J Autism Dev Disord
DOI 10.1007/s10803-014-2334-0

BRIEF REPORT

Brief Report: Development of a Robotic Intervention Platform for Young Children with ASD

Zachary Warren · Zhi Zheng · Shuvajit Das · Eric M. Young · Amy Swanson · Amy Weitlauf · Nilanjan Sarkar

Abstract Increasingly researchers are attempting to develop robotic technologies for children with autism spectrum disorder (ASD). This pilot study investigated the development and application of a novel robotic system capable of dynamic, adaptive, and autonomous interaction during imitation tasks with embedded real-time performance evaluation and feedback. The system was designed to incorporate both a humanoid robot and a human examiner. We compared child performance within system across these conditions in a sample of preschool children with ASD (n = 8) and a control sample of typically developing children (n = 8). The system was well-tolerated in the sample, children with ASD exhibited greater attention to the robotic system than the human administrator, and for children with ASD imitation performance appeared superior during the robotic interaction.

Keywords Autism spectrum disorder · Robotics · Technology · Imitation

Z. Warren (✉)
Department of Pediatrics, Psychiatry, and Special Education, Vanderbilt Kennedy Center, Treatment and Research Institute for Autism Spectrum Disorders, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37206, USA
e-mail: zachary.warren@vanderbilt.edu

Z. Zheng · N. Sarkar
Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA

S. Das
University of Michigan, Ann Arbor, MI, USA

E. M. Young · N. Sarkar
Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

A. Swanson
Vanderbilt Kennedy Center, Treatment and Research Institute for Autism Spectrum Disorders, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37206, USA

A. Weitlauf
Department of Pediatrics, Vanderbilt Kennedy Center, Treatment and Research Institute for Autism Spectrum Disorders, Vanderbilt University, 230 Appleton Pl., Nashville, TN 37206, USA

Introduction

Given the prevalence (CDC 2014) and lifespan costs (Amendah et al. 2011) associated with Autism Spectrum Disorder (ASD), there is an urgent need for the development and application of novel and more efficacious treatment paradigms. Given findings that many very young children with and at-risk for ASD often preferentially orient to nonsocial contingencies, videos, and arrays rather than biological motion or videos (see Annaz et al. 2012; Falck-Ytter et al. 2013), researchers have increasingly proposed advanced technologies, including robotic systems, as potential mechanisms for addressing the substantial limits of current ASD intervention practices (see Diehl et al. 2012). While several recent approaches have highlighted the potential of robotic tools for immediately enhancing brief social interactions (e.g., Scassellati et al. 2012), the ability of systems to target and meaningfully impact core developmental skills and vulnerabilities in a lasting and generalized manner has been significantly limited.

A major historical limit of technological intervention systems has been a reliance on static performance-based protocols (e.g., narrow set of responses, required remote operation) with researchers only recently offering "closed-

Published online: 14 December 2014

loop" systems. "Closed-loop" interaction refers to the ability of a technological system to dynamically interact (i.e., adapt) with a child based on his response in real-time, as opposed to an "open-loop" system where the system behaves in a limited, pre-programmed way. Realizing within-system changes in response to detected behaviors is likely extremely important for individualization of meaningful technological interventions, as effective treatment of young children with ASD often requires extended and meaningful interactions (Yoder and McDuffie 2006).

Although several works have described robust systems for adaptive interaction in children with ASD (e.g., Feil-Seifer and Matarić 2011), to date offered paradigms have had little direct relevance to the core deficits of ASD at young ages, instead focusing on simple task and game performance (e.g., playing basketball) with older school-aged children. More recently Bekele et al. (2012) described the development of an autonomous humanoid robotic system capable of providing joint attention prompts to young preschool children with ASD. The system demonstrated increased attention and focus during robotic interaction as well as brief improved response to joint attention prompts over time (Warren et al. 2013). While such research furthers the argument for closed-loop robotic intervention systems, the work also highlighted the need to move away from technologies that require the use of wearable sensors/technologies (i.e., estimated 40 % non-completion rate) for effective application to young children.

In the current work, we describe the development and initial application of a non-invasive intelligent robotic intervention system capable of dynamic, individualized, targeted interaction for young ASD children with hypothesized potential relevance to improving imitation skills. Imitation is a critically important social communication skill that emerges early in life, is often impaired in children with ASD, and is theorized to play an important role in the development of cognitive, language, and social skills (see Ingersoll 2012). While the technology embedded within the current system could be modified to test impact in other core areas of deficit over time, we primarily focused on early imitation skill learning due to this potential relevance to ASD and the fact that this allowed us to test the potential appeal and capacity for embodied movement with a humanoid robot.

Ultimately, the current work was explicitly motivated by a desire to develop a potential *accelerant robotic intervention technology* (i.e., platform that accelerates learning in core deficit area) with relevance to an important early ASD vulnerability and with the potential for future availability outside of highly specified research environments (i.e., use of commercially available products). In what follows, we describe the development of the non-invasive co-robotic imitation interaction system and then present results of our initial user study. We hypothesized that children (both typically developing and ASD) would tolerate the system quite well and that children with ASD would demonstrate increased attention to the robot during the sessions as compared to a human counterpart. We also conducted preliminary analyses exploring group differences regarding successful imitation of presented gestures (i.e., would children with ASD do better with robot prompts).

Methods

Participants

All children were recruited through existing university-based registries. Eight children with ASD (age $m = 3.83$, $SD = 0.54$; see Table 1) and eight typically developing children (TD: $m = 3.61$, $SD = 0.64$) completed the study. An additional four children with ASD and two TD children were unable to complete the study due to exhibited distress during the protocol. Among the ASD children who failed to complete the study, two exhibited distress in response to the robot with the other children having difficulty sitting in the rifton chair utilized for the experiment without distress. The two TD participants who failed the protocol were siblings who exhibited distress in response to initiating the protocol. All children with ASD received a clinical diagnosis based on DSM-IV-TR (APA 2000) criteria, met the spectrum cut-off on the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2012), and had existing cognitive data with the research registry (Mullen Scales of Early Learning; Mullen 1995). All parents in both groups also completed both the Social Communication Questionnaire (SCQ; Rutter et al. 2010) and the Social Responsiveness Scale (SRS; Constantino and Gruber 2002) to screen for clinically significant ASD symptoms in the TD group and as an index of current symptoms in the ASD group. Although not selected a priori, varying levels of baseline abilities regarding functional and symbolic imitation were present in the sample (ADOS Module 1: Item C2; $m = 1.88$; $SD = 1.13$).

Procedural and System Overview

Children participated in a single research session which involved assessing imitation skills both during brief interactions with a humanoid robot and a human social partner. They were seated in a chair directly facing the robot/human and a Microsoft Kinect sensor (see Fig. 1). During these sessions children's motor behavior was briefly mirrored by the administrator (i.e., either the robot or the human), and

Table 1 Participant characteristics

Mean (SD)	ADOS CS	SRS-2	SCQ	MSEL	Age (years)
ASD mean (SD)	7.63 (1.69)	75.29 (12.62)	17.88 (6.58)	64.75 (22.11)	3.83 (.54)
TD mean (SD)	–	42.75 (10.08)	3.88 (2.95)	–	3.61 (.64)

ADOS CS Autism Diagnostic Observation Schedule Comparison Score, *MSEL* Mullen Scales of Early Learning, Early Learning Composite, *SRS-2* Social Responsiveness Scale—Second Edition, T-Score; *SCQ* Social Communication Questionnaire Lifetime Total Score

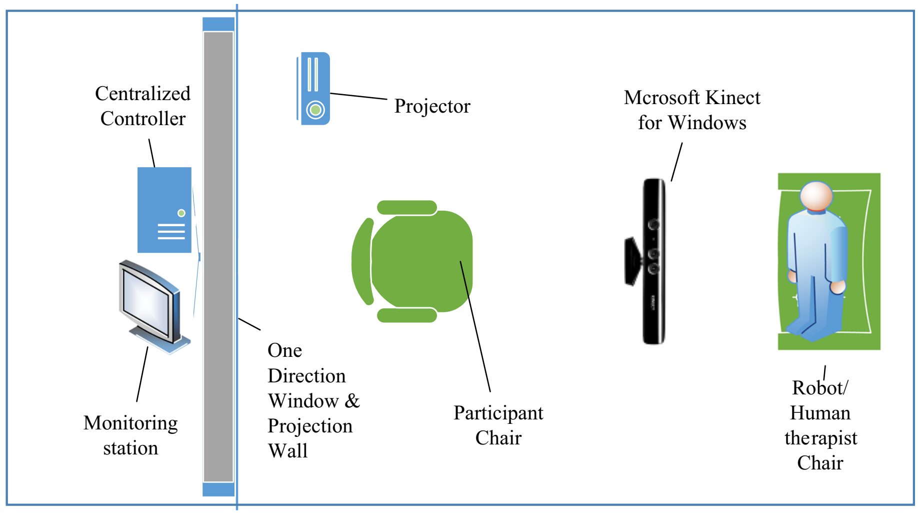

Fig. 1 Experiment room setup

the children were asked to complete a series of four simple motor movements. The Kinect based motion sensing module automatically detected and determined whether or not the child's movements were "correct" and in turn provided feedback in terms of praise or extra prompts. In the humanoid robot condition this information was directly fed to the humanoid robot who would offer praise or additional prompts (i.e., approximating child movements and then completing the gesture). In the human administrator condition instructions were visually conveyed (i.e., projected on wall) such that the human administrator would follow the exact same protocol.

Apparatus

The system was designed and implemented with Microsoft.Net Framework as a component-based distributed architecture capable of interacting via network in real-time (for technical development, gesture mapping, and system validation see Zheng et al. 2014). The humanoid robot, Microsoft Kinect motion sensing module, and projected prompt system for the human administrator were integrated with a supervisory control system (e.g., designed intelligent software) for real-time closed-loop interaction. The system control components were distributed on different threads to achieve parallel control and operation. The robot utilized, NAO, is a commercially available (Aldebaran Robotics Company) child-sized plastic bodied humanoid robot (58 cm tall, 4.3 kg). Microsoft Kinect has an infrared (IR) emitter, an IR depth sensor, and a 1,280 × 960 resolution RGB camera that ultimately gather data that can be translated into skeletal tracking and 3D head pose estimation. We utilized this head pose information as a coarse marker for estimating the participant's attention on robot and human therapist via defining an 85.77 cm × 102.42 cm box around the robot and the upper body of human as target attention regions.

Gesture Selection, Mapping, and Recognition

Four gestural movements were chosen for this study: (1) raising one hand, (2) raising two hands, (3) waving, and (4) reaching arms out to the side. These four gestures were intentionally selected due to the low motor skill requirements they presented to participating children and to avoid motor limitations of the humanoid robot (e.g., challenges crossing midline). In order for the robot to map a participant's gesture, each frame's skeleton data (30 frames per second) were transferred to the robot and displayed by the robot in real-time. We used a model-free rule based method for gesture recognition specifically to avoid challenges acquiring vast amounts of training data needed for statistical model-based gesture recognition methods and to provide a method easily translatable to qualitative categorization of performance. In order to capture a gesture that can vary in speed among participants, we defined five sliding time windows ranging from 1 to 5 s so that any gesture completed within 5 s would be captured. Ultimately, in the current work a "correct" imitated gesture was operationally defined by a sequence of trajectory constraints under a precondition (e.g., regional constraints of the gesture and basic joint positions). The system's embedded supervisory controller continuously monitored the Kinect and determined whether the child's imitative attempt fell within a movement band deemed as "sufficient," and subsequently instructed the robot either to move towards another gesture or aid the child in reinforcement components and approximations of the gestures within their motor movements. The supervisory controller continued this procedure in a closed-loop manner for a specified duration of time and collected data to evaluate the efficacy of the trials towards optimal performance of specific gestures. Previous system validation procedures (Zheng et al. 2014) indicated gesture recognition agreements with clinical observation greater than 98 %.

Trial Procedures

Each child participated in a single experimental session with two human-administered sub-sessions and two robot-administered sub-sessions during which gestures were tested in a randomized order. Prior to the demonstration of each gesture, the robot and human administrator initiated a period of mirroring with the verbal prompt "Let's play! I will copy you!" Subsequently both the robot and human, respectively, imitated the child's movements for 15-s. After mirroring the child, the system started a trial. The robot or human administrator first gave the verbal prompt, "Okay! Now you copy me. Look at what I am doing!" They then demonstrated a gesture two times and said, "You do it!" If correct, the child was presented with praise and asked to imitate the gesture correctly again. If incorrect, the system copied the child's movements and then transferred from this position to the correct display of the gestural movement. The system then asked the child to imitate the presented gesture. Gesture recognition was initiated immediately upon gesture demonstration and ended prior to presentation of subsequent prompts. In the human administrator administered sub-sessions, the human replicated the robot-administered trials via prompts delivered and projected to examiner by the system (i.e., visual cues regarding performance as well as next positions/prompts). This presentation occurred outside of the field of view of the participating children in order to ensure it was not a distraction.

Results

Two primary outcomes of interest, beyond feasibility and tolerability, were examined in the current work: (1) inferred attention to administrator and (2) successful completion of imitation trials. For both outcomes we compared differences between the TD and ASD groups as well as performance differences within groups across conditions (e.g., percentage of time looking at human vs. time looking at robot; performance with robot vs. human). Given the quasi-randomized blocked presentation of trials across the human and robot conditions, we reduced data across all robot trials and all human trials for individual participants. Pre-assuming non-normality of the distributed responses within this small sample, we initially utilized a non-parametric analytic approach to examine within and across group differences. Specifically, within group and specific ASD versus TD group differences were examined via Wilcoxon's and Mann–Whitney test, respectively. However, given the limited power corresponding to this sample size and conservative non-parametric approach, we also examined effect sizes and set a priori statistical significance as either $p < .05$ or Cohen's $d \geq .50$ (e.g., a medium effect size) for composite group differences.

Attention to Administrator

The system continuously gathered head pose data as an inferred marker of gaze during the experimental procedure. In order to guard against confounds of time associated with increased or decreased time necessary for correct performance we examined percentages of time spent directed toward the administrator during sessions as our primary outcome. A two-sided Wilcoxon test did not demonstrate a statistically significant difference between attention paid to the robot and human administrator for either group. However, as seen in Table 2 children with ASD spent a much smaller percentage of time looking at the human administrator than TD children ($p = .046$; $d = 0.79$). Although children with ASD did, as expected, spend a larger percentage of time (11 %) looking toward the robot than human administrator this effect fell just below our significance threshold ($p = .07$; $d = .48$).

Imitation Trial Performance

We also examined performance, or actual demonstrated imitation skills between the human and robot administrator

Table 2 Time spent looking toward robot and human across trials

	Robot session			Human session		
	Attention on target time (s)	Total session time (s)	Ratio (%)	Attention on target time (s)	Total session time (s)	Ratio (%)
ASD mean (SD)	55.01 (28.42)	105.52 (24.47)	52 (24)	43.3 (25.47)	104.35 (23.51)	41 (21)*
TD mean (SD)	61.35 (28.89)	99.69 (29.76)	64 (24)	47.02 (17.3)	86.67 (28.64)	62 (29)*

* $p < .05$ for ASD versus TD in the human session

Table 3 Imitation performance across trials

Group	Graded score		Binary performance	
	Robot session	Human session	Robot session	Human session
ASD mean (SD)	27.31 (32.07)[a]	19.75 (13.64)[b]	2.13 (3.04)[c]	0.88 (1.36)[c,d]
TD mean (SD)	43.75 (28.26)[a]	44.79 (31.98)[b]	3.13 (3.04)	3.75 (3.77)[d]

[a] d > .5 for ASD versus TD in the robot session with graded score

[b] d > 1 for ASD versus TD in the human session with graded score

[c] d > .5 for robot session versus human session binary performance of ASD group

[d] d > 1 for ASD versus TD binary performance in human session

conditions. To evaluate performance, every gesture was scored along a scale from 0 to 10 based on the components of the target skill demonstrated (i.e., segmented intervals approximating the complete movement). This allowed us to evaluate not just binary success (e.g., correct = 10 or incorrect <10), but also partial success and approximations towards the desired target imitative skill (i.e., graded success). Table 3 shows the mean overall gestural performance scores across groups for both graded and binary success. Not surprisingly, TD children were more successful than ASD children imitating the target gestures across both human (d = 1.01 graded) and robot (d = .54 graded) conditions. As in the attention task, children with ASD demonstrated higher imitation performance (38 % more points graded; 17 correct vs. 7 correct binary) in the robot versus the human administrator condition with binary performance differences exceeding the cutoff for meaningful significant effect size (d = .53). Examining individual child performance within the ASD group (see Fig. 2) indicated that some three of the eight participating children demonstrated substantially better imitation performance in the robot condition as compared the human administrator condition (mean performance increase of 42.5 %, range 27.5–62.5 % for these children. Only one child demonstrated improved performance difference in the binary scored human condition (1 vs. 0 correct). Individual performance within the TD group revealed fairly equivalent performance or human trial advantage for a majority of children.

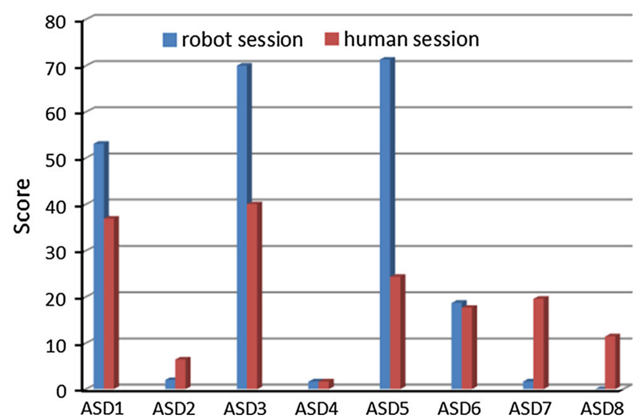

Fig. 2 Individual session performance regarding graded imitation score for children with ASD

Discussion

In the current pilot study, we studied the development and application of an innovative closed-loop adaptive robotic system with potential relevance to a core area of early social deficit of ASD (i.e., imitation skills). In terms of feasibility, we were able to realize an autonomous system capable of administering gestures to the children, assessing the imitated gestures, and providing feedback, all in real-time. Deployment of the system was well-tolerated across both our young ASD and TD samples (i.e., completion rates of 66 % ASD and 80 % TD). Further, children's failure to participate in the protocol was often catalogued as a reaction to the presented stimuli (e.g., robot) or distress in being asked to sit in a rifton chair in a novel treatment room, rather than in relation to an aspect of the technological feedback system itself. In terms of performance within the system, children with ASD spent relatively more time looking at the humanoid robot than the human administrator, a finding replicating previous work suggesting attentional preferences for robotic interactions over brief intervals of time (see Bekele et al. 2012). Preliminary results also suggest many young children with ASD may demonstrate enhanced initial performance in response to robotic prompts than those delivered by human counterparts (i.e., more trials correctly imitated during robot condition).

There are several methodological limitations of the current study that are extremely important to highlight. The small sample size examined and the limited time frame of interaction are potent limits of the current preliminary study. In addition, it is important to note that while we posit this system as a potential framework for learning and teaching imitation skills, we did not systematically intend to assess learning or provide intervention within the scope of the current work. Further, the brief exposure of the current paradigm, in combination with unclear baseline skills of participating children, ultimately cannot answer questions as to whether heightened attention paid to the robotic system or performance differences in conditions displayed during the study are simply the artifact of novelty or of a more characteristic pattern of preference that could be harnessed over time. As such, we are left with data suggesting the potential of closed-loop application across humanoid robot and human partner conditions, but the utilized methodology potently restricts our ability to realistically comment on the value and ultimate clinical utility of this system as applied to young children with ASD. Another important technical limitation was the approximation of attention with head pose which is merely a coarse proxy and does not necessarily equate to actual eye gaze or attention. Systems incorporating these limitations represent realistic next steps for future works and systems, with the current early findings suggesting such work may yield valuable results.

Despite significant limitations, this work is the first to our knowledge to design and empirically evaluate the usability, feasibility, and limited preliminary efficacy of a non-invasive closed-loop interactive robotic technology potentially capable of modifying response based on within system measurements of performance on imitation tasks with young children with ASD. The current work also represents a move towards realization of cost-effect technologies for intervention. Technological intervention systems are often criticized as being unrealistic for application to real world settings due to addition fiscal, operational, and human resources necessary for deployment. Within the current work, we attempted to overcome this critique by designing a robotic system with explicit capacity for autonomous, individualized closed-loop interaction and by utilizing widely available hardware components (e.g., common gaming system, desktop computer, monitor). While it is certainly unlikely that humanoid robots will be present within a majority of homes for children with ASD, the other components of the system are already present in millions of households. In this manner, we attempted to draw closer to systems that could in fact be present in typical households without prohibitive cost in terms of technology or expertise. This potential opportunity is further supported by the system's demonstrated capacity to utilize a human, rather than the humanoid robot, as the primary administrator working in concert with the motion detection technology. Utilizing humans as potential social partners operating in concert with closed-looped intervention platforms may represent a much more proximal strategy for realizing meaningful technological intervention for young children with ASD. Ultimately, as systems with capacities for detecting and responding meaningfully to human behavior (e.g., movement, gaze, facial expression, physiological signals) in non-invasive formats become widely available, traditional pragmatic critiques of technological systems become less challenging.

Ultimately, the current work was explicitly motivated by a desire to develop a potential *accelerant robotic intervention technology* with relevance to an important early ASD vulnerability and with the potential for future availability outside of a highly specified research environment. Movement in this direction introduces the possibility of technological intervention tools that are not simple response systems, but systems that are capable of sophisticated adaptations. Systems capable of such adaptation may ultimately be used to promote meaningful change related to the complex and important social communication impairments of the disorder itself. While we are hopeful that future sophisticated clinical applications of adaptive robotic technologies may demonstrate meaningful improvements for young children with ASD, it is important to note that it is both unrealistic and unlikely that such technology will constitute a sufficient intervention paradigm addressing all areas of impairment for all individuals with the disorder. However, if we are able to discern measurable and modifiable aspects of adaptive robotic intervention with meaningful effects on skills seen as tremendously important to neurodevelopment, or tremendously important to caregivers, we may realize transformative accelerant robotic technologies with pragmatic real-world application of import.

Acknowledgments This study was supported by in part by a Grant from the Vanderbilt Kennedy Center (Hobbs Grant), a Vanderbilt University Innovation and Discovery in Engineering and Science (IDEAS) Grant, the National Science Foundation under Grants 0967170 and 1264462, and the National Institute of Health under Grants 1R01MH091102-01A1 and R21 MH103518. Work also includes core support from NICHD (P30HD15052) and NCATS (UL1TR000445-06). The authors would also express great appreciation to the participants and their families for assisting in this research.

References

Aldebaran Robotics. (2014). http://www.aldebaran-robotics.com/en/.

Amendah, D., Grosse, S. D., Peacock, G., & Mandell, D. S. (2011). Economic costs of autism: A review. In D. Amaral, D. Geschwind, & G. Dawson (Eds.), *Autism spectrum disorders* (pp. 1347–1360). New York, NY: Oxford University Press.

American Psychiatric Associate (APA). (2000). *Diagnostic and statistical manual of mental disorders*—fourth edition. Washington, DC.

Annaz, D., Campbell, R., Coleman, M., Milne, E., & Sweetenham, J. (2012). Young children with autism spectrum disorder do not preferentially attend to biological motion. *Journal of Autism and Developmental Disorders, 42*, 401–408.

Bekele, E., Lahiri, U., Swanson, A., Davidson, J., Warren, Z. E., & Sarkar, N. (2012). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21*(2), 289–299.

Centers for Disease Control and Prevention. (2014). *MMWR weekly: Prevalence of autism spectrum disorders. Autism and developmental disabilities monitoring network*. Retrieved May 4, 2014. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6103a1.htm.

Constantino, J., & Gruber, C. (2002). The Social Responsiveness Scale. *Los Angeles: Western Psychological Services,* 2002.

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

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders, 5*(1), 28. doi:10.1186/1866-1955-5-28.

Feil-Seifer, D., & Matarić, M. (2011). Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. In *Proceedings of the 6th international conference on human–robot interaction* (pp. 323–330). New York, NY: ACM Press.

Ingersoll, B. (2012). Effect of a focused imitation intervention on social functioning in children with autism. *Journal of Autism and Developmental Disorders, 42*, 1768–1773.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2)*. Torrance, CA: Western Psychological Services.

Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Services Inc.

Rutter, M., Bailey, A., & Lord, C. (2010). *The Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

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

Warren, Z. E., Zheng, Z., Swanson, A., Bekele, E., Zhang, L., Crittendon, J., …, & Sarkar, N. (2013). Can robotic interaction improve joint attention skills? *Journal of Autism and Developmental Disorders.* doi:10.1007/s10803-013-1918-4

Yoder, P. J., & McDuffie, A. S. (2006). Treatment of responding to and initiating joint attention. In T. Charman & W. Stone (Eds.), *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention* (pp. 88–114). New York, NY: Guilford.

Zheng, Z., Das, S., Young, E., Swanson, A., Warren, Z. E., & Sarkar, N. (2014). Autonomous robot-mediated imitation learning for children with autism. In *Proceedings of IEEE international conference on robotics and automation (ICRA 2014)*, June 5, 2014. Hong Kong, China. doi:10.1109/ICRA.2014.6907247