was also predicted that the intervention group would display an increased amount of socially directed gaze with the robot, as measured by time spent viewing the robot's face, compared to the control group following the intervention (Hypothesis 2.2).

Aim 3: To analyze whether the robot-based intervention improves generalized social-emotional skills in children with ASD immediately following the intervention program. It was hypothesized that the intervention group would show improved performance on a standardized measure of affect recognition compared to the control group following the intervention (Hypothesis 3.1). It was further predicted that the intervention group would show improved performance on a standardized measure of theory of mind compared to the control group following the intervention (Hypothesis 3.2).

METHODS

Participants

Of the 23 children who were initially enrolled, one withdrew before completing study procedures. Twenty-two children between the ages of 5 and 12 completed the study (M = 8.45 years, SD = 1.90). Of the 22 participants, 13 were male. Fifteen participants were Caucasian (and not of Hispanic origin), five were African American (and not of Hispanic origin), one was Hispanic, and one was of mixed descent. Parental education ranged from high school to advanced degree (M = 16.23 years, SD = 2.43). Participants were randomly assigned to either the intervention (n = 11) or waitlist control group (n = 11). Fisher's exact tests and independent samples t-tests were conducted to examine potential differences across groups for demographic variables. There were no differences in the demographic composition of the two groups for age, $t(20) = 0.896$, $p = .381$, gender, $p > .05$, ethnicity, $p > .05$, or parental education, $t(20) = -1.150$, $p = .264$. Demographic characteristics are reported in Table 2.

Table 2

Demographic and Baseline Characteristics of Study Participants

	Intervention (*n* = 11)	Control (*n* = 11)	Overall (N = 22)
Variable	*M (SD)*	*M (SD)*	*M (SD)*
Age	8.09 (1.70)	8.82 (2.09)	8.45 (1.90)
Parental Education	16.82 (2.32)	15.64 (2.50)	16.23 (2.43)
ADOS-2			
Social Affect	10.09 (2.74)	11.91 (3.59)	11.00 (3.25)
RRB	1.36 (0.92)	2.18 (1.78)	1.77 (1.45)
Overall Total	11.45 (3.14)	14.09 (3.67)	12.77 (3.60)
SCQ Total	18.82 (6.56)	18.45 (7.08)	18.64 (6.66)
KBIT-2			
Verbal IQ	103.55 (11.32)	97.91 (20.29)	100.73 (16.29)
Nonverbal IQ	99.18 (12.99)	101.27 (14.49)	100.23 (13.47)
IQ Composite	101.73 (11.46)	99.73 (17.43)	100.73 (14.43)
	Freq. (%)	Freq. (%)	Freq. (%)
Gender			
Male	7 (63.6)	6 (54.5)	13 (59.1)
Female	4 (36.4)	5 (45.5)	9 (40.9)
Ethnicity			
African American	3 (27.3)	2 (18.2)	5 (22.7)
Caucasian	7 (63.6)	8 (72.7)	15 (68.2)
Hispanic	1 (9.1)	0 (0)	1 (4.5)
Mixed	0 (0)	1 (9.1)	1 (4.5)

Note. RRB = Restricted and Repetitive Behavior.

Participants were recruited from organizations in the Birmingham area that offer resources and specialized services for families of children with ASD (e.g., Autism Society of Alabama, Mitchell's Place). Inclusion criteria included a previous diagnosis of ASD according to the criteria specified by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) by a licensed community professional (American Psychiatric Association, 2013). Diagnostic eligibility was confirmed by experimenter administration of the Autism Diagnostic Observation Schedule – Second Edition (ADOS-2; Lord et al., 1989) and parent completion of the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). Nineteen participants met ADOS-2 criteria for autism, and three for autism spectrum disorder (M = 12.77, SD = 3.60). All but one participant met criteria for likelihood of autism on the SCQ (M = 18.64, SD = 6.66). Based on ADOS-2 scores for this participant and experimenter consensus, the child was deemed eligible and included in study procedures and analyses.

To ensure that cognitive skills were adequate to participate in the study protocol, additional inclusion criteria included mild-to-no cognitive impairment. Cognitive eligibility was confirmed by experimenter administration of the Kaufman Brief Intelligence Test – Second Edition (KBIT-2; Kaufman & Kaufman, 1990). All participants achieved an overall IQ composite score within the range of mild-to-no cognitive impairment (M = 100.73, SD = 14.43). Lastly, participants were not permitted to have uncorrected vision problems, given that it may interfere with their ability to view stimuli and interact with the robot. Independent samples t-tests were conducted to examine potential differences across groups for descriptive measures. There were no baseline differences in the two groups for the ADOS-2 overall total, $t(20) = 1.809$, p =

.086, SCQ total, $t(20) = -0.125$, $p = .902$, or KBIT-2 composite IQ, $t(20) = -0.318$, $p = .754$. Baseline characteristics are reported in Table 2.

Procedure

The protocol for this study was approved by the Institutional Review Board at UAB. Recruitment was completed by posting and distributing flyers to families of prospective participants through local organizations specializing in ASD-related resources and services. Researchers spoke with interested parents to increase awareness of the study and provide answers to any questions about participation. After full explanation of study procedures, parents who wished to enroll gave written informed consent and written child assent was obtained from children ages 7 and older.

Enrolled participants subsequently underwent group assignment. A randomized controlled trial experimental design was utilized to assess the effectiveness of the intervention. Due to the small sample size, a block randomization scheme was used to randomize participants into intervention and waitlist control groups. This method ensured a balance in sample size across groups over time as additional participants enrolled in the study. A block size of four was used to more easily control balance. All possible balanced combinations of assignment within the block were calculated, and blocks were randomly generated to determine each participant's group assignment. All participants, regardless of group placement, continued to receive any preexisting school-based and/or private services during the course of the study.

Study procedures were completed over the course of eight weeks. Participants completed demographic, descriptive, and outcome measures at a baseline assessment prior to beginning the intervention (T1). Children were introduced to SAM and completed the baseline session of the robot-based intervention program at this time. T1 assessments lasted approximately 140 minutes. During the intervention phase, participants assigned to the intervention group received once weekly 15-25 minute sessions of the robot-based intervention for six weeks. Participants assigned to the waitlist control condition did not interact with SAM during this time; these children were given the opportunity to receive the intervention at the end of the study. Following completion of the intervention phase, participants completed outcome measures, including the post-intervention session of the robot-based intervention program, at a post-intervention assessment (T2). T2 assessments lasted approximately 45 minutes. See Table 3 for timing of specific measures.

Table 3

Length and Timing of Measures

Measure	T1	T2
Parent		
Consent Form	10 min	
Demographic Questionnaire	10 min	
SCQ	20 min	
Total Minutes to Complete:	40 min	
Child		
Assent Form	10 min	
KBIT-2	25 min	
ADOS-2	60 min	
NEPSY-II	20 min	20 min
SAM Interaction/faceLAB™ Eye Tracking	15 min	15 min
Enjoyment Questionnaire	10 min	10 min
Total Minutes to Complete:	140 min	45 min

Measures

Demographic and Descriptive Measures

Demographic Questionnaire. Parents completed a demographic questionnaire regarding their child's age, ethnicity, grade, gender, diagnosis, household composition, area of residence, and service history (e.g., past and present social skills training). They also reported on parent characteristics including education level and marital status.

Autism Diagnostic Observation Schedule – Second Edition. The ADOS-2 was used to corroborate previous diagnosis of ASD. The ADOS-2 is a standardized, semi-structured assessment measure that evaluates social interaction, communication, imaginative play skills, and restricted and repetitive behaviors (Lord et al., 1989). It was developed for the assessment of ASD across a wide range of chronological and developmental ages and consists of five different modules, each developed for a different developmental stage and language level. Classifications of autism (total score ≥ 9), autism spectrum (total score = 7-8), or non-spectrum (total score < 7) are based on scores assigned to observed behaviors and skills during the ADOS-2 and are used to assist in determining a diagnosis. Based on the language and developmental level of the sample, Module 3 was completed for all participants.

Social Communication Questionnaire. In concert with the ADOS-2, the SCQ was used to corroborate previous diagnosis of ASD. The SCQ is a brief instrument that evaluates the communication skills and social functioning of an individual who may have ASD (Rutter, Bailey, & Lord, 2003). Designed as a companion for the longer and more thorough Autism Diagnostic Interview – Revised (ADI-R), this 40-item parent-report questionnaire provides a dimensional measure of ASD symptomology with a cutoff score that can be used to indicate the likelihood that an individual has ASD. The SCQ has been shown to be an effective screener for discriminating between children with and without ASD (Chandler et al., 2007). While a cutoff score of 15 is suggested by the test developers, it is stated that researchers may wish to use a lower cutoff to increase sensitivity. Based on previous research, a cutoff score of 12 was used for the purpose of this study (Allen, Silove, Williams, & Hutchins, 2007; Corsello et al., 2007).

Kaufman Brief Intelligence Test – Second Edition. The KBIT-2 was used to obtain an estimate of cognitive functioning. The KBIT-2 assesses general cognitive abilities and generates verbal, nonverbal, and composite domain scores as well as verbal and nonverbal age equivalents of cognitive functioning (Kaufman & Kaufman, 1990). The IQ composite score on the KBIT-2, defined as a standard score (M = 100, SD = 10), was used as global measure of cognitive functioning. For the purpose of this study, an IQ Composite score of 80 or higher was indicative of mild-to-no cognitive impairment and an IQ Composite score below 80 was indicative of moderate-to-severe cognitive impairment. The KBIT-2 has strong psychometric characteristics, with an internal consistency coefficient of .92 and a test–retest reliability coefficient of .90. This test has also demonstrated a high correlation with Wechsler intelligence tests (r = .84; Kaufman & Kaufman, 1990).

Outcome Measures

Robot-Based Emotion Skills Knowledge. All participants completed baseline and post-intervention sessions of the robot-based intervention program, regardless of group placement. As a measure of knowledge of emotion skills explicitly taught in the robot-based intervention, participants were asked to identify the emotions displayed in schematic face drawings (MacDonald, Kirkpatrick, and Sullivan, 1996) and photos of human faces (Ekman and Friesen, 1975) that corresponded with SAM's emotional facial expressions. During the Robot-Based Emotion Skills Knowledge (R-BESK) task, six possible answer choices depicting different emotions (e.g., via drawings or photos) appeared on a touchscreen tablet, allowing the child to touch the desired image to indicate his or her response. Accuracy for matching emotions was recorded via responses made on the touchscreen tablet, with a maximum total correct score of 24. This measure was used to assess whether the child paid attention, learned, and remembered the skills taught during the intervention. Emotional expressions analyzed included happiness, sadness, anger, fear, surprise, and disgust. This measure has strong test-retest reliability (r = .79) with a slight upward shift at retest, based on the control group in this sample.

FaceLAB™ Eye-Tracking Technology. During baseline and post-intervention sessions of the robot-based intervention, participants' eye gaze patterns were monitored and recorded by the faceLAB™ eye-tracking system (Version 5; Seeing Machines, 2009). Eye gaze is often described as an important nonverbal component of social engagement, and viewing patterns were analyzed as an objective measure of engagement during the interaction. During the session, eye-tracking cameras sat non-intrusively on a table between the child and the robot. Cameras tracked the child's eye movements by measuring infrared pupillary and corneal reflections. Coordinates of relevant stimuli in the environment (e.g., SAM's face, SAM's body, touchscreen tablet), deemed "regions of interest" (ROIs), were defined within the faceLAB™ system. Estimates of overall engagement based on time spent viewing ROIs were generated. Estimates of socially directed gaze based on time spent viewing SAM's face were also calculated. Many studies assessing engagement in robotic research utilize recordings of participant eye gaze (Castellano, Pereira, Leite, Paiva, & McOwan, 2009; Michalowski, Sabanovic, & Simmons, 2006; Peters, Asteriadis, Karpouzis, & de Sevin, 2008) and past research has found consistency between eye gaze patterns and subjective ratings of engagement made by observers (Lahiri, Warren, & Sarkar, 2011).

Developmental Neuropsychological Assessment – Second Edition. The Developmental Neuropsychological Assessment – Second Edition (NEPSY-II) is a standardized neuropsychological battery of measures designed to assess a variety of neurocognitive processes (Korkman, Kirk & Kemp, 2007). Subtests within the Social Perception domain of the NEPSY-II, including Affect Recognition (AR) and Theory of Mind (TOM), were used as measures of generalized social-emotional processing skills. Subtest scores are defined as scaled scores (M = 10, SD = 3). The NEPSY-II AR subtest assesses emotion perception by requiring the child to identify and match different emotional facial expressions. The NEPSY-II TOM subtest is comprised of Verbal and Contextual TOM tasks. The Verbal task assesses understanding of others' thoughts and feelings through stories and pictures, and the Contextual task assesses recognition of emotions in contextual situations. NEPSY-II reliability coefficients range from .83-.93 for the Social Perception subtests based on a special groups sample. NEPSY-II test-retest reliabilities show generally adequate stability across time with coefficients ranging from .46-.66 for AR and .76-.84 for TOM. The NEPSY-II Social Perception subtests are also sufficiently predictive of performance on other social-emotional measures, such as the

Self/Social Awareness subtest of the Bracken Basic Concept Scale (r = .25-.32; Korkman, Kirk & Kemp, 2007).

Enjoyment Questionnaire. To assess overall response to SAM and the robot-based intervention, participants completed a brief enjoyment questionnaire. The enjoyment questionnaire uses rating scales to assess three domains: how happy the child felt while working with SAM, how comfortable the child felt during the interaction, and how much the child would like to interact with SAM on another occasion. Self-reported levels of enjoyment were measured using 10-point Likert-type scale items, where higher ratings indicated increased enjoyment, comfort, and desire for future interactions. For the purpose of this study, ratings between 8-10 were indicative of high favorability, ratings between 4-7 were indicative of moderate favorability, and ratings between 0-3 were indicative of low favorability across the three domains. To promote understanding of the task, a thermometer was used to illustrate the 10-point rating scale for each domain.

RESULTS

A Priori Analyses

Sample Size Considerations

A priori power analysis using G*Power was performed to determine an appropriate sample size for the study (Faul, Erdfelder, Lang, & Buchner, 2007). The estimated effect size for this study was garnered from previous research in the STAR lab conducted by Hopkins et al. (2011) assessing a computer-administered avatar-based social skills intervention in a group of school-age children with ASD. Based on this work, the effect size was expected to be large (d = 1.01). Power analysis was conducted for a repeated measures, between factors analysis of variance (ANOVA) using the determined effect size with alpha set at .05. Power analysis indicated that including 22 participants in this study should be sufficient to maintain power of .80. As this study proposed two groups (intervention v. control), the obtained N was divided by two to include 11 individuals in each group.

Engagement and Enjoyment

Hypothesis 1.1

Overall engagement with SAM was measured based on faceLAB^{TM} eye-tracking data for percentage of time spent viewing ROIs during baseline and post-intervention sessions for participants in the intervention group. A paired-samples equivalence test (Weber & Popova, 2012) was conducted to assess whether there was statistically significant equivalence in engagement from baseline to post-intervention, with percentage of time spent viewing ROIs as the dependent variable. Assumptions of this test were met. There were no outliers in the data and the assumption of normality was not violated, as assessed by Shapiro-Wilk's test (p = .182). Analyses revealed that participants in the intervention group did not spend significantly equivalent amounts of time viewing ROIs across T1 and T2, t(10) = 3.23, p = .815, δ = .50 (Table 4). Rather, participants spent significantly less percentage of intervention time engaged at T2 (M = 38.94%, SD = 12.72) than at T1 (M = 58.37%, SD = 16.48), representing an overall mean decrease of 19.43% engagement.

Table 4

Engagement and Enjoyment Ratings for the Intervention Group across T1 and T2

	T1		T2				95% CI	
Variable	*M*	*SD*	*M*	*SD*	*t*(10)	*p*	*LL*	*UL*
Overall % engagement	58.37	16.48	38.94	12.72	3.23	.815	6.02	32.84
Happiness	8.64	2.06	8.64	2.06	0.00	.020*	-0.60	0.60
Comfort	7.82	2.60	9.55	1.51	-2.30	.567	-3.40	-0.05
Desire for future interactions	9.45	0.31	9.18	0.48	0.54	.065	-0.86	1.40

Note. n = 11. δ = .50. CI = confidence interval; *LL* = lower limit; *UL* = upper limit. *p < .05.

Hypothesis 1.2

Enjoyment while interacting with SAM was measured by self-reported ratings for happiness, comfort, and desire for future interactions on the Enjoyment Questionnaire (maximum rating = 10) for participants in the intervention group. A series of paired-samples equivalence tests (Weber & Popova, 2012) were conducted to assess for statistically significant equivalence in enjoyment from baseline to post-intervention, with ratings for happiness, comfort, and desire for future interactions as the dependent variables.

Regarding happiness ratings, two outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of their values did not reveal them to be extreme and they were kept in the analysis. Shapiro-Wilk's test revealed the change scores from T1 to T2 were not normally distributed ($p < .001$); however, the analysis was continued as planned given adequate skewness (0.000 ± 0.661) and the robustness of the t-test to violations of this kind (Boneau, 1960). Analyses revealed that participants in the intervention group rated statistically equivalent feelings of happiness from T1 ($M = 8.64$, $SD = 2.06$) to T2 ($M = 8.64$, $SD = 2.06$), $t(10) = 0.00$, $p < .05$, $\delta = .50$ (Table 4). Ratings were indicative of consistently high favorability across time.

For comfort ratings, there were no outliers in the data, as assessed by inspection of a boxplot. Shapiro-Wilk's test revealed the change scores from T1 to T2 were not normally distributed ($p < .001$); however, the analysis was continued as planned given adequate skewness (0.899 ± 0.661) and the robustness of the t-test to violations of this kind (Boneau, 1960). Analyses revealed that participants in the intervention group did not rate significantly equivalent feelings of comfort across T1 and T2, $t(10) = -2.30$, $p = .567$,

$\delta = .50$ (Table 4). Rather, participants reported feeling more comfortable at T2 ($M = 9.55$, $SD = 1.51$) than at T1 ($M = 7.82$, $SD = 2.60$), representing an overall mean increase of 1.73 points. Ratings increased from moderate to high favorability across time.

Regarding ratings of desire for future interactions, three outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. Two outliers were not considered extreme and were kept in the analysis. One extreme outlier was identified; the outlier's value was modified to just larger than the next largest value to help remove its negative effect. Shapiro-Wilk's test revealed the change scores from T1 to T2 were not normally distributed ($p < .05$); however, the analysis was continued as planned given adequate skewness (-0.561 ± 0.661) and the robustness of the t-test to violations of this kind (Boneau, 1960). Analyses revealed that participants in the intervention group did not rate statistically equivalent desire for future interactions across T1 and T2, $t(10) = 0.54$, $p = .065$, $\delta = .50$ (Table 4). Rather, participants reported slightly less desire for future interactions at T2 ($M = 9.18$, $SD = 0.48$) than at T1 ($M = 9.45$, $SD = 0.31$), representing an overall mean decrease of 0.27 points. Ratings were, however, indicative of consistently high favorability across time. See Figure 4 for summary of enjoyment ratings.

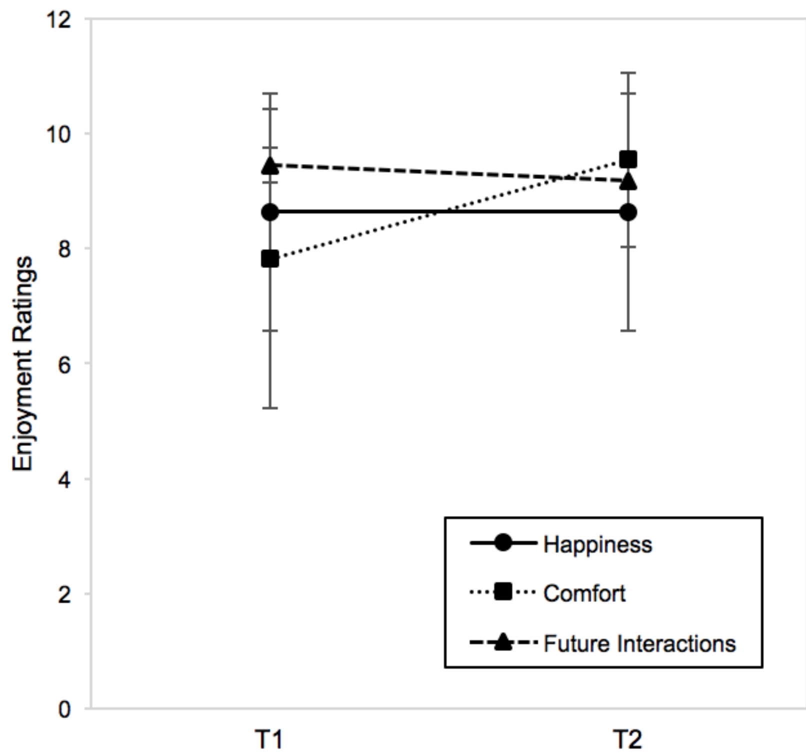

Figure 4. Enjoyment Ratings for the Intervention Group across T1 and T2

Robot-Based Social-Emotional Skills

Hypothesis 2.1

Accuracy for matching robotic emotional facial expressions was measured using raw scores on the R-BESK (maximum score = 24) during baseline and post-intervention sessions with SAM. A two-way mixed ANOVA was conducted with group placement (intervention v. control) as the between-groups factor and time (baseline v. post-intervention) as the within-groups factor, and with R-BESK accuracy scores as the dependent variable. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The data were normally distributed, as assessed by Normal Q-Q Plot. Levene's test revealed a minor violation of the assumption of homogeneity of variances at T1 (p = .027); however, the analysis was continued as planned given the robustness of the ANOVA to violations of this kind, particularly when samples sizes are equal (Sawilowsky & Blair, 1992). Notably, when the analysis was run following a reflect and logarithmic transformation, the assumption of homogeneity of variances was met and results were consistent with those performed with non-transformed data. As such, analyses were continued with non-transformed data. There was homogeneity of covariances (p = .015), as assessed by Box's M test.

Analyses revealed a statistically significant interaction between the intervention and time on R-BESK scores, $F(1, 20) = 5.511$, $p < .05$, partial $\eta^2 = .216$ (Table 5). Follow-up analyses for simple main effects indicated a statistically significant difference in R-BESK scores between groups at T2, $F(1, 20) = 5.818$, $p < .05$, partial $\eta^2 = .225$. Additionally, there was a statistically significant effect of time on R-BESK scores for the intervention group, $F(1, 10) = 26.738$, $p < .001$, partial $\eta^2 = .728$. While R-BESK scores at T1 did not differ between the intervention group (M = 19.55, SD = 2.16) and control group (M = 18.73, SD = 4.76), the intervention group displayed significantly improved performance at T2 (M = 22.91, SD = 1.51) as compared to the control group (M = 19.45, SD = 4.50). See Figure 5 for summary of changes in R-BESK scores across groups over time.

Table 5

Group Comparisons for Social-Emotional Measures across T1 and T2

	Intervention (n = 11)	Control (n = 11)	Time x Group Interaction		Time Effect		Group Effect	
Variable	*M (SD)*	*M (SD)*	F(1, 20)	*p*	F(1, 20)	*p*	F(1, 20)	*p*
R-BESK			5.511	.029*				
T1	19.55 (2.16)	18.73 (4.76)					.270	.609
T2	22.91 (1.51)	19.45 (4.50)					5.818	.026*
Intervention					26.738[a]	.000**		
Control					0.631[a]	.445		
Socially Directed Gaze			0.543	.470	0.033	.859	0.269	.610
T1	30.33 (17.45)	30.67 (12.56)						
T2	27.17 (12.89)	32.59 (17.65)						
NEPSY-II AR			1.729	.203	0.098	.757	0.001	.971
T1	9.73 (2.45)	8.73 (3.20)						
T2	9.00 (3.82)	9.91 (3.73)						
NEPSY-II TOM			0.178	.677	2.575	.124	0.030	.864
T1	7.55 (2.88)	7.55 (3.39)						
T2	8.64 (3.56)	8.18 (3.40)						

Note. [a]F(1, 10). *p < .05. **p < .001.

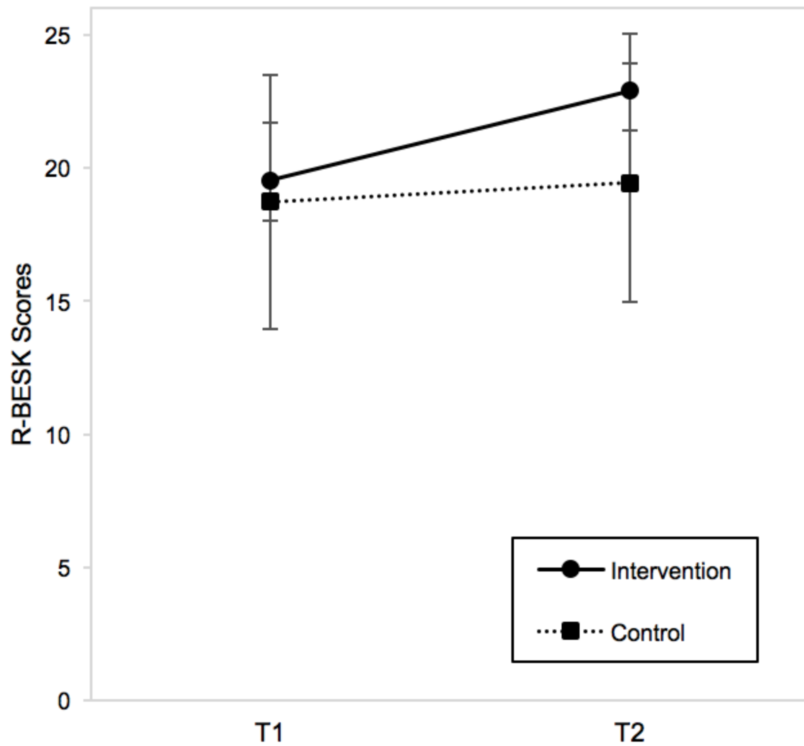

Figure 5. Changes in R-BESK Scores across Groups from T1 to T2

Hypothesis 2.2

Socially directed gaze was measured based on faceLAB^{TM} eye-tracking data for percentage of engagement time spent viewing SAM's face during baseline and post-intervention sessions. A two-way mixed ANOVA was conducted with group placement (intervention v. control) as the between-groups factor and time (baseline v. post-intervention) as the within-groups factor, and with percentage of engagement time spent viewing SAM's face as the dependent variable. Assumptions of this test were met. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The data were normally distributed, as assessed by Normal Q-Q Plot. There was homogeneity of variances ($p > .05$) and covariances ($p = .488$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

Analyses revealed there was no statistically significant interaction between the intervention and time on socially directed gaze, $F(1, 20) = 0.543$, $p = .470$, partial $\eta^2 = .026$ (Table 5). The intervention group, on average, spent 30.33% (*SD* = 17.45) of engagement time looking at SAM’s face at T1 and 27.17% (*SD* = 12.89) at T2, while the control group spent 30.67% (*SD* = 12.56) of engagement time at T1 and 32.59% (*SD* = 17.65) at T2. The main effect of time showed no difference in socially directed gaze across T1 and T2, $F(1, 20) = 0.033$, $p = .859$, partial $\eta^2 = .002$. Similarly, the main effect of group indicated no difference in socially directed gaze between the groups, $F(1, 20) = 0.269$, $p = .610$, partial $\eta^2 = .013$.

Generalized Social-Emotional Skills

Hypothesis 3.1

Affect recognition skills were measured by performance on the NEPSY-II AR subtest. A two-way mixed ANOVA was conducted with group placement (intervention v. control) as the between-groups factor and time (baseline v. post-intervention) as the within-groups factor, and with NEPSY-II AR scaled scores as the dependent variable. Assumptions of this test were met. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The data were normally distributed, as assessed by Normal Q-Q Plot. There was homogeneity of variances ($p > .05$) and covariances ($p = .573$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

Analyses revealed there was no statistically significant interaction between the intervention and time on NEPSY-II AR scores, $F(1, 20) = 1.729$, $p = .203$, partial $\eta^2 = .080$ (Table 5). The intervention group performed within the average range across T1 and T2 with mean scaled scores of 9.73 ($SD = 2.45$) and 9.00 ($SD = 3.82$), respectively, as did the control group with scores of 8.73 ($SD = 3.20$) and 9.91 ($SD = 3.73$), respectively. The main effect of time showed no difference in NEPSY-II AR scores across T1 and T2, $F(1, 20) = 0.098$, $p = .757$, partial $\eta^2 = .005$. Similarly, the main effect of group indicated no difference in NEPSY-II AR scores between the groups, $F(1, 20) = 0.001$, $p = .971$, partial $\eta^2 = .000$.

Hypothesis 3.2

Theory of mind knowledge was measured by performance on the NEPSY-II TOM subtest. A two-way mixed ANOVA was conducted with group placement (intervention v. control) as the between-groups factor and time (baseline v. post-intervention) as the within-groups factor, and with NEPSY-II TOM scaled scores as the dependent variable. Assumptions of this test were met. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The data were normally distributed, as assessed by Normal Q-Q Plot. There was homogeneity of variances ($p > .05$) and covariances ($p = .305$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively.

Analyses revealed there was no statistically significant interaction between the intervention and time on NEPSY-II TOM scores, $F(1, 20) = 0.178$, $p = .677$, partial $\eta^2 = .009$ (Table 5). The intervention group performed within the average range across T1 and

T2 with mean scaled scores of 7.55 (*SD* = 2.88) and 8.64 (*SD* = 3.56), respectively, as did the control group with scores of 7.55 (*SD* = 3.39) and 8.18 (*SD* = 3.40), respectively. The main effect of time showed no difference in NEPSY-II TOM scores across T1 and T2, $F(1, 20) = 2.575$, $p = .124$, partial $\eta^2 = .114$. Similarly, the main effect of group indicated no difference in NEPSY-II TOM scores between the groups, $F(1, 20) = 0.030$, $p = .864$, partial $\eta^2 = .002$.

DISCUSSION

This study examined the effectiveness, feasibility, and acceptability of a robot-based intervention program designed to improve social-emotional skills in school-age children with ASD. The first set of analyses focused on whether participants with ASD who completed the intervention found the robot-based intervention to be consistently engaging and enjoyable throughout the course of the program. Contrary to initial hypothesis, results indicated that participants in the intervention group displayed a decrease in overall percent engagement from baseline to post-intervention. While these children spent approximately 60% (*M* = 58.37, *SD* = 16.48) of the initial intervention session visually attending to SAM and the touchscreen tablet, they spent only 40% (*M* = 38.94, *SD* = 12.72) of the session engaged at post-intervention.

This drop in overall engagement can likely be explained in many ways. At the baseline assessment, children were introduced to SAM for the first time, and were likely intrigued by the robot and the novelty of the situation. It is possible that the children in the intervention group habituated to SAM and the interaction environment over the course of the 8-week study, and therefore spent less time visually exploring the robot at post-intervention. This drop in engagement may also be explained by skill improvement over time. It is possible that children in the intervention group, having spent multiple weeks engaged in the training protocol learning specific emotion skills, simply required less time looking at SAM and the tablet to identify and select correct responses during the post-intervention session, thereby decreasing their overall engagement based on the eye-tracking measure.

While participants in the intervention group did not remain consistently engaged from baseline to post-intervention, results indicated consistently high levels of happiness throughout the intervention protocol. These children rated their feelings of happiness during interactions with SAM as highly favorable across baseline ($M = 8.64$, $SD = 2.06$) and post intervention ($M = 8.64$, $SD = 2.06$). Their desire for additional interactions decreased only slightly from baseline ($M = 9.45$, $SD = 0.31$) to post-intervention ($M = 9.18$, $SD = 0.48$), and were consistently in the highly favorable range. Contrary to hypothesis, participants in the intervention group reported feeling more comfortable with SAM at post-intervention ($M = 9.55$, $SD = 1.51$) than they did at baseline ($M = 7.82$, $SD = 2.60$), representing an increase from moderate to high favorability.

Overall, results of self-reported happiness ratings on the Enjoyment Questionnaire were indicative of high acceptability of SAM and the robot-based intervention protocol. Children in the intervention group continued to report high feelings of happiness after participating in weekly sessions with SAM, and noted high desirability for additional interactions following completion of the study protocol. Children in the intervention group also grew more comfortable working with SAM after several weeks, likely due to repeated exposure to SAM and the structure of the intervention sessions, which fits with the notion that children with ASD feel more comfortable in simplified, predictable, and reliable environments where the complexity of interaction can be controlled (Moore, McGrath, & Thorpe, 2000; Robins, Dickerson, Stribling, & Dautenhahn, 2004). Taken together, these findings further support the notion that decrease in overall engagement was likely not due to boredom or rejection of the intervention program, but rather habituation and possibly more efficient responding due to expertise in emotion skill knowledge.

The second set of analyses explored whether the robot-based intervention improved robot-based social-emotional skills in children with ASD. Results indicated significant improvement in accuracy for identifying face drawings and photos that corresponded with robotic emotional facial expressions on the R-BESK at post-intervention for individuals in the intervention group. Participants who completed eight sessions with SAM improved from correctly matching an average of 20/24 emotions (M = 19.55, SD = 2.16) at baseline to 23/24 emotions (M = 22.91, SD = 1.51) at post-intervention. Participants in the control group matched an average of 19/24 emotions across baseline (M = 18.73, SD = 4.76) and post-intervention (M = 19.45, SD = 4.50), therefore not demonstrating significant skill improvement.

Contrastingly, individuals in the intervention group did not display increased socially directed gaze with SAM compared to the control group at post-intervention. Rather, the intervention group spent approximately 30% of engagement time looking at SAM's face across baseline (M = 30.33, SD = 17.45) and post-intervention (M = 27.17, SD = 12.89), as did the control group across baseline (M = 30.67, SD = 12.56) and post-intervention (M = 32.59, SD = 17.65). This indicates that participants in the intervention group, despite spending several weeks with SAM reviewing emotional facial expressions, did not increase their looking to SAM's face when matching facial expressions at post-intervention. Notably, socially directed gaze is defined in this study as a function of overall engagement, such that data represent percent of engagement time spent looking at

SAM's face. As such, while participants in the intervention group spent similar percentages of time looking at SAM's face across baseline and post-intervention when engaged in the session, decreased overall engagement therefore indicates decreased total amount of socially directed gaze at post-intervention.

Findings regarding robot-based social-emotional skills have significant implications for the use of SAM as a tool to teach specific and identifiable skills. Participants in the intervention group clearly benefited from the content of the intervention program and were able to demonstrate learned skills during post-intervention assessment. When these results are considered along with data indicating decreased overall engagement and therefore decreased socially directed gaze at post-intervention, this suggests that children in the intervention group were able to learn and remember presented information throughout the course of several sessions with SAM, such that they did not necessarily need to examine and compare SAM's face and the images on the touchscreen tablet to deduce correct responses at post-intervention.

The third and final set of analyses examined whether the robot-based intervention improved generalized social-emotional skills in children with ASD. Contrary to hypotheses, results indicated that participants in the intervention group did not show improved performance on standardized measures of affect recognition and theory of mind from the NEPSY-II compared to the control group following the intervention. Rather, participants in both groups demonstrated average performance on the NEPSY-II AR across baseline and post-intervention with mean scaled scores of 9.73 (SD = 2.45) and 9.00 (SD = 3.82) in the intervention group, respectively, and scores of 8.73 (SD = 3.20) and 9.91 (SD = 3.73) in the control group, respectively. Similarly, participants in both groups displayed average performance on the NEPSY-II TOM across baseline and post-intervention with mean scaled scores of 7.55 (*SD* = 2.88) and 8.64 (*SD* = 3.56) in the intervention group, respectively, and scores of 7.55 (*SD* = 3.39) and 8.18 (*SD* = 3.40) in the control group, respectively.

These findings may result from multiple factors. It is possible that the specific robot-based emotion skills learned during the intervention simply did not generalize when assessing others' thoughts and emotional expressions in external pictures and stories. Another potential explanation includes insufficient sensitivity of the NEPSY-II to reflect the minor improvement in emotion recognition skills at post-intervention, particularly given the small sample size and associated decreased power. Notably, it is important to consider the overall strong performance of participants at the baseline assessment on both subtests of the NEPSY-II. Participants in both groups achieved baseline scores within the average range on AR and TOM, indicating performance consistent with what would be expected compared to typically developing same-age peers. In this regard, the data may have reached a ceiling effect, allowing minimal room for improvement at post-intervention. This has important implications for interpretation. While the robot-based intervention did not improve skill generalization to a level above what would be expected based on group norms, this does not necessarily indicate that learned skills cannot generalize to external tasks up to the average range.

Summary of Findings

Overall, results of the current study indicate high acceptability of SAM and the robot-based intervention within a sample of school-age children with ASD and mild-to-no cognitive impairment. This is evidenced by high favorability ratings at post-intervention on an Enjoyment Questionnaire, including consistently high ratings of happiness, increased comfort ratings, and only slightly decreased ratings of desire for future interactions across time. Notably, these findings are particularly important for promoting successful outcomes, as social skills are developed over the course of several sessions and disinterest and attrition can pose a problem in intervention programs. Results also indicate that the robot-based intervention is successful in promoting knowledge and understanding of robot-based emotion recognition skills explicitly taught in the intervention protocol. As such, SAM's ability to create an enjoyable interaction environment conducive to learning and skill development clearly makes it a useful tool for intervention within this population.

Notably, while children with ASD who received the intervention displayed improved knowledge of skills explicitly reviewed in the program, the function of their learning remains somewhat unclear. The finding that participants in the intervention group spent significantly less time visually attending to SAM's face and the touchscreen tablet at post-intervention, yet demonstrated increased accuracy on the R-BESK, suggests a shift from effortful processing and active in-session problem solving to more automatic responding as a result of their training. It is also possible that skill improvement was related to direct learning of presented facts and not necessarily development of generalized emotion decoding and understanding. In other words, while the children learned to accurately identify the appropriate emotion images as a function of the intervention, it is unclear whether they truly learned the appropriate steps to recognizing an individual's emotion based on facial expression.

Results indicating lack of improved scores for the intervention group on the NEPSY-II AR and TOM subtests also query the function of learning following the intervention. While the intervention group improved in their ability to identify emotion images within the robot-based intervention, they did not demonstrate improved applied knowledge in identifying and matching emotions when presented through pictures, illustrations, and stories on the NEPSY-II. Notably, these results may be partially due to a ceiling effect by which average scores at baseline left minimal room for improvement at post-intervention. This indicates that the robot-based intervention did not improve skill generalization to a level above what would be expected based on group norms, not necessarily that skill generalization could not be improved from below average to the average range. Taken together, these data raise a question regarding the appropriateness of this particular intervention for the population included in this study, which will be further discussed in the limitations.

Strengths and Limitations

The current study had a number of strengths. This study implemented a large sample size compared to previous robotic research for individuals with ASD, which has relied largely on case examples or small groups of participants with three or four children. It also used standardized, norm-based outcome measures as estimates of social-emotional skills. This study was one of the first to utilize a randomized controlled trial experimental design with an intervention and control group, allowing for group analyses regarding skill improvement over time. The robot-based intervention program also represents a strength of the study. Given the scripted stimulus-response style of the intervention, this allowed for a high level of consistency across different users, such that all participants who received the intervention were exposed to identical content. Additionally, SAM was designed using low-cost materials with a great deal of readily accessible code in the online electronics community, which was in line with the overarching principle of creating an easily reproducible robot for the ASD community.

The study also had limitations. One such limitation includes the homogeneity of the sample in regard to cognitive functioning. Given average NEPSY-II scores and fairly high R-BESK scores across both groups at baseline, the level of content presented in the robot-based intervention may not have been appropriately challenging for the sample included in this study. While the intervention focused on core emotion recognition skills for the six universally identified emotions (Ekman & Friesen, 1975), participants with ASD and mild-to-no cognitive impairment displayed an adequate understanding of basic emotion concepts prior to intervention, as evidenced by baseline scores on social-emotional tasks. As such, the content included in the protocol may have been unfit for their level of cognitive and social-emotional functioning. It is likely that the participants enrolled in this study would have benefitted more from an intervention focused on the learning and application of more advanced emotion concepts. Similarly, it is possible that children with ASD and below average cognitive functioning may have more room for growth in terms of emotional development, and may benefit more significantly from the direct teaching of basic skills presented through the robot-based intervention.

Additionally, the current study did not control for exposure to the robot across groups throughout the course of the study. While participants in the intervention group had once weekly 15-25 minute sessions with SAM during the 6-week intervention phase, participants in the control group did not have any interaction with SAM during this time. Notably, the implementation of a waitlist control group is commonly used in randomized controlled trial designs. However, it is not clear the extent to which simple exposure to SAM and the program structure led to improvement in robot-based emotion recognition skills as compared to the actual content of the intervention sessions.

Another limitation involves the present definition of engagement. Notably, while eye gaze is considered an important nonverbal component of social engagement, visual attention is not the sole indicator of overall engagement. As such, the current findings do not account for other factors such as auditory attention, verbal output, and active thinking that may or may not have remained consistent within the intervention group from baseline to post-intervention. Thus, while we understand that participants in the intervention group spent less time visually attending to ROIs at post-intervention, we cannot firmly conclude that they did not spend that time engaged in some other manner (e.g., talking with the robot, mentally working through the presented problems).

Additionally, this study incorporated only one measure of generalized social-emotional skills. While the NEPSY-II provides a standardized estimate of affect recognition and theory of mind skills based on understanding of concepts not directly presented within the intervention program, it still represents a laboratory-administered measure that may not necessarily reflect the application of skills in the natural environment. Notably, while one benefit of technology-based intervention is the ability to control the complexity of the interaction for learning purposes, true human-to-human social interactions are innately complex in nature. To promote meaningful skill improvement, learned social skills from the intervention program must apply in real-life social situations that are naturally unstructured and chaotic. The collection of additional data regarding participants' daily social functioning, through naturalistic observation as well as parent- and teacher-reported rating scales, would provide valuable information for assessing true generalizability of skills.

Finally, another notable limitation includes the overall sample size included in the study. As noted, while the current study enrolled a much larger group of children with ASD than other robotic studies that rely on case examples or small groups of participants, the sample remains small compared to the standards set by broadbased psychological research. An assessment of the robot-based intervention utilizing a larger sample size, with an increased yet equal number of participants enrolled within both intervention and control groups, would increase power in the completed analyses and likely clarify and strengthen results.

Future Directions

The current study was the first to examine the effectiveness and acceptability of the SAM robot within a population of children with ASD, and was conducted not only to provide a critical assessment of this prototypal version, but to uncover avenues for continued research and further development of the robot and intervention protocol. Future work within the STAR lab will continue to focus on the practicality, efficacy, and ultimate benefit of this tool. Immediate plans include exploring the effectiveness of the robot-based intervention within a group of individuals with ASD and below average cognitive skills. Notably, this population, which arguably has the most to benefit from interventions of this nature, is frequently left out of therapeutic studies due to the difficulty of recruitment and assessment. However, this should not deter investigators from conducting research with this population. Future research will also utilize a variety of outcome measures to examine whether learned social skills generalize to settings outside the laboratory in the absence of SAM. Measures will include parent- and teacher-reported rating scales of social skills as well as direct observation of participant behavior in the natural environment using pre-established behavior-coding schemes.

The STAR lab will also continue to further examine the design of SAM and explore additional areas of focus for intervention. Future versions of SAM will aim to increase the number of emotions discussed and improve upon the display of complex emotions by modifying and emphasizing salient facial features. For example, the research team has discussed adding an additional axis to each eyebrow and lip to allow for increased flexibility of expression. SAM's appearance may also be manipulated to meet specific goals, such as dressing SAM in human-like clothing and using a human-like facial mask to explore potential differences in engagement and acceptability, as well as generalization of skills to naturalistic human-to-human interactions. Additionally, the content of the intervention will be expanded to include previously undiscussed emotions and to improve variability in presented photos, drawings, and social scenarios. Other areas for intervention content will also be explored. For example, for children with ASD and mild-to-no cognitive impairment, the STAR lab will consider ways to improve the intervention to foster increased emotion decoding and understanding. For children with ASD and below average cognitive skills, interventions focused on other areas for social skills improvement, such as eye contact and joint attention, will also be explored.

LIST OF REFERENCES

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders*, 37(7), 1272-1278.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders DSM-5* (5th ed.). Arlington, VA: American Psychiatric Association.

Anderson, D. K., Liang, J. W., & Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(5), 485–494.

Ashwin, C., Chapman, E., Colle, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience, 1,* 349– 363.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg & D. Cohen (Eds.) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (2nd edn) (pp. 3–20). Oxford, UK: Oxford University Press.

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philosophical Transactions of the Royal Society B: Biological Sciences, 364,* 3567–3574.

Baron-Cohen, S., Golan, O., Chapman, E., & Granader, Y. (2007). Transported to a world of emotion. *The Psychologist, 20*(2), 76-77.

Bauminger, N., & Kasari, C. (2000). Loneliness and friendship in high-functioning children with autism. *Child Development*, *71*(2), 447–456.

Bellack, A. S., & Hersen, M. (1979) *Research and Practice in Social Skills Training.* New York: Plenum Press.

Bellini, S. (2004). Social skill deficits and anxiety in high-functioning adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *19*(2), 78–86.

Bellini, S., Peters, J. K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. *Remedial and Special Education, 28*(3), 153-162.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin, 57*(1), 49.

Buitelaar, J. K., van Engeland, H., de Kogel, K. H., de Vries, H., & van Hooff, J. A. (1991). Differences in the structure of social behaviour of autistic children and non-autistic retarded controls. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *32*(6), 995–1015.

Bukowski, W. M., Newcomb, A. F., & Hartup, W. W. (Eds.). (1998). *The Company They Keep: Friendships in Childhood and Adolescence*. Cambridge University Press.

Castellano, G., Pereira, A., Leite, I., Paiva, A., & McOwan, P. W. (2009, November). Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, 119-126.

Centers for Disease Control and Prevention. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summary, 65*(3), 1-23.

Chamberlain, B. O. (2001). *Isolation or involvement? The social networks of children with autism included in regular classes.* (Unpublished doctoral dissertation). University of California, Los Angeles.

Chandler, S., Charman, T., Baird, G., Simonoff, E., Loucas, T., Meldrum, D., Scott, M., & Pickles, A. (2007). Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(10), 1324-1332.

Cho, S. J., & Ahn, D. H. (2016). Socially assistive robotics in autism spectrum disorder. *Hanyang Medical Reviews*, *36*(1), 17-26.

Church, C., Alisanski, S., & Amanullah, S. (2000). The social, behavioral, and academic experiences of children with asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, *15*(1), 12–20.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932-940.

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition, 12*, 1-35.

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders, 6*(1), 249-262.

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots, 24*(2), 147–157.

Eaves, L. C., & Ho, H. H. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 739–747.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues.* Englewood Cliffs, NJ: Prentice-Hall.

Esbensen, A. J., Seltzer, M. M., Lam, K. S. L., & Bodfish, J. W. (2009). Age-related differences in restricted repetitive behaviors in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(1), 57–66.

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feil-Seifer, D., & Matarić, M. J. (2009). Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In *Experimental Robotics* (pp. 201-210). Springer Berlin Heidelberg.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, 27,* 321–342.

Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Wimsatt, F. C., & Biasini, F. J. (2011). Avatar assistant: Improving social skills in students with an ASD through a computer-based intervention. *Journal of Autism and Developmental Disorders, 41*(11), 1543-55.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Wadsworth.

Howlin, P., & Goode, S. (1998). Outcome in adult life for people with autism, asperger syndrome. In F. R. Volkmar (Ed.), *Autism and Pervasive Developmental Disorders* (pp. 209 – 241). New York: Cambridge University Press.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(2), 212–229.

Huskens, B., Verschuur, R., Gillesen, J., Didden, R., & Barakova, E. (2013). Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. *Developmental Neurorehabilitation*, *16*(5), 345-356.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman brief intelligence test*. Circle Pines, MN: American Guidance Service.

Kennedy, D. P., & Adolphs, R. (2014). Violations of personal space by individuals with autism spectrum disorder. *PLoS ONE*, *9*(8), e103369.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2012). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders, 43*(5), 1038-1049.

Koch, S.A., Clesi, C.D., Lebersfeld, J.B., Stevens, C.E., Parker, A.G., McNew, M.E., Hopkins, M.I., Amthor, F.R., & Biasini, F.J. (2015, May). *Socially animated machine (SAM): A novel design for robotic research in children with autism spectrum disorders.* Poster presented at the International Meeting for Autism Research, Salt Lake City, UT.

Koch, S.A., Lebersfeld, J.B., Clesi, C.D., Hopkins, M.I., & Biasini, F.J. (2015, May). *Social scenarios for affect development in children with autism spectrum disorders: Validation in adult and child populations.* Poster presented at the Association for Psychological Science, New York City, NY.

Koegel, L. K., Koegel, R. L., Harrower, J. K., & Carter, C. M. (1999). Pivotal response intervention I: Overview of approach. *Research and Practice for Persons with Severe Disabilities*, *24*(3), 174-185.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II*. San Antonio, TX: Pearson.

Kozima, H., & Nakagawa, C. (2006). Interactive robots as facilitators of children's social development. In A. Lazinica (Ed.), *Mobile Robots towards New Applications* (p. 784). Germany.

Laerd Statistics (2015). Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Lahiri, U., Warren, Z., & Sarkar, N. (2011). Design of a gaze-sensitive virtual social interactive system for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(4), 443-452.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, *19*(2), 185-212.

MacDonald, P. M., Kirkpatrick, S. W., & Sullivan, L. A. (1996). Schematic drawings of facial expressions for emotion recognition and interpretation by preschool-aged children. *Genetic, Social, and General Psychology Monographs, 122*, 373–388.

Mazurek, M. O. (2014). Loneliness, friendship, and well-being in adults with autism spectrum disorders. *Autism*, *18*(3), 223–232.

Mehrabian, A. (1968). Communication without words. *Psychological Today*, *2*, 53-55.

Michalowski, M. P., Sabanovic, S., & Simmons, R. (2006). A spatial model of engagement for a social robot. In *9th IEEE International Workshop on Advanced Motion Control* (pp. 762-767).

McConnell, S. R. (2002). Interventions to facilitate social interaction for young children with autism: Review of available research and recommendations for educational intervention and future research. *Journal of Autism and Developmental Disorders, 32,* 351–372.

McEvoy, M. A., Odom, S. L., & McConnell, S. R. (1992). Peer social competence intervention for young children with disabilities. In S. L. Odom, S. R. McConnell, & M. A. McEvoy (Eds.), *Social competence of young children with disabilities: Issues and strategies for intervention* (pp. 113–133). Baltimore, MD: Paul Brookes.

Moore, D., McGrath, P. & Thorpe, J. (2000). Computer-aided learning for people with autism - A framework for research and development. *Innovations in Education and Teaching International*, *37*(3), 218–228.

Mori, M. (1970). The uncanny valley. *Energy, 7*, 33–5.

Neurobehavioral Systems (2004). Presentation, Version 17.2 [Computer software]. Retrieved from www.neurobs.com

Paul, R. (2008). Interventions to improve communication in autism. *Child and Adolescent Psychiatric Clinics of North America, 17*(4), 835–856.

Peters, C., Asteriadis, S., Karpouzis, K., & de Sevin, E. (2008, October). Towards a real-time gaze-based shared attention for a virtual agent. In *Workshop on Affective Interaction in Natural Environments (AFFINE), ACM International Conference on Multimodal Interfaces (ICMI'08)*.

Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., & De Rossi, D. (2005). An android for enhancing social skills and emotion recognition in people with autism*. IEEE Trans. on Neural Systems and Rehabilitation Engineering, 13*(4), 507-515.

Pioggia, G., Igliozzi, R., Sica, M. L., Ferro, M., Muratori, F., Ahluwalia, A., & De Rossi, D. (2008). Exploring emotional and imitational android-based interactions in autistic spectrum disorders. *Journal of CyberTherapy & Rehabilitation, 1*(1), 49-61.

Ricks, D. J., & Colton, M. B. (2010, May). Trends and considerations in robot-assisted autism therapy. In *Proceedings of the IEEE International Conference on Robots and Automation* (pp. 4354-4359).

Robins, B., Dautenhahn, K., & Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies, 7*(3), 479-512.

Robins, B., Dautenhahn, K., Te Boekhorst, R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society, Heidelberg, 4*(2), 105-120.

Robins, B., Dickerson, P., Stribling, P., & Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction Studies*, *5*(2), 161-198.

Rutter, M., Bailey, A., & Lord, C. (2003). *SCQ: Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Sawilowsky, S. S. & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. *Psychological Bulletin, 111*(2), 352-360.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2011). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience, 7*(4), 413–422.

Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. *International Journal of Robotics Research, 28*, 552-563.

Scassellati, B., Admoni, H., & Mataric, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering, 14,* 275-294.

Seeing Machines (2009). FaceLAB, Version 5 [Computer software]. Retrieved from www.seeingmachines.com

Shaked, M. & Yirmiya, N. (2003). Understanding Social Difficulties. In M. R. Prior (Ed.), *Learning and Behavior Problems in Asperger Syndrome* (pp. 104-125). Guilford Press.

Stanton, C. M., Kahn Jr., P. H., Severson, R. L., Ruckert, J. H., & Gill, B. T. (2008). Robotic animals might aid in the social development of children with autism. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 271-278). IEEE.

Stevens, C. E., Koch, S. A., Amthor, F. R., Biasini, F. J., & Hopkins, M. I. (2015, May). *S.A.M. (socially animated machine): A custom-built robot for use in child ASD interventions.* Poster session presented at the meeting of the Association for Psychological Science, New York City, NY.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1430), 303-314.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (pp. 335–364). John Wiley & Sons, Inc.

Tantam, D. (2003). The challenge of adolescents and adults with asperger syndrome. *Child Adolescence and Psychiatric Clinics of North America, 12*, 143–163.

Volkmar, F. R., Paul, R., Klin, A., & Cohen, D. J. (2005). *Handbook of Autism and Pervasive Developmental Disorders, Diagnosis, Development, Neurobiology, and Behavior*. John Wiley & Sons.

Weber, R. & Popova, L. (2012). Equivalence testing in communication research: Theory and application. *Communication Methods and Measures, 6*(3), 190-213.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders*, *37*(10), 1858-1868.

White, S. & Roberson-Nay, R. (2009). Anxiety, social deficits, and loneliness in youth with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(7), 1006–1013.

Wong, C., Odom, S., Hume, K., Cox, A., Fettig, A., Kucharczyk, S., … Schultz, T. (2014). *Evidence-based practices for children, youth, and young adults with autism spectrum disorder.* Chapel Hill, NC: Frank Porter Graham Child Development Institute, Autism Evidence Based Practice Review Group.

APPENDIX

ETHICAL APPROVAL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Institutional Review Board for Human Use

Form 4: IRB Approval Form
Identification and Certification of Research
Projects Involving Human Subjects

UAB's Institutional Review Boards for Human Use (IRBs) have an approved Federalwide Assurance with the Office for Human Research Protections (OHRP). The Assurance number is FWA00005960 and it expires on November 8, 2021. The UAB IRBs are also in compliance with 21 CFR Parts 50 and 56.

Principal Investigator:	KOCH, SARAH A
Co-Investigator(s):	BIASINI, FRED J HOPKINS, MARIA I
Protocol Number:	**F131210001**
Protocol Title:	*Robot-Based Social Skills Intervention for Children with Autism Spectrum Disorders*

The IRB reviewed and approved the above named project on 1/25/2017. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services. This Project will be subject to Annual continuing review as provided in that Assurance.

This project received FULL COMMITTEE review.

IRB Approval Date: 1/25/2017

Date IRB Approval Issued: Jan. 27, 2017

IRB Approval No Longer Valid On: Jan. 25, 2018

Identification Number: IRB00000726

HIPAA Waiver Approved?: N/A

Ferdinand Urthaler, MD / SBH

Ferdinand Urthaler, M.D.
Chairman of the Institutional Review
Board for Human Use (IRB)

Partial HIPAA Waiver Approved?: N/A

Investigators please note:

The IRB approved consent form used in the study must contain the IRB approval date and expiration date.

IRB approval is given for one year unless otherwise noted. For projects subject to annual review research activities may not continue past the one year anniversary of the IRB approval date.

Any modifications in the study methodology, protocol and/or consent form must be submitted for review and approval to the IRB prior to implementation.

Adverse Events and/or unanticipated risks to subjects or others at UAB or other participating institutions must be reported promptly to the IRB.

470 Administration Building
701 20th Street South
205.934.3789
Fax 205.934.1301
irb@uab.edu

The University of
Alabama at Birmingham
Mailing Address:
AB 470
1720 2ND AVE S
BIRMINGHAM AL 35294-0104