Research in Developmental Disabilities 86 (2019) 62–75

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Who is a better teacher for children with autism? Comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition

Wing-Chee So*, Miranda Kit-Yi Wong, Wan-Yi Lam, Chun-Ho Cheng, Sin-Ying Ku, Ka-Yee Lam, Ying Huang, Wai-Leung Wong

Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Intransitive gestures
Children with ASD
Robot-based intervention
Human-based intervention

ABSTRACT

Background: Individuals with autism spectrum disorder (ASD) tend to show deficits in engaging with humans. Previous findings have shown that robot-based training improves the gestural recognition and production of children with ASD. It is not known whether social robots perform better than human therapists in teaching children with ASD.
Aims: The present study aims to compare the learning outcomes in children with ASD and intellectual disabilities from robot-based intervention on gestural use to those from human-based intervention.
Methods and procedures: Children aged six to 12 with low-functioning autism were randomly assigned to the robot group (N = 12) and human group (N = 11). In both groups, human experimenters or social robots engaged in daily life conversations and demonstrated to children 14 intransitive gestures in a highly-structured and standardized intervention protocol.
Outcomes and results: Children with ASD in the human group were as likely to recognize gestures and produce them accurately as those in the robot group in both training and new conversations. Their learning outcomes maintained for at least two weeks.
Conclusions and implications: The social cues found in the human-based intervention might not influence gestural learning. It does not matter who serves as teaching agents when the lessons are highly structured.

What this paper adds

Previous findings have shown that gestural recognition and production of children with ASD improve after robot-based training. Additionally, this kind of training may reduce the gestural delay in children with ASD in their early childhood. Yet, it is not known whether social robots are more effective than human therapists in teaching children with ASD.

The present study compared the learning outcomes of robot-based intervention on gestural use to those of human-based intervention in children with ASD and intellectual disabilities. We designed and implemented an intervention protocol, which involved humans and robots as the teachers of children with ASD. In this study, human experimenters or social robots engaged in daily life conversations and demonstrated to children 14 intransitive gestures in a standardized and highly-structured intervention protocol.

* Corresponding author at: Room 208, Ho Tim Building, Department of Educational Psychology, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region.
E-mail address: wingchee@cuhk.edu.hk (W.-C. So).

https://doi.org/10.1016/j.ridd.2019.01.002
Received 15 June 2018; Received in revised form 16 November 2018; Accepted 4 January 2019
Available online 21 January 2019

Children were randomly assigned to the robot group (N = 12) and the human group (N = 11).

The human group was as likely as the robot group to recognize gestures and produce them accurately in the training conversations. Similar patterns were found in non-training conversations and two weeks later when no training was provided. Thus, there were no significant differences in the learning outcomes in either group of children. The comparable learning outcomes in the human- and robot-based intervention might be attributable to the fact that the training sessions in both conditions were highly structured, thereby yielding positive learning outcomes for children with ASD. When the lessons are highly structured, it does not matter who serve as teaching agents.

1. Introduction

Gestures are spontaneous hand movements produced for communication (e.g., hand waving for bye-bye; hand swiping forehead for an expression of feeling hot). Out of various kinds of gestures, Ham, Bartolo, Corley, Swanson, and Rajendran (2010) proposed that individuals with autism spectrum disorder (ASD) may have selective delayed development in one type of gesture, namely, intransitive gestures (actions without objects but with symbolic meanings). These gestures convey socio-communicative intent (e.g., waving a hand to say goodbye; giving a thumbs-up for a great job; opening arms wide to welcome others). A case report by Ham et al. (2010) found that an 11-year-old child with ASD had difficulty in producing intransitive gestures. In their later work, they found that children with ASD had greater difficulty in recognizing intransitive gestures than gestures that involved actions with objects (Ham et al., 2011). Additionally, a study by Stone and colleagues found that children with ASD have greater difficulty with imitating body movements than actions with objects in comparison to their peers without ASD (Stone, Ousley, & Littleford, 1997). Their results are in line with previous findings, in which young and school-aged children with ASD have specific difficulties in producing markers (i.e., gestures that carry culturally specific meaning for communication, such as the raised thumb for hitchhiking; Charman, Drew, Baird, & Baird, 2003; Luyster, Lopez, & Lord, 2007; Mastrogiuseppe, Capirci, Cuva, & Venuti, 2016; So, Wong, Lui, & Yip, 2015; Wetherby et al., 2004).

Intervention studies designed to teach children with ASD the use of intransitive gestures, however, are scarce. Gesture is an unaided alternative and augmentative communication (AAC) for these children. It can supplement (i.e., augment) existing speech or become the primary (i.e., alternative) method of expressive communication in children with ASD, especially for those with limited verbal abilities (see discussion of manual signs in Mirenda, 2003). Approximately half of them do not attain fluent speech, and a quarter of them do not have functional speech (Frea, Arnold, & Vittimberga, 2001; Hart & Banda, 2009; Ploog, Scharf, Nelson, & Brooks, 2013). This problem is more prominent in children with ASD who also have intellectual disabilities (i.e., IQ < 70; Volkmar & Wiesner, 2009). This group of children thus encounter significant challenges in their basic communication, such as partaking in social interchanges with others (Curcio & Paccia, 1987; Wetherby, Prizant, & Hutchinson, 1998). Therefore, we should provide this group of children with an alternative channel to express their ideas and thoughts.

Recently, So et al. conducted the first multi-phase robot-based intervention study on gestural use for Chinese-speaking school-aged children with ASD and intellectual disabilities (So et al., 2016). These children were taught to recognize, imitate, and produce 20 gestures of different types, demonstrated by a robot animation in three phases. Their results reported that children recognized and imitated more gestures as well as producing them in appropriate social contexts after training. There were significant differences between the pretests and posttests across the three phases. The children also generalized their acquired gestural skills to a novel setting with a human researcher. While robot animation may be initially exciting, children with low-functioning ASD (especially those with short attention span) may quickly lose interest. Significant coaching and encouragement are required to make them stay focused. In addition, the animated robot is not able to interact with children participants. A possible back and forth interaction between the robot and the child is necessary during intervention. Therefore, in order to maximize the impact of intervention, So, Wong, Lam, Lam et al.'s study (2018) used a real social robot, which is more engaging than an animated one. In this study, they used a real social robot (as opposed to robot animation) to teach Chinese-speaking school-aged children with ASD and intellectual disabilities to recognize and produce eight intransitive gestures that express feelings and needs (e.g., NOISY) (So, Wong, Lam, Cheng et al., 2018), in two phases. Compared to the students in the wait-list control group who had not received training, students in the intervention group were more capable of recognizing and producing the eight gestures produced in the trained and non-trained scenarios. Significant differences between the pretests and posttests were found across the two phases. Even more promising, these students could recognize the same gestures produced by human experimenters. However, there was no strong evidence showing that the children in the intervention group could generalize the acquired gestural production skills to interactions with human experimenters. A recent study by So and colleagues provided gestural training to preschoolers with autism (So, Wong, Lam, Cheng et al., 2018). Their findings showed that after training these children could catch up to the level of gestural production found in children with typical development. In addition, they were also more likely to produce verbal markers while gesturing than those in the wait-list control group.

At present, previous findings have shown that the gestural recognition and production of children with ASD improve after receiving robot-based training. Additionally, this kind of training may reduce the gestural delay in children with ASD in their early childhood. These results are encouraging, as social robots in educational and clinical settings provide an effective treatment for gestural communication in children with ASD. Yet, it is not known whether social robots perform *better* than human therapists in teaching gestural communication skills to children with ASD. The present study aimed to compare the learning outcomes of robot-based intervention on gestural use to those of human-based intervention in children with ASD and intellectual disabilities.

According to the social motivation theory of autism, individuals with ASD show deficits in orienting toward social stimuli, engaging with humans, and maintaining social relations (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). Empirical evidence has supported this theory showing that individuals with ASD tend to have low interest in other humans and have a weaker understanding of the interpersonal world than of the object-related world (Klin & Jones, 2006; Klin, Lin, Gorrindo, Ramsay, & Jones, 2009). In addition, they find it challenging to pay attention to multiple cues during social interactions with humans (Koegel, Koegel, Shoshan, & McNerney, 1999). Thus, they are not sensitive to other people's behaviors (Lee, Takehashi, Nagai, Obinata, & Stefanov, 2012). Therefore, some studies have shown that individuals with ASD may find it difficult to learn social skills from human therapists. For example, Dewey, Cantell, and Crawford (2007) showed that five to 18 year old individuals with ASD (with a wide range of IQs) were less likely to imitate isolated gestures, which were demonstrated by a human experimenter, and produce gestures on command than children with other developmental disorders.

In contrast to human therapists, social robots may be more suitable for individuals with ASD. Based on the empathizing-systemizing theory (Baron-Cohen, 2009), robots are operated on predictable and lawful systems, thereby providing children with ASD with a highly structured learning environment and helping them to focus on the relevant stimuli (Bölte, 2009; Duquette, Michaud, & Mercier, 2008). Additionally, children with ASD do not need to consider socio-emotional expectations when interacting with robots (Silver & Oakes, 2001), thus reducing their social anxiety (Mitchell, Parsons, & Leonard, 2007). Social robots have been widely used in therapy for individuals with ASD over the past decade (Cabibihan, Javed, Ang, & Aljunied, 2013; Fong, Nourbakhsh, & Dautenhahn, 2003; Li, Cabibihan, & Tan, 2011). Children with ASD treat their talking robot as a social agent, which attracts their attention (Kozima, Michalowski, & Nakagawa, 2009; Miyamoto, Lee, Fujii, & Okada, 2005). Social robots are also found to arouse interest in children, thereby eliciting positive and productive responses from them (Scassellati, Admoni, & Matarić, 2012), which in turn helps them to develop joint attention behaviors, self-initiated interactions, non-verbal communication skills, and an ability to make eye contact (Ricks & Colton, 2010; Werry, Dautenhahn, Ogden, & Harwin, 2001).

Several pilot studies have been conducted to evaluate whether individuals with ASD respond to social robots more favorably than to human beings. Previous studies have shown that children with ASD respond faster (Bird, Leighton, Press, & Heyes, 2007; Pierno, Mari, Lusher, & Castiello, 2008), manifest eye contact more frequently (Tapus et al., 2012), and need fewer prompts (Vanderborght et al., 2012) when interacting with a robot than when interacting with a human. Shamsuddin et al. (2013) even reported that children with ASD produced fewer stereotyped behaviors (e.g., avoiding eye contact, turning in circles, rocking back and forth) during the interaction session with a robot than in the regular classroom setting in which a human teacher was involved. Michaud et al. (2007) also reported similar findings. Additionally, they found that children with ASD who interacted with a robot were more likely to show shared attention and to imitate facial expressions than those who interacted with a human mediator. A study by Kim et al. (2013) showed that children with ASD spoke more when interacting with a social robot than when interacting with an adult during a computer game. Lee and Obinata (2015) further used the robot to provide feedback to children with ASD in a task and found that these children accomplished more than their peers who received feedback from a caregiver and a laptop. On the other hand, Wainer, Dautenhahn, Robins, and Amirabdollahian (2014) found that children with ASD who played with an autonomous robot in the first session were more engaged in the video game with a human partner in the second session, whereas those who played with a human in the first session did not show any increase in engagement in the second session. This result was replicated in their later study (Wainer, Robins, Amirabdollahian, & Dautenhahn, 2014). These results suggest that children with ASD might have gained some social skills when interacting with the autonomous robots and they transferred those skills to human-to-human interactions.

However, some opposite findings have also been reported. Some studies found no difference in the speech and social behaviors elicited by children with ASD subjected to robot or human conditions (Huskens, Verschuur, Gillesen, Didden, & Barakova, 2013; Kim et al., 2013; Pop, Pintea, Vanderborght, & David, 2014). A recent study by Simut, Vanderfaeillie, Peca, Van de Perre, and Vanderborght (2016) reported that children with ASD subjected to robot or human conditions did not differ in the initiation of joint attention, number of verbal utterances, and displays of positive affect. Yet children had more eye contact with the robot than with the human. Duquette et al. (2008) and Wainer, Ferrari, Dautenhahn, and Robins (2010) even found that children with ASD perform more poorly when interacting with a robot than with a human.

Therefore, there is no conclusive evidence supporting the findings that children with ASD respond to social robots more favorably than to human beings. Even more importantly, one should note that previous research was in the form of pilot studies looking at how well children with ASD engaged with social robots during play or interaction. While some studies showed that social robots are more likely to increase engagement in children with ASD than humans, it is not certain from these studies whether robots can effectively *teach* these children social and communication skills. This question remains under-studied because previous studies did not adopt an intervention protocol whereby children with ASD learned from social robots demonstrating certain social and communication skills and later had their learning outcomes assessed after training. The present study aimed to address this question by designing and implementing an intervention protocol that involved humans and robots as the teaching agents of children with ASD and compared the effectiveness of robot-based intervention to those of human-based intervention.

In this study, human experimenters or social robots engage in daily life conversations and demonstrate 14 intransitive gestures in a standardized intervention protocol to children with ASD and intellectual disabilities. Children were randomly assigned to the robot group and the human group. We measured children's abilities to recognize and produce the gestures before and after training in both groups. We also evaluated how well children in both groups engaged with the robot and human teachers by looking at whether they established eye gaze with the teachers and imitated the gestures demonstrated by them.

Besides standardized gestural recognition and production assessments, we also administered neuropsychological assessments. These assessments evaluated the children's fine motor skills, attention, and visual perception that could influence their gestural learning, accounting for individual variations in the learning outcomes. Individuals with ASD may have motor deficits due to dysfunction in the mirror neuron system (MNS) (Cattaneo et al., 2007), which may in turn influence their ability to imitate and produce gestures. Motor deficits in autism can be subdivided into two main categories: (a) deficits in basic motor control; and (b) difficulty with praxis performance. The latter is associated with social, communicative, and behavioral impairments that are typical of autism (Dowell, Mahone, & Mostofsky, 2009; Dziuk et al., 2007). Recent research has shown that praxis skills are found to be correlated to the imitation of gestures (Gizzonio et al., 2015), which requires the production of coordinated sequences of movement. Attention influences learning in general. Previous research has shown that children with low-functioning autism, who may be more easily distracted by irrelevant stimuli, tend to look at robots and follow their instructions more often than their peers do (Bekele, Crittendon, Swanson, Sarkar, & Warren, 2014; Duquette et al., 2008). Recent studies also reported that these children seem to be more focused when interacting with a social robot, NAO, than when interacting with teachers and peers (Shamsuddin, Yussof, Ismail, Mohamed, Mohamed et al., 2012; Shamsuddin, Yussof, Ismail, Mohamed et al., 2012; Shamsuddin, Yussof, Ismail, Hanapiah et al., 2012). Finally, visual perception may influence children's ability to learn the forms of gestures demonstrated by the robot or human. In addition to the neuropsychological assessments, we also administered theory of mind understanding tasks (Wellman & Liu, 2004; Wellman, Fang, Liu, Zhu, & Liu, 2006). Theory of mind understanding might be crucial in our gestural production assessments as children were required to take the perspective of the characters involved in the conversations and produce appropriate gestures.

In accordance with the social motivation theory of autism (Chevallier et al., 2012) and the empathizing-systemizing theory (Baron-Cohen, 2009), robot-based intervention should be found to be more effective than human-based intervention. If so, after controlling for fine motor skills, attention, visual perception, and theory of mind understanding, children in the robot group should recognize the meanings of intransitive gestures and accurately produce these gestures more often than those in the human group. Alternatively, based on the previous findings (Huskens et al., 2013; Kim et al., 2013; Pop et al., 2014; Simut et al., 2016), the robot- and human-based intervention should be found to be equally effective. If so, children in both groups should have similar performances in both gestural recognition and production. This study would provide us with insights into the effectiveness of robot- and human-based interventions on learning, specifically gestural use, in children with ASD. Our findings would shed light on the application of social robots in educational settings.

2. Method

2.1. Participants

A total of 23 Chinese-speaking (Cantonese-speaking) children aged six to 12 participated in this study. The participants were randomly assigned to two groups: the robot-based intervention group and the human-based intervention group. Children in the robot-based intervention group were taught by social robots, while those in the human-based intervention group were taught by human experimenters. The mean age of the children in the robot-based intervention group was 9.17 (two females; $SD = 1.29$; range 6.99–10.75) and that of the human-based intervention group was 8.92 (one female; $SD = 0.93$; range 6.83–9.67). The Man-Whitney test showed that there was no significant difference in age between the two groups, $U = 21$, $p < .75$.

The children participating in the study had been diagnosed with autism or another autistic disorder between the ages of 18 and 60 months ($M = 36.27$; $SD = 18.54$) by pediatricians at the Child Assessment Center for the Department of Health in Hong Kong. All the children were attending the Hong Chi Morninghill School, Tsui Lam, Tseung Kwan O. This is a school in Hong Kong for children diagnosed with ASD and mild to moderate intellectual disabilities. Their ASD diagnoses were further confirmed by clinical psychologists and pediatricians from the Pamela Youde Child Assessment Center, Hong Kong, through standard clinical interviews with their parents and based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (APA, 2010). All the procedures were approved by the institutional review board of the first author's university, in compliance with the Declaration of Helsinki (Reference no. 14600817). We obtained the parents' informed consent prior to the study. The children also gave their consent to participate in this study.

The children's IQs were assessed by qualified clinical psychologists from the Pamela Youde Child Assessment Center. Ten children had their IQs assessed using the Wechsler Intelligence Scale for Children®, Fourth Edition (Hong Kong; WISC IV-HK); their IQs ranged from 57 to 74 ($M = 65.56$; $SD = 6.28$). Other children were not capable of completing the WISC IV-HK subtests. Two children had their IQs assessed using the Stanford Binet Intelligence Scale (Fourth Edition; SB:FE); their IQs were 46 to 51 respectively. The remaining children took the Wechsler Preschool and Primary Scale of Intelligence – Revised; their IQs ranged from 47 to 70 ($M = 48.25$; $SD = 1.23$).

At the beginning of the experiment, all the children with ASD took neuropsychological tests. Specifically, they had their fine motor skills, attention skills, and visual perception assessed, as these skills could influence their gestural learning. Children's fine motor skills were assessed by the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT™-2; Bruininks & Bruininks, 2005). We examined their fine motor skills, as opposed to gross motor skills, because fine motor skills are related to the production of the intransitive gestures taught in the present study. We report the Fine Motor Composite standardized score here, which consists of the following subtests: Fine Motor Precision, Fine Motor Integration, Manual Dexterity, and Upper-Limb Coordination. There was no significant difference in the composite scores of both groups of children, $U = 62.00$, $p < .81$.

The children's attention skills were measured by the Attention Network Test (ANT) (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Rueda et al., 2004). This test lasts for half an hour and provides a nonverbal measure of the efficiency of the attentional networks involved in alerting, orienting, and executive attention across all ages in both typical and atypical populations. We focused on executive attention, which requires the children to respond by pressing two keys indicating the direction (left or right) of a central arrow surrounded by congruent, incongruent, or neutral flankers. Thus, it evaluates one's ability to focus on the relevant stimulus while ignoring the distracting but irrelevant stimuli. For each child, we averaged the proportion of the trials to which he or she responded correctly in the congruent and incongruent conditions. There was no significant difference between the groups, $U = 52.50$,

Table 1
Performance of various pre-assessment tasks in both groups of children.

Groups	Descriptive Statistics	Chronological Age	Standardized score in BOT	Proportion of accurate trials in ANT	Score in VP	Score in ToM
Robot-based intervention (N = 12)	Mean	9.14	63.59	0.59	6.83	2.58
	SD	1.29	8.03	0.27	2.04	1.92
	Minimum	6.99	49.00	0.13	4.00	0.00
	Maximum	10.69	73.00	0.97	11.00	6.00
Human-based intervention (N = 11)	Mean	8.88	64.63	0.50	6.27	2.45
	SD	0.97	8.79	0.28	3.34	1.69
	Minimum	6.83	49.00	0.13	1.00	0.00
	Maximum	9.95	75.00	0.94	14.00	5.00

Notes: BOT = Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT™-2; Bruininks & Bruininks, 2005); ANT = Attention Network Task (Rueda et al., 2004); VP = Beery Visual Perceptual Subtest (Beery & Beery, 2004); ToM = Theory of Mind Understanding Task (Wellman & Liu, 2004).

$p < .41$.

Their visual perception was assessed using the Beery Visual Perceptual Subtest (VP), which required the children to match geometric shapes (Beery & Beery, 2004). We evaluated the children's visual perception in accordance with the scoring manual (Martin, 2006). The maximum possible score is 16. There was no significant difference between the groups, $U = 57.50$, $p < .61$.

Besides administering neuropsychological assessments, we also measured the children's theory of mind understanding, which was assessed by the Chinese version of six tasks: Diverse Desire, Diverse Belief, Knowledge Assess, Content False Belief, Explicit False Belief, and Hidden Emotion. These tasks were developed by Wellman and Liu (2004) and Wellman et al. (2006). The maximum possible score is 6. There was no significant difference between the groups, $U = 64.00$, $p < .92$.

The order of these assessments was counterbalanced across the participants. Table 1 shows the descriptive statistics of the performance in each assessment for both groups of children.

2.2. Stimuli

2.2.1. Target gestures

A total of 14 intransitive gestures that are commonly used in daily life were taught in the robot- and human-based intervention programs. The findings of a study by Cabibihan, So, and Pramanik (2012) showed that these gestures are easily recognized by speakers in Chinese society. Additionally, typically developing individuals could assign similar meanings to the same gestures produced by the NAO robot and by the human experimenter (Cronbach's alpha = .83) (So, Wong, Lam, Cheng et al., 2018).

2.2.2. Social robot

Two NAO (Aldebaran Robotics Company) robots were programmed to speak and / or produce the 14 gestures (see Fig. 1). They performed role plays in the conversations, with one robot acting Mary, and another Sally. The NAO robot has been widely used in autism therapy. It is 50 cm tall and anthropomorphic. It was deployed in the present study because it might facilitate children with ASD generalizing the acquired imitation and social skills to human-to-human interactions (Cabibihan et al., 2013). Besides, unlike other robots, NAO robots can produce a wide range of gestures. The NAO robot contains 25° of freedom (DOF) from 15 joints and actuators. Our gestures were accomplished by 14 DOF from nine joints and actuators. Each of the gestural movements required two DOF in the neck, two DOF in each shoulder, two DOF on each elbow, and one DOF on each wrist. Each gesture lasted for three to four seconds, which captured the period of the time the NAO robot prepared to produce the gestures, moved its body parts, and resumed to its original resting position).

2.2.3. Conversations

In the training sessions, two NAO robots acting different characters engaged in a series of conversations (S1, training conversations). The NAO robot acting Ann gestured while speaking. For example, one of the S1 conversations, which was happened in a classroom setting, was as follows:

NAO robot acting Mary: "Ann, what would you say to the teacher after school?"
NAO robot acting Ann: "I would say bye-bye to the teacher after school."

Each conversation in S1 contained two sentences and one gesture (e.g., hand waving for the BYE-BYE gesture in the previous example). Each gesture only appeared once across all the conversations. Thus, there were 14 conversations covering all 14 gestures. This set of conversations, S1, was used for training and assessment in the standardized pretests and posttests.

The conversations in S1 were held between two female Cantonese speakers (one acting Mary and another Ann) and their narrations were recorded as audio clips. To make the recordings sound like the speech produced by robots, robotic effects were added and the speech rate was reduced using an audio editor (Audacity, v. 2.1.0, the Audacity Team, US state). A total of 14 audio clips were made, each containing one conversation, and these were imported into two NAO robots. The mean length of each clip was 15.71 s

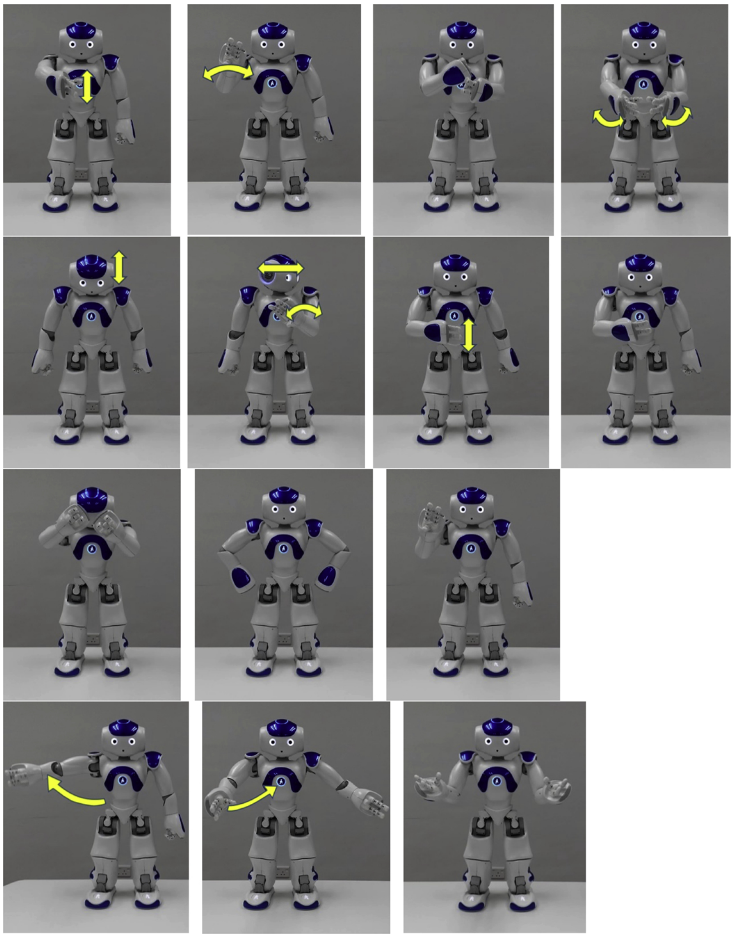

Fig. 1. Gestures performed by NAO robot. From the upper left corner and proceeding from left to right, the following gestures are: (first row) hello, bye, wrong, hurray, (second row) yes, no, hungry, me, (third row) annoyed, angry, wait, (forth row) welcome, come and where.

(ranging from 10 to 20 s). The robots then played the audio clips and gestured during the narration. The gesture and its accompanying speech started at the same time. Thus, the children with ASD in the robot-based intervention condition watched the gestures while listening to the stories. For example, in the previous conversation, the NAO robot acting Ann said and gestured: "*I would say bye-bye (RIGHT HAND WAVES) to the teacher after school.*" In addition to the verbal narration, the background images of the conversations (e.g., a classroom) were visually displayed on the laptop screen, which was placed next to the NAO robots. One picture was shown for each conversation.

The training conversations (including verbal narrations and gestures) demonstrated in the human-based condition were exactly the same as those in the robot-based condition except that two female human demonstrators (instead of robots) engaged in the conversations. Both human demonstrators behaved naturally. They spoke with each other with their usual intonation and facial expressions during the training.

Another set of conversations (S2, non-training conversations) contained 14 conversations that were different from those in S1. S2 was presented during the assessment (but not training) sessions to examine the generalization effects of the intervention in the new context. The conversations in S2 were held between two female Cantonese speakers.

2.3. Procedures

The experiment was conducted in a treatment room at Hong Chi Morninghill School, Tsui Lam, Tseung Kwan O, in Hong Kong. The treatment room was often used by the children for school activities. The room was equipped with two NAO robots, a laptop, and a camera in front of the child (see Fig. 2). The camera videotaped the hand movements the child produced in the sessions.

The robot- and human-based interventions lasted for nine weeks. They consisted of pretests (one for each set of conversations: S1, S2), four training sessions for S1 (with two 30-minute sessions per week), immediate posttests (which were the same as the pretests), and the same follow-up posttests after two weeks. The posttest for S1 assessed the training effects, while that for S2 assessed the generalization effects. For each training or assessment session, the participants were accompanied by a teacher. The training and

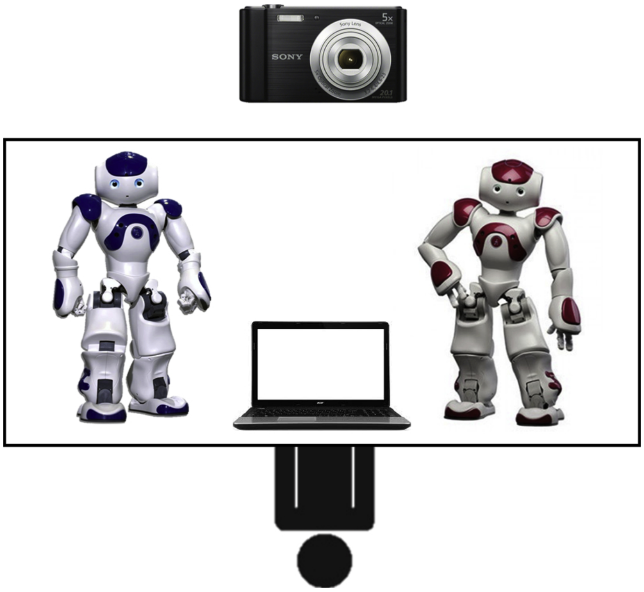

Fig. 2. The experimental setting for training in the intervention condition.

assessment sessions were administered by a researcher, who was either the assistant or one of the authors. A small reward by way of positive reinforcement (snacks or access to toys) was offered by the teacher at the end of each pretest, posttest, and training session. All the sessions were videotaped. Each session lasted for approximately 30 min. Details of the intervention program are provided below.

There were three pretests assessing the recognition and production of 14 intransitive gestures. At the beginning of each pretest, the human researcher first greeted the participant, then gave instructions. The researcher then asked the participant whether he or she understood the instructions. If the participant indicated that he or she did not understand the instructions, the researcher would repeat them. Otherwise, the human researcher proceeded to administer the pretest.

Regarding the pretest for gestural recognition, we followed the protocol administered in previous studies (So et al., 2016; So, Wong, Lam, Lam et al., 2018; So, Wong, Lam, Cheng et al., 2018). We videotaped a human producing the 14 intransitive gestures that were taught in the present study. The human researcher then presented the video clips to the children. Each time, a child was shown a gesture (e.g., both hands clapping) and was asked to choose one of three options that best identified its meaning (e.g., HELLO, AWESOME, WELCOME). One of the options was the correct answer while the remaining two were the distractors. The three choices were presented verbally and in text. Its order was counterbalanced across participants. For each child we reported the proportion of trials he or she was able to recognize as the correct meanings of the gestures.

The two pretests of gestural production evaluated the children's ability to produce intransitive gestures in different conversations in S1 and S2 respectively. Two human experimenters (they were not the human demonstrators in the human-based intervention condition) played the roles of the characters in the conversations in S1 and S2, with the corresponding background pictures sequentially displayed on the laptop screen. After demonstrating a conversation once, the human researcher asked the child to demonstrate the corresponding gesture (e.g., "*What are the hand movements for expressing bye-bye?*"). The child was given 10 s to respond. The researcher prompted the child if he or she gave no response and gave the child another 10 s to respond. Upon receiving the child's response, the researcher then judged the accuracy of the gesture produced (based on four parameters, which are further elaborated in the following session) and provided feedback ("Your hand movements are correct" or "Your hand movements are incorrect"). Then the two human experimenters proceeded to the next conversation. The two pretests were completed after the child had been asked to demonstrate the individual gestures in the conversations in S1 and S2. Each pretest lasted for approximately 30 min. A small reward by way of positive reinforcement was provided after the pretest.

After the pretest, children in the robot-based intervention group proceeded to training. One of the robots greeted the child followed by giving an instruction, "Today we are going to demonstrate conversations and actions. After each demonstration, we will request you to imitate the action. Do you understand?" Training started when the child indicated that he or she understood the instructions. Otherwise, the robot repeated them. During training, two robots engaged in a series of conversations in S1 (with each demonstrated twice) while producing appropriate gestures. Each time, the child was asked to imitate the gesture. A prompt was given if he or she did not imitate it. The training was completed after each of the 14 conversations had been demonstrated twice and the corresponding gestures had been imitated by the child in S1. Those in the human-based intervention group went through the same training procedure except that they watched two human demonstrators perform the role-plays. The two human demonstrators participating in the training sessions were the same for all the children but they were not involved in the pretests and posttests.

All children then took the posttests immediately after the training and the delayed posttests two weeks after the training. The human experimenters and researcher administering the posttests were the same as in the pretests. The procedures in both posttests were the same as those in the pretests. All children were able to pay attention during the training and assessments. A short break was given to the participants if requested. None of them were absent from the sessions.

2.4. Coding and scoring

We asked a research assistant, who did not know the objectives of the study and was unaware of the research questions of the present study, to watch the videos of the participants and count the number of trials in which they recognized the gestures and in which they produced the gestures correctly, according to four parameters (Stokoe, 2005): use of hand / hands (e.g., placing right / left hand against the head vs. using both hands); hand-shape (e.g., open palm vs. curled palm vs. fist); direction of movement (e.g., head nods vs. head shakes; moving hand from left to right vs. moving it up and down); and placement (e.g., hand placed on the head vs. on the chest). The following gestures were considered incorrect: using only the left hand to produce the WHERE gesture (reason: incorrect use of hands); making a fist when producing the WAIT gesture (reason: incorrect hand-shape); moving the right hand downward when producing the NOT ALLOWED gesture (reason: incorrect direction of movement); and covering the face when producing the MYSELF gesture (reason: incorrect placement).

Besides looking at the performance of assessments in both groups of children, we also examined whether they established eye contact with their teachers, that is to say, with human demonstrators or with NAO robots, and whether they imitated the gestures demonstrated by the teachers. We counted the number of times in which eye contact and gestural imitation occurred respectively. For each gesture, the child was granted eight times to learn it from their teachers (each gesture was taught twice in a conversation in each training session, and altogether eight times in four training sessions). A child was considered to establish eye contact with the teacher if he or she looked at the robots or human demonstrators for at least 2 seconds in each conversation. A child was considered to imitate the gesture if he or she did so without prompting.

We then trained a second coder, who equally did not know the objectives of the study and was unaware of the research questions of the present study, to code gestures. She watched 20% of the videos. The inter-observer agreement in an evaluation of the accuracy of gesture production was .93 (N = 194; Cohen's Kappa = .90, $p < .001$), and in the coding of eye gaze and gestural imitation it was .92 (N = 258, Cohen's Kappa = .90, $p < .001$) and .92 (N = 258, Cohen's Kappa = .90, $p < .001$) respectively.

3. Results

This study examined whether the robot-based intervention on gestural learning was as effective as human-based intervention. We compared the gestural recognition and production performance of children with ASD in the robot-based intervention group to that of children in the human-based intervention. If robot-based intervention was found to be more effective than human-based intervention, the proportion of trials with gestures recognized and produced accurately in the robot-based intervention would be higher than those in the human-based intervention. If both kinds of interventions were found to be equally effective, the proportions would be comparable in both groups. Appendix A shows the results of individual children in their gestural recognition and production tasks as well as their performance in the neuropsychological assessment and theory of mind understanding tasks.

All children produced the intransitive gestures when instructed or prompted by the human researcher. We first report the correlations between chronological age, fine motor skills, attention skills, visual perception, gestural recognition performance in the pretests and posttests, and gestural production performance in the pretests and posttests (the data in both S1 and S2 were collapsed). This is followed by the proportion of trials in which the children with ASD in the robot-based intervention condition and those in the human-based intervention condition accurately recognized and produced gestures in the pretests and immediate and delayed posttests. Finally, we report on the frequency in both groups with which children established eye contact with their teachers during training and imitated the gestures demonstrated by them.

There were significant correlations between attention, visual perception, ToM, and gestural recognition and gestural production performances in the pretests and / or posttests. Chronological age and fine motor skills were not correlated to the gestural production and recognition performance, and, therefore, they were not included in the later analyses. See Table 2.

We then examined whether children in the robot-based intervention group performed better in the gestural recognition and production assessments after training than those in the human-based intervention group (see Fig. 3). Separate generalized linear mixed effects model analyses (GLME model; Bates et al., 2014) were conducted for gestural recognition and production using the statistical analysis tools R (R Core Team, 2012). The first GLME model analysis examined whether the children in the robot-based intervention group showed significant differences in gestural recognition across all time points (pretest, immediate posttest, delayed posttest), in comparison to the children in the human-based group. An interaction between the time point and intervention was expected and, as such, an interaction term was included in the model as a fixed effect. Additionally, we entered attention, visual perception, and ToM, which were significantly correlated to gestural recognition, as fixed effects. For random effects, intercepts were created for individual participants and gesture items, as were by-subject and by-gesture random slopes for the effects of both time point and intervention. Children's performance for each recognition trial was the binomial dependent variable. After controlling for attention, visual perception, and ToM, the time point x group interaction was not significant, $\beta = .03$, $SE = .02$; $t = 1.21$, $p < .23$. The time point was found to affect gestural recognition for both groups, $\beta = .15$, $SE = .01$; $t = 6.92$, $p < .001$. Thus, human-based intervention was found to be as effective as robot-based intervention in training on gestural recognition. Both groups showed a significant improvement in the immediate and delayed posttests for gestural recognition after receiving training.

The second GLME model analysis examined whether the children in the robot-based intervention group showed significant differences in gestural production across all time points (pretest, immediate posttest, delayed posttest), in comparison to the children

Table 2
Correlations between chronological age, fine motor skills, attention skills, visual perception, theory of mind understanding, gestural recognition and gestural production accuracy in pretests and immediate and delayed posttests.

	Chronological age	Fine motor skills	Attention skills	Visual perception	ToM	Gestural recognition accuracy in pretest	Gestural recognition accuracy in immediate posttest	Gestural recognition accuracy in delayed posttest	Gestural production accuracy in pretests	Gestural production accuracy in immediate posttests	Gestural production in delayed posttests
Chronological age	–										
Fine motor skills	.08	–									
Attention skills	−.02	−.18	–								
Visual perception	.001	.21	.48*	–							
ToM	−.02	.26	.58**	.27	–						
Gestural recognition accuracy in pretest	.36	.34	.59**	.34	.81***	–					
Gestural recognition accuracy in immediate posttest	.13	.20	.70***	.30	.65**	.71***	–				
Gestural recognition accuracy in delayed posttest	.19	.12	.70***	.44*	.65**	.72***	.90***	–			
Gestural production accuracy in pretests	.44	.18	.44*	.49*	.42*	.39	.52**	.48**	–		
Gestural production accuracy in immediate posttests	.02	.10	.56**	.37	.48*	.57**	.78***	.74***	.69**	–	
Gestural production in delayed posttests	−.008	.12	.68***	.36	.49*	.56**	.82***	.75***	.67**	.92***	–

Notes:
*** $p < .001$.
** $p < .005$.
* $p < .05$.

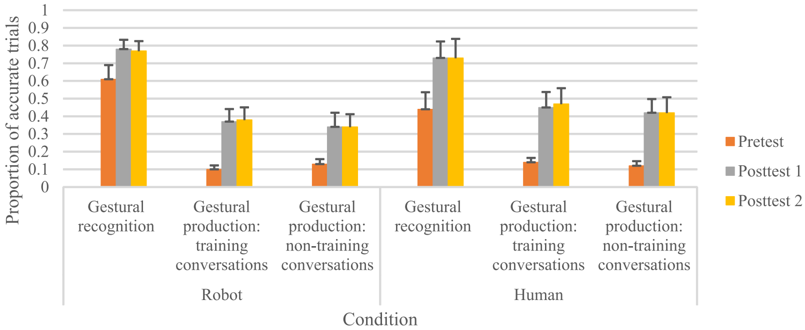

Fig. 3. The proportions of accurate trials in the gestural recognition pretests and immediate and delayed posttests and those in the gestural production pretests and immediate and delayed posttests of both training and non-training conversations in children with ASD in both conditions.

in the human-based intervention group, in both sets of conversations (S1, S2). Children's performance for each production trial was the binomial dependent variable. This model was similar to the first one except that conversations was added as a fixed effect. The time point x group interaction was not significant, $\beta = .01$, $SE = .06$; $t = 0.21$, $p < .83$. Conversations and its interaction with group and time point were not significant, $ps < 0.53$. The time point was found to affect gestural production for both groups, $\beta = .18$, $SE = .04$; $t = 3.86$, $p < .001$. Similar to gestural recognition, human-based intervention was found to be as effective as robot-based intervention in training on gestural production. Both groups showed a significant improvement in the immediate and delayed posttests for gestural production in both training and non-training conversations. Yet, the teaching agent did not play a significant role here.

The third GLME model analysis focused on the percentage of parameters erroneously produced (see Fig. 4). This model was similar to the second one except that the dependent variable was continuous. Similar to the results of the second model, the time point x group interaction was not significant, $\beta = .001$, $SE = .06$; $t = 0.02$, $p < .96$. Conversations and its interactions with group and time point were not significant, $ps < .40$. Intervention, either human-based or robot-based, was found to significantly reduce the percentage of parameters erroneously produced in the immediate and delayed posttests for children with ASD, $\beta = .18$, $SE = .04$; $t = 4.41$, $p < .001$.

Although the learning outcomes in gestural production and recognition did not differ between the robot- and human-based intervention groups, we examined how well children with ASD in both groups engaged with their teachers during training by looking at the number of trials in which they established eye contact with the teachers and imitated the gestures as instructed (see Fig. 5). To address these issues, we conducted two GLME model analyses on eye gaze and gestural imitation respectively. Regarding eye gaze, we entered group, attention, visual perception, and ToM as the fixed effects and individual participants and gesture items as the random effects. The number of trials in which children established eye contact with their teachers for at least two seconds was the continuous dependent variable. After controlling for attention, visual perception, and ToM, the group effect was significant, $\beta = .96$, $SE = .24$; $t = 2.21$, $p < .04$, suggesting that children in the robot-based intervention group were more likely to establish eye contact with the teachers than those in the human-based intervention group. We conducted a similar analysis for gestural imitation, with the number of trials in which children imitated gestures as the continuous dependent variable. However, the group effect was not significant in this scenario, $\beta = .56$, $SE = .51$; $t = 1.19$, $p < .24$.

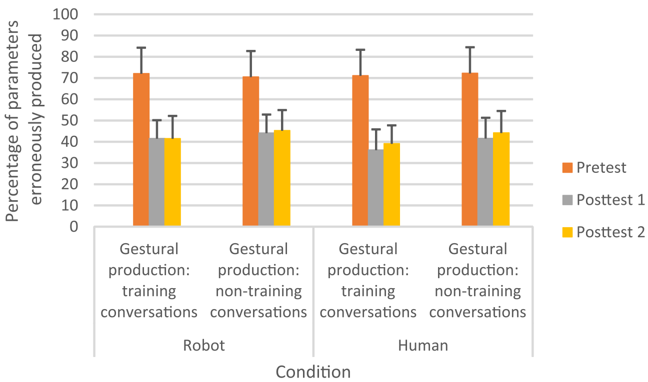

Fig. 4. The percentages of parameters erroneously produced in the gestural production pretests and immediate and delayed posttests of both training and non-training conversations in children with ASD in both conditions.

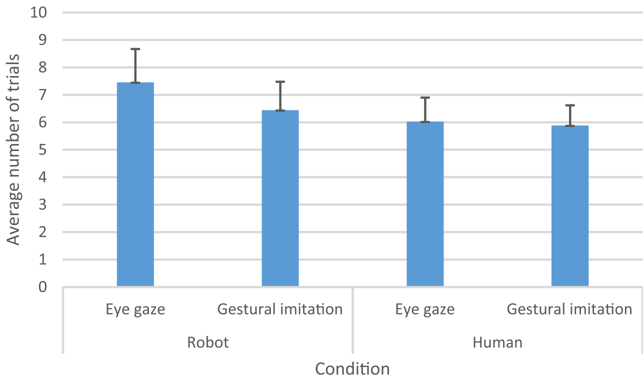

Fig. 5. Average number of trials with eye gaze and gestural imitation in both groups of children during training sessions.

4. Discussion

To summarize, children with ASD who received human-based gestural training were as likely to recognize gestures and produce these gestures accurately in the training conversations as those who received robot-based gestural training. Similar patterns were found in non-training conversations, suggesting that, for both groups of children, the acquired gestural production skills could be generalized to new situations. Additionally, the positive learning outcomes were maintained for at least two weeks for both groups of children when no training was provided. Taken together, there were no significant differences in the learning outcomes between the groups of children. Nevertheless, our results have shown that children in the robot-based intervention group were more likely to establish eye contact with their teachers (the social robots) than those in the human-based intervention group.

Our results suggested that both robot- and human-based intervention programs overall are effective in promoting gestural recognition and production skills in children with low-functioning autism. However, as shown in Appendix A, there were individual variations in the effectiveness of the interventions. For example, Subjects 6, 7, and 10 in the robot condition got a greater improvement in the gestural production posttests than Subjects 4 and 9 in the same condition. Further study should examine whether the individual variations in learning outcomes are attributable to their severity in autism, cognitive functioning and communication skills. One might also contend that children's performance in the gestural recognition and production tasks improved because the feedback given after each trial in the pretests helped them to remember correct and incorrect answers. If so, the improvement from the pretests to the posttests would be attributable to the feedback rather than to the training effects. We could not exclude this possibility as we did not introduce a wait-list control group, in which the children did not receive training but were assessed in the same ways as those in the robot- and human-based intervention. Having said that, it is unlikely that our positive learning outcomes were attributable to the feedback. Two of our previous intervention studies compared the pre- and post-test performance between the intervention and wait-list control groups and we provided feedback to both groups of the children during assessments (So, Wong, Lam, Cheng et al., 2018; So, Wong, Lam, Lam et al., 2018). The results have shown that, compared to the students in the wait-list control group who had not received training, students in the intervention group were more capable of recognizing and producing the taught gestures.

Human teachers seem to be as effective as robot teachers in teaching children with autism gestural recognition and production skills. Much research has shown that individuals with ASD respond to social robots more favorably than to human beings (Bird et al., 2007; Kim et al., 2013; Pierno et al., 2008; Tapus et al., 2012; Vanderborght et al., 2012). Additionally, they produced fewer stereotyped behaviors when interacting with a robot than when interacting with a human teacher (Michaud et al., 2007; Shamsuddin et al., 2013). However, none of these studies examined whether social robots can effectively teach social skills to children with ASD, when compared to human teachers. Only a few studies have addressed this issue indirectly (Wainer, Dautenhahn et al., 2014; Wainer, Robins et al., 2014). Our study was the first evaluating the learning outcomes of robot- and human-based interventions.

Our results may be explained by at least two reasons. First, children with autism might be simply "interested" in the social robots. According to the social motivation theory of autism, individuals with ASD show deficits in engaging with humans (Chevallier et al., 2012). In line with the previous findings (Tapus et al., 2012; Simut et al., 2016), our results have shown that children with ASD in the robot-based intervention group were more likely to establish eye contact with their teachers than those in the human-based intervention group, suggesting that these children found social robots more engaging than human teachers. However, simply engaging well with social robots might not necessarily lead to better learning outcomes.

Second, our human teachers who behaved naturally during training sessions taught the same content as the social robots. In every training session in both conditions, the human and robot teachers first greeted the child, followed by giving an instruction and demonstrating the conversations and gestures. On the initial viewing, the contents of gestural training in the human-based intervention were perceived as neither more nor less variable than those in the robot-based intervention. Two social robots or human demonstrators demonstrated the same gestures twice, in a training session. Besides, both groups of teachers provided the same instructions and prompts when requesting the children to imitate the gestures. Additionally, the number and duration of training sessions were the same in both intervention groups. Finally, human demonstrators were not allowed to modify the teaching content and instructions. We watched the videos of human- and robot-based intervention and undertook a manipulation check. The human demonstrators did follow the instructions in at least 90% of the training sessions (they sometimes added a few more words, which did not modify the teaching content and instructions). The robots were 100% compliant with the procedures. The comparable learning outcomes in the human- and robot-based interventions might be attributable to the fact that the training sessions in both conditions were highly structured.

According to the empathizing-systemizing theory (Baron-Cohen, 2009), a highly structured learning environment leads to positive learning outcomes in children with ASD. Under this view, children with ASD may learn effectively when they are placed in a highly structured learning environment in which the lessons are offered in a systematic and predictable manner (e.g., the new information is repeated a few times, each time sharing the same content; the procedures of the learning activities are consistent across lessons). Treatment and Education of Autistic and related Communication handicapped CHildren (TEACCH; Mesibov & Shea, 2010; Mesibov, Shea, & Schopler, 2005), in which structured teaching is a key component, is widely used in the education of children and adults with autism (Green et al., 2006).

When the lessons are highly structured, it does not matter whether social robots or human beings serve as teaching agents. Our human demonstrators spoke with each other with their usual intonation and facial expressions. In other words, the social cues found in the human-based intervention might not influence gestural learning. This finding is consistent with Chetcuti, Hudry, Grant, and Vivanti (2019) who found that imitative performance was similar when tasks were demonstrated by a socially responsive or aloof model for children with ASD. Taken together, children with ASD may not show differential sensitivity to the social contextual signals.

In conclusion, the present study designed comparable robot- and human-based intervention programs and found that the learning outcomes for both programs were similar. Our findings pave the way for conducting future research on suitable learning environments for children with ASD. They also shed light on the application of social robots in teaching children with ASD nonverbal and verbal communication skills

Funding

This research has been fully supported by grants to the WCSO from the Education Bureau (Project no. QEF6904338).

Acknowledgements

We acknowledge the help of our research assistants Fanny Lee and Angel Pang. Special thanks go to all the children and their parents for their help and dedication to education.

Appendix A

Condition	Subject	Chronological age	Proportion of trials with accurate gestural production						Proportin of trials with accurate gestural recognition			Standardized score in BOT	Proportion of accurate trials in ANT	Score in VP	Score in ToM
			Pretest		Posttest 1		Posttest 2		Pretest	Posttest 1	Posttest 2				
			S1	S2	S1	S2	S1	S2							
Robot	1	6.99	0.14	0.00	0.21	0.21	0.29	0.21	0.21	1.00	0.93	71	0.41	5	1
	2	9.41	0.07	0.07	0.14	0.07	0.21	0.21	0.21	0.50	0.57	63	0.13	7	0
	3	8.48	0.07	0.14	0.43	0.29	0.43	0.36	0.64	0.71	0.86	54	0.59	6	4
	4	10.39	0.07	0.14	0.21	0.21	0.14	0.07	0.50	0.86	0.86	71	0.52	6	2
	5	9.05	0.00	0.14	0.29	0.29	0.43	0.36	0.29	1.00	1.00	55	0.56	4	0
	6	9.23	0.21	0.21	0.79	0.93	0.86	0.86	0.64	1.00	1.00	64	0.97	10	1
	7	7.58	0.29	0.29	0.57	0.57	0.71	0.64	0.86	0.86	0.71	57	0.91	7	6
	8	8.41	0.07	0.07	0.36	0.29	0.36	0.43	1.00	0.64	0.50	70	0.97	5	4
	9	8.48	0.07	0.14	0.14	0.00	0.07	0.07	0.57	0.50	0.71	49	0.41	7	3
	10	10.39	0.07	0.29	0.86	0.79	0.64	0.57	1.00	0.71	0.50	70	0.75	8	5
	11	10.69	0.07	0.07	0.21	0.29	0.21	0.21	0.79	0.71	0.64	73	0.53	11	3
	12	9.89	0.07	0.00	0.21	0.21	0.21	0.07	0.64	0.93	1.00	66	0.31	6	2
Human	1	8.55	0.07	0.07	0.79	0.79	0.79	0.79	0.57	1.00	1.00	57	0.41	3	3
	2	6.83	0.29	0.21	0.79	0.64	0.79	0.79	0.79	0.36	0.50	74	0.91	14	5
	3	8.51	0.07	0.00	0.14	0.21	0.07	0.07	0.14	1.00	1.00	69	0.25	1	2
	4	7.76	0.14	0.14	0.36	0.36	0.29	0.21	0.07	0.07	0.00	49	0.5	8	1
	5	8.57	0.21	0.14	0.79	0.50	0.57	0.36	0.71	0.50	0.57	70	0.22	6	3
	6	10.33	0.29	0.29	0.71	0.57	0.86	0.64	0.93	0.93	1.00	66	0.66	7	3
	7	10.74	0.07	0.00	0.07	0.07	0.07	0.07	0.07	0.93	1.00	70	0.25	6	0
	8	7.28	0.14	0.07	0.71	0.79	0.71	0.79	0.29	1.00	1.00	63	0.94	8	2
	9	8.72	0.07	0.07	0.14	0.07	0.07	0.07	0.14	0.71	0.79	51	0.13	4	0
	10	7.90	0.07	0.21	0.14	0.14	0.36	0.21	0.36	1.00	1.00	67	0.56	7	3
	11	7.86	0.14	0.14	0.36	0.50	0.57	0.57	0.86	0.50	0.14	75	0.72	5	5

References

American Psychiatric Association (2010). *Diagnostic and statistical manual of mental disorders (DSM-4)* (4th ed.). United States: American Psychiatric Publishing.

Baron-Cohen, S. (2009). Autism: The empathizing–systemizing (E-S) theory. *Annuals of the New York Academy of Sciences, 1156*(1), 68–80.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lm: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), e4.

Beery, K. E., & Beery, N. A. (2004). *The Beery-Buktenica developmental test of visual-motor integration: Administration, scoring, and teaching manual* (5th ed.). Minneapolis, MN: NCS Pearson.

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism, 18*, 598–608.

Bird, G., Leighton, J., Press, C., & Heyes, C. M. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorder. *Proceedings of the Royal Society of London. Series B, 274*, 3027–3031.

Bölte, S. (2009). *Autismus. Spektrum, ursachen, diagnostik, intervention, perspektiven.* Bern: Huber, Hogrefe AG.

Bruininks, R., & Bruininks, B. (2005). *Bruininks-Oseretsky test of motor proficiency* (2nd ed.). Minneapolis, MN: NCS Pearson.

Cabibihan, J.-J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey of the roles and benefits of social robots in therapy of children with autism. *International Journal of Social Robotics, 5*, 593–618.

Cabibihan, J.-J., So, W. C., & Pramanik, S. (2012). Human-recognizable robotic gestures. *IEEE Transactions on Autonomous Mental Development, 4*(4), 305–314.

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Rizzolatti, G., ... Cossu, G. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proceedings of the National Academy of Sciences of the United States of America, 104*(45), 17825–17830.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*(1), 213–236.

Chetcuti, L., Hudry, K., Grant, M., & Vivanti, G. (2019). Object-directed imitation in autism spectrum disorder is differentially influenced by motoric task complexity, but not social contextual cues. *Autism, 23*(1), 199–211.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory. *Trends in Cognitive Sciences, 16*, 231–239.

Curcio, F., & Paccia, J. (1987). Conversations with autistic children: Contingent relationships between features of adult input and children's response adequacy. *Journal of Autism and Developmental Disorders, 17*(1), 81–93.

Dewey, D., Cantell, M., & Crawford, S. G. (2007). Motor and gestural performance in children with autism spectrum disorder, developmental coordination disorder, and/or attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society, 13*(2), 246–256.

Dowell, L. R., Mahone, E. M., & Mostofsky, S. H. (2009). Associations of postural knowledge and basic motor skill with dyspraxia in autism: Implication for abnormalities in distributed connectivity and motor learning. *Neuropsychology, 23*(5), 563–570.

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 Robot, 24*, 147–157.

Dziuk, M. A., Larson, J. C. G., Apostu, A., Mahone, E. M., Denckla, M. B., & Mostofsky, S. H. (2007). Dyspraxia in autism: Association with motor, social, and communicative deficits. *Developmental Medicine and Child Neurology, 49*, 734–739.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*, 143–166.

Frea, W., Arnold, C., & Vittimberga, G. (2001). A demonstration of the effects of augmentative communication on the extreme aggressive behavior of a child with autism within an integrated preschool setting. *Journal of Positive Behavior Intervention, 3*, 194–198.

Gizzonio, V., Avanzini, P., Campi, C., Orivoli, S., Piccolo, B., Cantalupo, G., ... Fabbri-Destro, M. (2015). Failure in pantomime action execution correlates with the severity of social behavior deficits in children with autism: A praxis study. *Journal of Autism and Developmental Disorders, 45*(10), 3085–3097.

Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities, 27*(1), 70–84.

Ham, H. S., Bartolo, A., Corley, M., Rajendran, G., Szabo, A., & Swanson, S. (2011). Exploring the relationship between gestural recognition and imitation: Evidence of dyspraxia in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*(1), 1–12.

Ham, H. S., Bartolo, A., Corley, M., Swanson, S., & Rajendran, G. (2010). Case report: Selective deficit in the production of intransitive gestures in an individual with autism. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 46*(3), 407–409.

Hart, S., & Banda, D. (2009). Picture exchange communication system with individuals with developmental disabilities: A meta-analysis of single subject studies. *Remedial and Special Education, 31*(6), 476–488.

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

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., ... Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders, 43*(5), 1038–1049.

Klin, A., & Jones, W. (2006). Attributing social and physical meaning to ambiguous visual displays in individuals with higher-functioning autism spectrum disorder. *Brain and Cognition, 61*(1), 40–53.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature, 459*(7244), 257–261.

Koegel, L., Koegel, R., Shoshan, Y., & McNerney, E. (1999). Pivotal response intervention II: Preliminary long-term outcome data. *Journal of the Association for Persons with Severe Handicaps, 24*, 186–198.

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics, 1*(1), 3–18.

Lee, J., & Obinata, G. (2015). Interactive educational material for children with ASD. *26th 2015 international symposium on micro-nanomechatronics and human science (MHS2015)*.

Lee, J., Takehashi, H., Nagai, C., Obinata, G., & Stefanov, D. (2012). Which robot features can stimulate better responses from children with autism in robot-assisted therapy? *International Journal of Advanced Robotic Systems, 9*, 1–6.

Li, H., Cabibihan, J.-J., & Tan, Y. (2011). Towards an effective design of social robots. *International Journal of Social Robotics, 3*(4), 333–335.

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorder using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*(3), 623–654.

Martin, N. (2006). *Test of visual perceptual skills* (3rd ed.). USA: Academic Therapy Publications.

Mastrogiuseppe, M., Capirci, O., Cuva, S., & Venuti, P. (2016). Gestural communication in children with autism spectrum disorder during mother-child interaction. *Autism, 19*(4), 469–481.

Mesibov, G. B., & Shea, V. (2010). The TEACCH program in the era of evidence-based practice. *Journal of Autism and Developmental Disorders, 40*(5), 570–579.

Mesibov, G. B., Shea, V., & Schopler, E. (2005). *The TEACCH approach to autism spectrum disorder.* New York: Springer.

Michaud, F., Salter, T., Duquette, A., Mercier, H., Lauria, M., Larouche, H., ... Larose, F. (2007). *Assistive technologies and children-robot interaction. AAAI spring symposium: Multidisciplinary collaboration for socially assistive robotics*45–49.

Mirenda, P. (2003). Toward functional augmentative and alternative communication for students with autism manual signs, graphic symbols, and voice output communication aids. *Language, Speech, and Hearing Services in Schools, 34*(3), 203–216.

Mitchell, P., Parsons, S., & Leonard, A. (2007). Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 589–600.

Miyamoto, E., Lee, M., Fujii, H., & Okada, M. (2005). *How can robots facilitate social interaction of children with autism? Possible implications for educational environments.*

*Proceedings of the 5th international workshop on epigenetic robotics: Modeling cognitive development in robotic systems*145–146.
Pierno, A., Mari, M., Lusher, D., & Castiello, U. (2008). Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia, 46*(2), 448–454.
Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(2), 301–322.
Pop, C., Pintea, S., Vanderborght, B., & David, D. D. (2014). Enhancing play skills, engagement and social skills in a play task in ASD children by using robot-based interventions. A pilot study. *Interaction Studies, 15*, 292–320.
Ricks, D. J., & Colton, M. B. (2010). *Trends and considerations in robot-assisted autism therapy. IEEE international conference on robotics and automation (ICRA)*4354–4359.
Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., ... Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*, 1029–1040.
Scassellati, B., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering, 14*, 275–294.
Shamsuddin, S., Yussof, H., Miskam, M. A., Hamid, A. C., Malik, N. A., & Hashim, H. (2013). Humanoid robot NAO as HRI mediator to teach emotions using game-centered approach for children with autism. *HRI 2013 workshop on applications for emotional robots*.
Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F., Mohamed, S., Piah, H., ... Zahari, N. (2012). *Initial response of autistic children in human-robot interaction therapy with humanoid robot Nao. Signal processing and Its applications (CSPA), 2012 IEEE 8th international colloquium*188–193.
Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., & Zahari, N. I. (2012). Humanoid robot NAO interacting with autistic children of moderately impaired intelligence to augment communication skills. *Procedia Engineering, 41*, 1533–1538.
Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., & Zahari, N. I. (2012). Initial response in HRI-a case study on evaluation of child with autism spectrum disorder interacting with a humanoid robot Nao. *Procedia Engineering, 41*, 1448–1455.
Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or asperger syndrome to recognize and predict emotions in others. *Autism: The International Journal of Research and Practice, 5*(3), 299–316.
Simut, R. E., Vanderfaeillie, J., Peca, A., Van de Perre, G., & Vanderborght, B. (2016). Children with autism spectrum disorder make a fruit salad with probo, the social robot: An interaction study. *Journal of Autism and Developmental Disorders, 46*(1), 113–126.
So, W., Wong, M., Lui, M., & Yip, V. (2015). The development of co-speech gesture and its semantic integration with speech in 6- to 12-year-old children with Autism spectrum disorder. *Autism: International Journal of Research and Practice, 19*(8), 956–968.
So, W. C., Wong, M. K. Y., Cabibihan, J.-J., Lam, K.-Y., Chan, Y.-Y., & Qian, H.-H. (2016). Using robot animation to promote gestural skills in children with autism spectrum disorder. *Journal of Computer Assisted Learning, 32*, 632–646.
So, W. C., Wong, M. K. Y., Lam, W. Y., Cheng, C. H., Yang, J. H., Huang, Y., ... Lee, C. C. (2018). Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder. *Molecular Autism, 9*, 34.
So, W. C., Wong, M. K. Y., Lam, K. Y., Lam, W. Y., Chui, T. F., Lee, T. L., ... Fok, C. W. (2018). Using a social robot to teach gestural recognition and production in children with autism spectrum disorder. *Disability and Rehabilitation: Assistive Technology, 13*(6), 527–539.
Stokoe, W. C. (2005). Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of Deaf Studies and Deaf Education, 10*(1), 3–37.
Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology, 25*(6), 475–485.
Tapus, A., Peca, A., Aly, A., Pop, C. A., Jisa, L., Pintea, S., ... David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot–A series of single case experiments. *Interaction Studies, 13*(3), 315–347.
Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., ... David, D. O. (2012). Using the social robot probo as a social story telling agent for children with ASD. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 13*, 348–372.
Volkmar, F. R., & Wiesner, L. A. (2009). *A practical guide to autism: What every parent, family member, and teacher needs to know.* Hoboken, N.J: John Wiley & Sons.
Wainer, J., Ferrari, E., Dautenhahn, K., & Robins, B. (2010). The effectiveness of using a robotics class to foster collaboration among groups of children with autism in an exploratory study. *Personal and Ubiquitous Computing, 14*(5), 445–455.
Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2014). A pilot study with a novel setup for collaborative play of the humanoid robot KASPAR with children with autism. *International Journal of Social Robotics, 6*(1), 45–65.
Wainer, J., Robins, B., Amirabdollahian, F., & Dautenhahn, K. (2014). Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development, 6*(3), 183–199.
Wellman, H. M., & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development, 75*, 523–541.
Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science, 17*(12), 1075–1081.
Werry, I., Dautenhahn, K., Ogden, B., & Harwin, W. (2001). Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy. *Lecture Notes in Computer Science, CT, 2001*, 57–74.
Wetherby, A. M., Prizant, B. M., & Hutchinson, T. A. (1998). Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders. *American Journal of Speech-Language Pathology, 7*, 79–91.
Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorder in the second year of life. *Journal of Autism and Developmental Disorders, 34*, 473–493.