

Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention

BIBI HUSKENS¹, RIANNE VERSCHUUR¹, JAN GILLESSEN², ROBERT DIDDEN³, & EMILIA BARAKOVA²

¹Research and Development, Dr Leo Kannerhuis, Doorwerth, The Netherlands, ²Faculty of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands, and ³Behavioral Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

(Received 9 October 2012; accepted 9 October 2012)

Abstract

Objective: The purpose of the present study was to investigate the effectiveness of an applied behaviour analysis (ABA)-based intervention conducted by a robot compared to an ABA-based intervention conducted by a human trainer in promoting self-initiated questions in children with autism spectrum disorder (ASD).

Methods: Data were collected in a combined crossover multiple baseline design across participants. Six children were randomly assigned to two experimental groups.

Results: Results revealed that the number of self-initiated questions for both experimental groups increased between baseline and the first intervention and was maintained during follow-up. The high number of self-initiated questions during follow-up indicates that both groups maintained this skill.

Conclusions: The interventions conducted by a robot and a human trainer were both effective in promoting self-initiated questions in children with ASD. No conclusion with regard to the differential effectiveness of both interventions could be drawn. Implications of the results and directions for future research are discussed.

Keywords: Autism spectrum disorders, robot intervention, high-functioning children

Introduction

Children with autism spectrum disorder (ASD) often have deficits in behaviours concerning the so-called pivotal areas [1]. Koegel et al. [1] conceptualize pivotal areas as areas that, when targeted during an intervention, also lead to improvements in other behaviours. So far, research has focused on five pivotal areas: responding to multiple cues, motivation, self-management, empathy, and self-initiations during social interactions [1–4].

Self-initiations include, for example, initiations in (meaningful) first words, question asking, child-initiated comments and attention-seeking initiations [5]. Several studies have shown that children with ASD often lack or show deficits in self-initiations [2, 6, 7]. However, the ability to engage in self-initiations predicts favourable long-term effects on several other behaviours [1], which could be explained by the fact that self-initiations result in

self-learning [1] and elicit teaching interactions from the child's environment [8]. Research has shown that self-initiations positively influence the child's level of expressive and receptive language [9, 10], may increase social communication abilities [2] and lead to increased participation in inclusive education and extra-curricular activities and positive interactions with peers [5]. Therefore, it is important that children with ASD are taught how to engage in self-initiations.

Several studies targeting self-initiated question asking have shown that children with ASD are able to engage in self-initiated question asking [10], and that self-initiated question asking can be generalized to the home situation [6] or the question taught can be generalized into other question forms [8].

Recently, more attention is being given towards developing and validating technology in interventions for children with ASD [11–13], for example, by using speech generating devices during communication interventions [14], computer games [15], virtual

reality [16], and robots [13, 17, 18]. There are several reasons for the use of robots in interventions for children with ASD. First, robots are engaging and motivate children with ASD to interact [18]. Technology is considered intrinsically appealing to children with ASD [19]. Second, robot behaviour is consistent and highly predictable, which is preferred by children with ASD [20, 21]. Third, because children with ASD have difficulties to pay attention to multiple cues during social interactions [1], social interactions with other people can be complex and difficult to understand. A robot is easily able to perform less complex interaction behaviour (i.e., use fewer cues), isolate interaction skills and adapt the level of complexity of its behaviour [12, 13, 17–20]. Finally, because a robot is able to create a predictable and simple social interaction situation, it reduces the stress and pressure that children with ASD often experience during interactions with humans, which leads to a more enjoyable and effective learning environment [12, 13].

Several studies show promising results of the use of robots in interventions for children with ASD. For example, Dautenhahn and Billard [17] observed that eight children with ASD played imitations games with a humanoid robot and seemed to enjoy the sessions. Feil-Seifer and Matarić [22] found that contingent behaviour of a non-humanoid robot led to an increase in self-initiated social interactions in three children with ASD, both with the robot and with a parent. Duquette et al. [23] investigated the use of a mobile robot to facilitate social interaction behaviours (e.g., shared focused attention and shared conventions) in four 5-year old children with low-functioning autism. Two children were pre-verbal, the other two were non-verbal. The children were assigned to two intervention groups: one group paired with the mobile robot and the other with a human mediator. The authors did not find a significant increase in the social interaction behaviours, but they found that the children paired with the robot showed more shared focused attention than the children paired with the human mediator and engaged in less repetitive or stereotyped behaviour. However, they also showed less imitative behaviour (shared conventions) than the children paired with the human mediator. Robins et al. [24] examined the utility of a humanoid robot to elicit social interaction behaviour in four children with ASD. No group analyzes were performed. Each child showed an increase in the overall duration of social interactions with the robot, but only two children showed an increase in the duration of the measured behaviours (eye gaze, touch, imitation, and proximity). Stanton et al. [25] examined whether a robotic pet elicited social behaviours (e.g., verbal engagement, affection, and reciprocal

and authentic interaction) in 11 children with ASD, aged between 5 and 8 years. The results failed to reach statistical significance, but suggested that the children showed more verbal engagement and reciprocal and authentic interaction with the robotic dog. To conclude, robots might be beneficial in promoting social interaction skills in children with ASD.

Although above studies offer some insight into the effects of robots on social interaction in children with ASD, many questions with regard to the role of robots in autism interventions remain. Diehl et al. [19] reviewed several studies on the application of robots in interventions for children with ASD and concluded that, although the results of the studies are promising, it is difficult to draw firm conclusions based on these studies, because most studies are only exploratory and have methodological limitations. Therefore, Diehl et al. stress the need for empirical studies, such as applied behaviour analysis (ABA), that compare the effectiveness of interventions conducted by robots with other interventions that target social interaction in children with ASD.

The purpose of the present study was to investigate the effectiveness of an ABA-based intervention conducted by a robot compared to an intervention conducted by a human trainer in promoting question-asking in children with ASD. The two research questions were: (a) are an intervention conducted by a robot and an intervention conducted by a human trainer effective in promoting question-asking in children with ASD? and (b) which of these two interventions is more effective in promoting question-asking in children with ASD?

Method

Participants

Six children, who attended a day treatment group or a clinical treatment facility of the Dr Leo Kannerhuis, were selected to participate in this study. The children of the day treatment facility attended this facility for 2 days a week. The children had to meet the following inclusion criteria: (a) a diagnosis of ASD according to the DSM-IV TR [26] criteria, (b) between 8 and 14 years of age, (c) a full-scale IQ above 80 on a standardized intelligence test, (d) not being able to initiate a question after a statement of a trainer, (e) being able to participate during the whole study, and (f) not being anxious about the robot. Informed consent was obtained from the parents of each child. Characteristics of the included children are displayed in Table I.

Table I shows that the ages of the children ranged from 8 years, 9 months to 12 years, 3 months with a mean of 11 years and 5 months. All children were boys. For all children the ASD diagnosis was

Table I. Participant characteristics.

Child	Age ^a	Gender	Diagnosis	SCQ-score ^b	IQ
Andy	10;10	Male	ASD, ADHD-NOS	24	85 (V) 107 (P)
Bryan	11;5	Male	PDD-NOS, subtype McDD, ADHD	18	94
Colin	11;10	Male	PDD-NOS	31	89
Dean	11;5	Male	ASD, Erb's palsy	28	111
Edward	12;3	Male	PDD-NOS	26	98
Finn	8;9	Male	Autism	32	–

Notes: ^aYear; month.

^bScore on the SCQ. A score > 15 is an indication for ASD.

confirmed by their score on the Social Communication Questionnaire (SCQ [27]; Dutch translation [28]). No interpretable full-scale IQ was available for Finn.

Setting and materials

For the children of the day treatment group, sessions were conducted in a room of the day treatment facility. For the children of the clinical treatment facility, sessions were conducted in a room at the children's school. During sessions, the child and the trainer sat at a table.

Several materials were used during sessions. A digital video camera was used to record all sessions. During baseline, intervention with the human trainer and follow-up, several play materials (e.g., LEGO®-blocks, toy cars, and plastic toy animals), pictures (e.g., pictures of animals and cars) and colouring pictures were used. The children were familiar with those materials.

During intervention sessions with the robot, a NAO robot from Aldebaran Robotics was used. This neutral-coloured robot is 57 cm tall, has 25 mechanical degrees of freedom and is equipped with digital cameras, microphones, speakers and touch sensors. The NAO robot is a humanoid robot; however, it has a "simple" face to prevent confusion or over-stimulation in the children. For the voice of the robot a female pre-recorded voice was used. This was done for two reasons. First, it appears that a pre-recorded human voice is preferred over synthetic speech [29]. Second, a female pre-recorded voice was used to reduce the difference between the robot intervention and the trainer intervention, which was conducted by a female trainer. A remote control designed for the present study was used to control the robot. A visual programming environment, called TiViPE [30], was used for programming the scenarios for the robot [11].

Design

A combined crossover multiple baseline design across participants was used to investigate the effectiveness

of an intervention conducted by a robot compared with an intervention conducted by a human trainer in promoting self-initiated questions in children with autism. The six children were randomly assigned to two intervention (experimental) groups. The mean age was 11; 4 years for experimental group 1 and 10; 10 years for experimental group 2. The full-scale IQ was 94 for experimental group 1 and 104 for experimental group 2. For experimental group 1, the mean SCQ-score was 24 and for experimental group 2, this score was 29.

Procedures

Introduction of the robot. Prior to baseline, the robot was introduced to all children who were part of the day treatment group for children and the clinical treatment facility. The introduction of the robot was thus not limited to the children who participated in the present study. The introduction took place during two sessions to ensure that all children were familiar with the robot. During the first session, the children met the robot in their regular group. The robot told the children that it felt sad, because it did not have a name. The children were then asked to think about a name for the robot together with the staff members. During this session, the staff members made clear that the children are only allowed to look at the robot. During the second session, the children met the robot again. This time the robot told the children its name and the children were thanked for their help.

Baseline 1. Baseline conditions consisted of three to five individual 10-min sessions. The children were trained by two child psychologists and one child psychologist in training. At the beginning of each session, the trainer told the child that he will practice question asking. The trainer then provided the child with four opportunities to ask a question by making a statement (e.g., "I have done something nice yesterday"). The child was not prompted to respond, but if he asked a correct question (i.e., a syntactically correct question), the trainer responded accordingly. If the child asked a question that implied an action,

the trainer carried out this action. If the child asked another correct question, the trainer verbally responded to this question. If the child did not respond within 7 s, the trainer said “Maybe some other time” and made the next statement, or ended the session if four statements were made.

Intervention phases. All six children received two interventions: one intervention conducted by the robot and one intervention conducted by the human trainer. These interventions were conducted in a different order for both groups of children. Each intervention consisted of four individual sessions and each intervention session took 10 min.

At the beginning of each intervention session with the robot, the trainer told the child that the robot will practice question asking with him. The robot was standing on the table within the child’s sight, but out of his reach. During each session, the robot provided the child with four opportunities to engage in asking questions by following a statement-question-action sequence. During the first statement-question-action sequence, the robot first greeted the child and attracted the child’s attention by saying “I want to tell you something.” During the other statement-question-action sequences, the robot only attracted the child’s attention. The robot then made a statement (e.g., “I can do a cool dance”). Next, the child was provided with an opportunity to engage in self-initiation by asking a question (e.g., “Could I see your dance?”). If the child asked a correct question (i.e., a syntactically correct question) within 7 s that implied carrying out an action, the robot subsequently performed its action (e.g., dancing). This was counted as a self-initiated question. When the robot had performed its action, it said “How nice that you’ve asked me something. This was it!” to complete the statement-question-action sequence during the first, second, and third statement-question-action sequence. During the last statement-question-action sequence, the robot said “How nice that you’ve asked me something. This was it! Till next time!”

The child could also ask a correct question that did not imply carrying out an action (e.g., “What kind of dance?”). The trainer paused the robot in that case and verbally answered the child’s question by partly repeating the robot’s statement (e.g., “A cool dance”). The trainer then provided the child with another 7 s to engage in a self-initiation by asking a question. If the child then asked a correct question that implied carrying out an action, the robot performed its action. If the child did not respond correctly or asked another question that did not imply carrying out an action, he was prompted (see below).

It was also possible that the child indicated that he is not interested in the robot’s action (e.g., by saying

“I don’t want to see it”). This was counted as an incorrect response. If the child indicated that he is not interested, the robot did not perform his action, but it responded by saying “It is good that you said that. Maybe some other time.” The statement-question-action sequence ended here.

If the child did not ask a syntactically correct question that implied carrying out an action within 7 s, he was prompted to respond following a least to most prompt hierarchy (increasing assistance). First, the robot used an open-question-prompt in which the robot asked the child “What could you ask me now?” If the child did not respond correctly within 7 s (i.e., no response, an incorrect response or by asking another question that does not implicate an action), a waiting-prompt was used, in which the robot made a “waiting gesture.” If the child did not respond correctly within 7 s, a tell-prompt was used, in which the robot told the child how he could respond by saying “You could ask me if you can see my dance.” If the child did not respond correctly within 7 s, a fill-in-prompt was used, in which the robot said a part of the correct response by saying “Can I see...”, for example. If the child did not respond correctly within 7 s, the robot said “Maybe some other time.” The state-question-action sequence ended here. However, if the child responded correctly (i.e., by asking a question that implicates an action) to one of these prompts, the robot performed his action (e.g., by dancing). When the robot has performed its action, it said “How nice that you’ve asked me this. Well, this was it! Bye!” to end the statement-question-action sequence. The procedures for the robot intervention and the prompt hierarchy are provided in Figure 1.

During the intervention sessions with the human trainer, the trainer told the child that he is going to practice question asking. The trainer then provided the children with four opportunities to engage in self-initiated questions by following exactly the same procedures (including the use of increasing assistance) as during the intervention with the robot.

Baseline 2. The second baseline consisted of four individual sessions. The procedures for baseline 2 were identical to the procedures for baseline 1.

Follow-up. Follow-up data were collected two weeks after the last intervention session. The follow-up consisted of four sessions. The procedures for the follow-up sessions were identical to the procedures for the first baseline.

Data collection

All 10-min baseline, intervention and follow-up sessions were recorded using a video camera. The

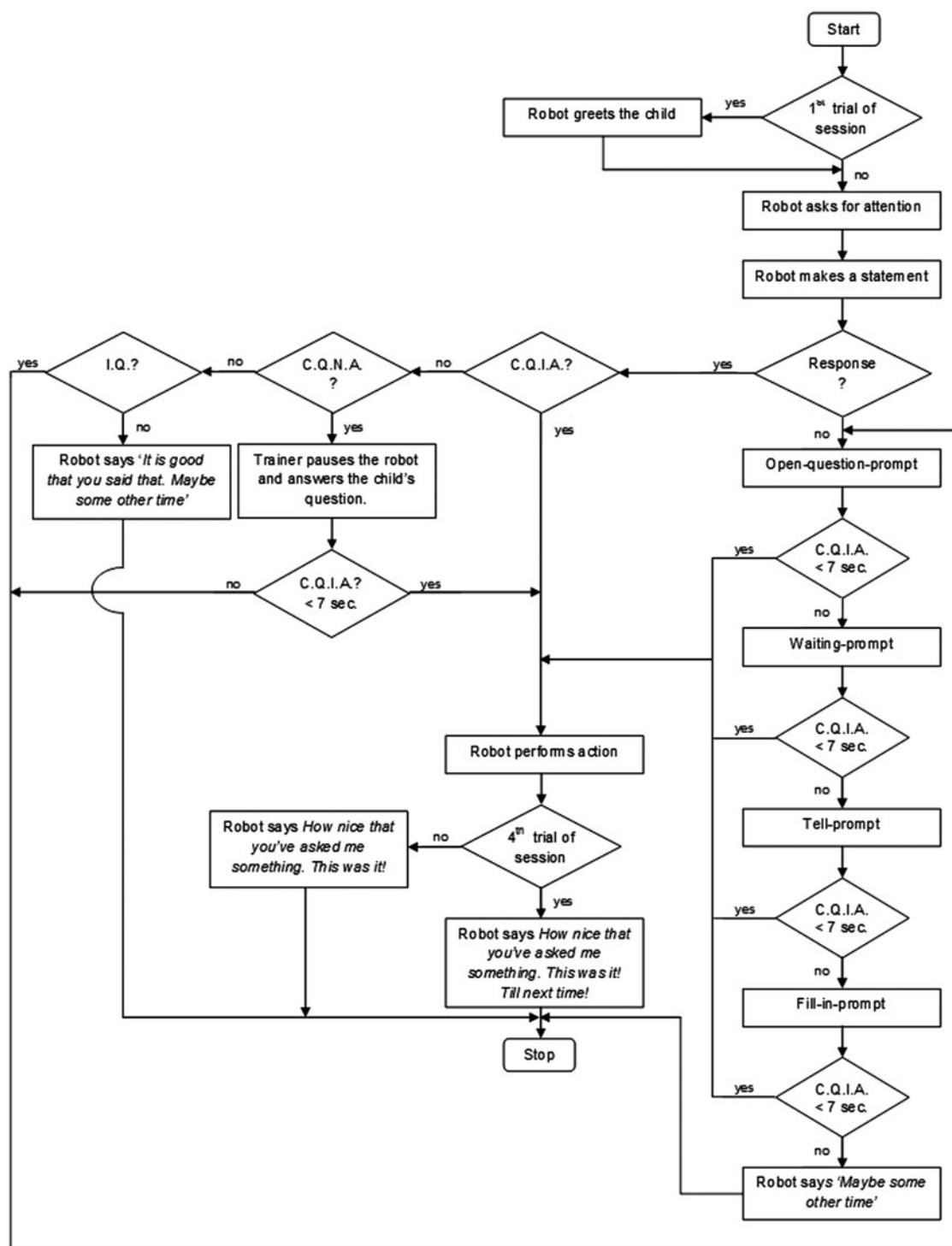


Figure 1. Flowchart of the intervention procedure and prompt hierarchy.

Note: I.Q. = incorrect question, C.Q.N.A. = correct question not implicating action, C.Q.I.A. = correct question implicating action.

videos were observed in a randomized order when data-collection was completed.

The videos of the sessions were recorded by the researcher using an event-recording system [31]. The following behaviours were recorded: (a) the robot/trainer makes a statement, (b) the child asks a syntactically correct question that implies carrying

out an action (i.e., a self-initiated question), (c) the human trainer reinforces the self-initiated question by carrying out the action, (d) the robot reinforces the self-initiated question by carrying out the action, (e) the child asks a syntactically correct question that does not imply carrying out an action, (f) the human trainer verbally answers the syntactically correct

Table II. Kappa coefficients.

Behaviour	Mean	SD
Statement	1.00	0.00
Syntactically correct question that implies an action	0.98	0.06
Reinforcement by trainer (action)	1.00	0.03
Reinforcement by robot (action)	0.97	0.16
Syntactically correct question that does not imply an action	0.87	0.30
Verbal reinforcement by trainer	0.97	0.16
Incorrect response	0.87	0.27
Open-question-prompt	1.00	0.00
Waiting-prompt	1.00	0.00
Tell-prompt	1.00	0.00
Fill-in-prompt	1.00	0.00

Table III. PABAK coefficients.

Behaviour	Mean	SD
Statement	1.00	0.00
Syntactically correct question that implies an action	0.99	0.05
Reinforcement by trainer (action)	1.00	0.02
Reinforcement by robot (action)	1.00	0.02
Syntactically correct question that does not imply an action	0.98	0.05
Verbal reinforcement by trainer	1.00	0.02
Incorrect response	0.97	0.06
Open-question-prompt	1.00	0.00
Waiting-prompt	1.00	0.00
Tell-prompt	1.00	0.00
Fill-in-prompt	1.00	0.00

question that does not imply carrying out an action, (g) the child does not respond or responds incorrectly, (h) an open-question-prompt is used, (i) a waiting-prompt is used, (j) a tell-prompt is used, and (k) a fill-in-prompt is used.

Inter-observer agreement and treatment integrity

Inter-observer agreement was calculated for 33% of the sessions (approximately evenly distributed across children and phases) and was determined by comparing the data of both observers on a behaviour-by-behaviour basis. Kappa coefficient [32] and prevalence-adjusted and bias-adjusted kappa (PABAK) [33] were calculated to determine inter-observer agreement. Kappa coefficient across children and behaviours varied between 0.78 and 1.00, with a mean of 0.97 ($SD = 0.05$). PABAK across children and behaviours varied between 0.94 and 1.00, with a mean of 0.99 ($SD = 0.02$), indicating excellent inter-observer agreement [34]. Kappa coefficients and

PABAK for the separate behaviour categories are presented in Tables II and III.

Data on treatment integrity were collected to check if the procedures were implemented correctly. Treatment integrity was measured during 33% of the sessions for each child. Treatment integrity varied between 88% and 100%, with a mean of 99% ($SD = 3.14$) across children and phases. Level of treatment integrity did not vary between robot intervention and trainer intervention. A second observer collected data on treatment integrity during 10% of the sessions that were observed for treatment integrity. Kappa coefficient for inter-observer agreement on treatment integrity varied between 0.67 and 1.00, with a mean of 0.86 ($SD = 0.14$).

Data-analysis

Data-analysis first involved visual inspection and the calculation of the mean number of self-initiated questions for both experimental groups during each phase with the corresponding standard deviations. Second, the increase in the number of self-initiations during phases was determined by calculating $\text{Tau}_{\text{novlap}}$ [35]. $\text{Tau}_{\text{novlap}}$ is an effect size for single-case research that examines data non-overlap between phases. $\text{Tau}_{\text{novlap}}$ indicates the proportion of all data pairs that improve over time [36]. $\text{Tau}_{\text{novlap}}$ and the corresponding standard deviation and p -value were calculated for each phase-contrast for each participant. Using WinPepi software [37], the overall $\text{Tau}_{\text{novlap}}$ and the corresponding standard deviation and p -value was calculated.

Results

Experimental group 1: Robot-trainer sequence

Figure 2 shows the number of self-initiated questions by Andy, Bryan, and Colin across phases (i.e., the number of syntactically correct questions that implied carrying out an action). Table IV presents the mean number of self-initiated questions during each phase and the corresponding standard deviations. During baseline 1, the number of self-initiated questions (out of a maximum of 4) was relatively low ($M = 1.17$, $SD = 1.40$). Colin, however, showed an increase in the number of self-initiated questions during the first baseline. The number of self-initiated questions increased during the robot intervention to a mean of 3.92 ($SD = 0.29$). Visual inspection shows no overlap in data points between baseline 1 and the robot intervention. The values of $\text{Tau}_{\text{novlap}}$ for the individual children are provided in Table V, showing a significant increase for all three children ($p < 0.02$). The corresponding overall value of

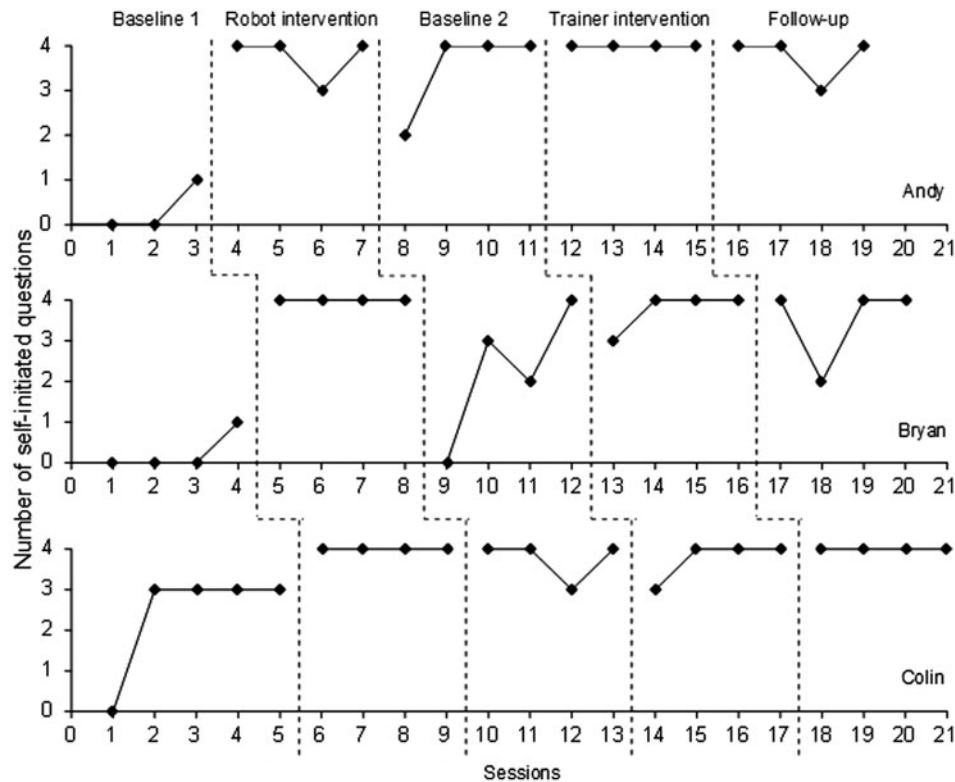


Figure 2 Number of self-initiated questions experimental group 1.

Table IV. Mean number of self-initiated questions during each phase and corresponding standard deviations for experimental group 1.

Phase	Mean	SD
Baseline 1	1.17	1.403
Robot intervention	3.92	0.289
Baseline 2	3.17	1.267
Trainer intervention	3.83	0.389
Follow-up	3.75	0.622

$\text{Tau}_{\text{novlap}}$ is 1.00 (SE = 0.17; 90% confidence interval: 0.73–1.00). As shown in Figure 2 and Table IV, the number of self-initiated questions remained at a high level during the second baseline for all three children ($M = 3.17$, $SD = 1.27$). The mean number of self-initiated questions did not increase during the trainer intervention ($M = 3.83$; $SD = 0.39$). During follow-up, the number of self-initiated questions remained at a high level ($M = 3.75$, $SD = 0.62$), showing that the self-initiations were maintained.

As shown in Figure 2 and Table IV, the number of self-initiated questions increased between baseline 1 and follow-up. The corresponding overall value of $\text{Tau}_{\text{novlap}}$ is 1.00 (SE = 0.24; 90% confidence interval: 0.61–1.00), indicating that the number of self-initiated questions increased during the

interventions conducted by a robot and a human trainer. The values of $\text{Tau}_{\text{novlap}}$ for the individual children are provided in Table V, showing a significant increase for all three children ($p < 0.02$).

Experimental group 2: trainer-robot sequence

Figure 3 shows the number of self-initiated questions by Dean, Edward, and Finn across phases. Table VI presents the mean number of self-initiated questions during each phase and the corresponding standard deviations. The number of self-initiated questions was generally low during baseline 1 ($M = 0.25$, $SD = 0.62$). However, Dean's data show a positive trend during baseline 1. The number of self-initiated questions increased during the trainer intervention ($M = 3.92$, $SD = 0.29$). Visual inspection shows no overlap in data points between baseline 1 and the trainer intervention. The values of $\text{Tau}_{\text{novlap}}$ for the individual children are provided in Table VII, indicating a significant increase for all three children ($p < 0.02$). The corresponding overall value of $\text{Tau}_{\text{novlap}}$ is 1.00 (SE = 0.11; 90% confidence interval: 0.81–1.00). Figure 3 and Table VII both show that the number of self-initiated questions generally remained at a high level during baseline 2 ($M = 3.33$, $SD = 1.23$). Therefore, the number of self-initiated questions could hardly increase during the robot intervention ($M = 3.75$, $SD = 0.62$). During follow-

Table V. Individual values of $\text{Tau}_{\text{novlap}}$ for the children of experimental group 1.

Child	Baseline1 – follow-up contrast			Baseline 1 – robot intervention contrast		
	$\text{Tau}_{\text{novlap}}$	SD	p	$\text{Tau}_{\text{novlap}}$	SD	p
Andy	1.00	0.45	0.02*	1.00	0.45	0.02*
Bryan	1.00	0.41	0.01**	1.00	0.40	0.01**
Collin	1.00	0.37	0.01**	1.00	0.37	0.01**

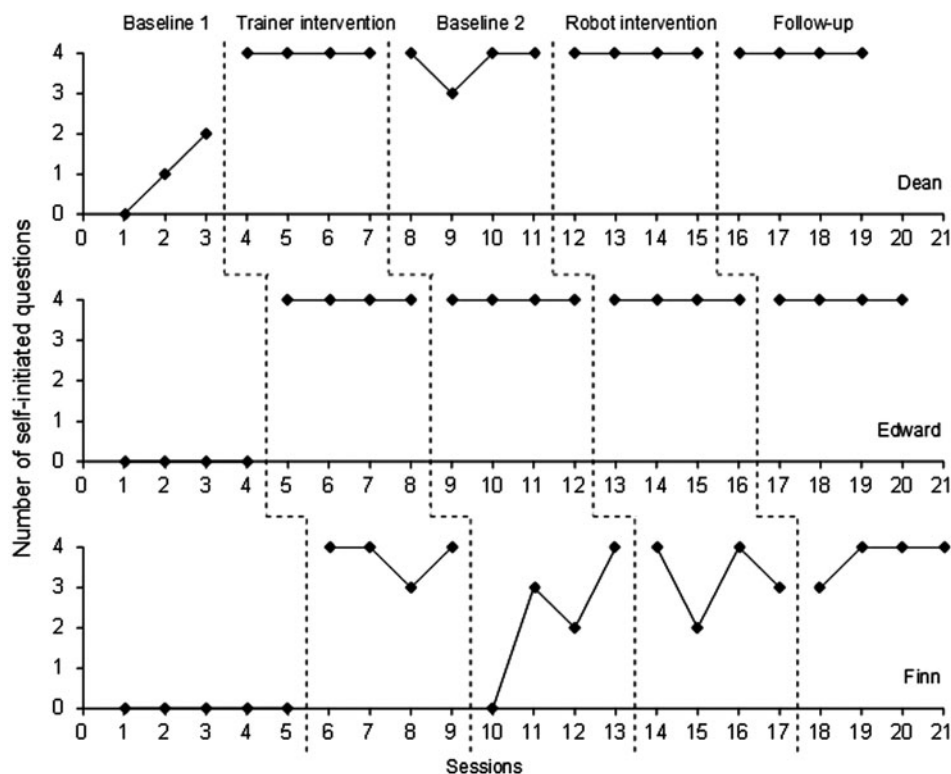
Notes: *Significant at $\alpha = 0.05$.**Significant at $\alpha = 0.01$.

Figure 3 Number of self-initiated questions experimental group 2.

up the number of self-initiated questions remained at a high level ($M = 3.92$, $SD = 0.29$), showing that the self-initiations were maintained.

As shown in Figure 3 and Table VI, the number of self-initiated questions increased between baseline 1 and follow-up. The values of $\text{Tau}_{\text{novlap}}$ for the individual children are provided in Table VII, indicating a significant increase for all children of experimental group 2 ($p < 0.02$). The corresponding overall value of $\text{Tau}_{\text{novlap}}$ is 1.00 ($SE = 0.13$; 90% confidence interval: 0.78–1.00), indicating that the number of self-initiated questions increased during the interventions conducted by a human trainer and a robot.

Comparison of experimental groups

To investigate which of the two interventions, namely the robot intervention or human trainer

intervention, was more effective in promoting self-initiated question asking, only the increase in the mean number of self-initiated questions during the first intervention phase of both experimental groups was compared. Because the number of self-initiated questions remained high during the second baseline for both experimental groups, the mean number of self-initiated questions could only hardly increase during the second intervention (i.e., failure to return to baseline).

The experimental groups differed with regard to the number of self-initiated questions during baseline 1: the children of experimental group 1 engaged in more self-initiated questions ($M = 1.17$) than the children of experimental group 2 ($M = 0.25$). During the first intervention, both experimental groups engaged in an equal number of self-initiated questions ($M = 3.92$).

Discussion

The purpose of the present study was to investigate the effectiveness of an ABA-based intervention conducted by a robot and the effectiveness of the same ABA-based intervention conducted by a human trainer in promoting self-initiated questions in children with ASD.

For both experimental groups, the number of self-initiated questions increased significantly between the first baseline and the first intervention, indicating that the interventions conducted by a robot and a human trainer were both effective in promoting self-initiated questions in children with ASD. Although the increased number of self-initiated questions during the first intervention was larger for the experimental group that first received the human trainer intervention, it cannot be concluded that the human trainer intervention was more effective than the robot intervention. Since the group that first received the robot intervention engaged in more self-initiated questions during the first baseline and both experimental groups performed near the maximum level during the first intervention, the number of self-initiated questions of the group that first received the robot intervention could increase less. Therefore, no conclusions with regard to the differential effectiveness of both interventions can be drawn. The high number of self-initiated questions during follow-up indicates that both experimental groups maintained this skill.

The intervention conducted by a human trainer was effective in promoting self-initiated questions in children with ASD. This result is in line with the

results of several other studies that targeted self-initiations (i.e., question asking) during human trainer interventions based on ABA [6, 8, 10]. However, the intervention conducted by a robot was also effective in promoting self-initiated questions in children with ASD. Until now, no studies have been published in which a robot and a human trainer both conducted the same intervention and in which both interventions led to an increase in the target behaviour. Only Duquette et al. [23] compared a condition with a robot to a condition with a human mediator, but their results were mixed. First, there was no significant increase in the social interaction skills of the children, when comparing pre- and post-intervention measures, suggesting that none of the conditions was effective in promoting social interaction skills. Second, when comparing both conditions, the children paired with the robot showed more shared focused attention and less repetitive or stereotyped problem behaviour, but they also showed less imitative behaviour than the children paired with human mediator. Although the results of Duquette et al.'s study should be interpreted with caution, several factors may have accounted for the mixed results when compared to findings of the present study. First, the children in the study by Duquette et al. were low-functioning (i.e., a deficit in imitation and a severe delay in receptive and expressive language), whereas the children in the present study had a full-scale IQ of at least 80. In general, children with lower IQ-scores show lower abilities and less improvement during interventions in various behaviours than children with higher IQ-scores [38–40]. Second, the robot used in the Duquette et al. study had limited physical abilities compared to the robot used in the present study and lacked reciprocity, which could explain the low number of imitative behaviours [41]. Third, the procedures of Duquette et al. were not explicitly based on evidence based intervention principles, like components of ABA.

Several systematic reviews have provided criteria to evaluate the certainty of evidence for a study [see e.g. 42–45]. Studies have to meet to following five criteria to be classified as “conclusive”: (1) the study used an experimental design (e.g., a group design

Table VI. Mean number of self-initiated questions during each phase and corresponding standard deviations for experimental group 2.

Phase	Mean	SD
Baseline 1	0.25	0.622
Trainer intervention	3.92	0.289
Baseline 2	3.33	1.231
Robot intervention	3.75	0.622
Follow-up	3.92	0.289

Table VII. Individual values of $\text{Tau}_{\text{novlap}}$ for the children of experimental group 2.

Child	Baseline 1 – follow-up contrast			Baseline 1 – trainer intervention contrast		
	$\text{Tau}_{\text{novlap}}$	SD	p	$\text{Tau}_{\text{novlap}}$	SD	p
Dean	1.00	0.43	0.02*	1.00	0.43	0.02*
Edward	1.00	0.38	0.01**	1.00	0.38	0.01**
Finn	1.00	0.37	0.01**	1.00	0.37	0.01**

Notes: *Significant at $\alpha = 0.05$. **Significant at $\alpha = 0.01$.

with random assignment, an ABAB design or a multiple baseline design), (2) adequate inter-observer agreement and treatment integrity were reported (i.e., measured during a minimum of 20% of the sessions with at least 80% agreement and treatment integrity), (3) operational definitions for dependent variables were provided, (4) sufficient details for replication were provided and (5) the study's design provided at least some control for alternative explanations for increases in the target behaviour (e.g., a multiple baseline design across participants in which the start of interventions was staggered and simultaneous interventions targeting the same behaviour were held constant [44]). The present study meets all of these criteria. First, an experimental design was used (i.e., a combined crossover multiple baseline design across participants with random assignment to experimental groups). Second, inter-observer agreement and treatment integrity were adequate (i.e., the mean overall kappa coefficient and PABAK for inter-observer agreement were 0.97, respectively, 0.99 and treatment integrity was 99%). Third, an operational definition of a self-initiated question was provided (i.e., the child asks a syntactically correct question that implies carrying out an action). Fourth, participant characteristics, setting, materials, procedures and data-collection were described in detail to enable replication. Fifth, the design of the present study controlled for alternative explanations. Therefore, the present study is the first study that measured children's behaviour during a robot intervention and that found evidence for the effectiveness of a robot intervention.

However, this study also has several limitations. First, the sample size was small and the children in the two experimental groups were not matched on age, IQ and social communication skills. Second, the data lacked variance, because the number of opportunities to engage in a self-initiated question was rather small, which restricted the possibilities for discriminating across phases and participants. Drawing a conclusion with regard to the differential effectiveness of the robot and human trainer was therefore impossible. Third, the programmed robot behaviours were rather limited, which made it difficult to adapt the robot's behaviour to the preferences and abilities of the children. Fourth, social validity was not assessed. Social validity is an important feature of intervention studies, as it indicates whether the goals, procedures and effects of an intervention are seen as desirable by children, parents, and practitioners. Socially invalid and unaccepted interventions are unlikely to be implemented [31, 46, 47].

There are several important directions for future research. First, there is a need for more rigorous studies that compare the differential effectiveness of

interventions conducted by a robot and a human trainer for various target behaviours. Second, robots should be programmed in such a way that they are better able to respond to the children's behaviours and that their behaviours could be easily adapted to the children's individual preferences and abilities. Third, future studies should be conducted with younger children and children with lower verbal abilities to investigate whether these children could also benefit from the robot interventions targeting self-initiations. This would broaden the clinical application of robots. Fourth, robot interventions should also target other social interaction skills (e.g., imitation, initiating, and maintaining conversations, offering help and cooperation). Furthermore, robots should be deployed as mediators by mediating the social interaction between the child with ASD and his or her parents or peers [13]. Finally, the social validity of interventions conducted by robots should be systematically assessed. Although a robot intervention might be effective in promoting certain target behaviour, the intervention will not be implemented in clinical practice if the children with ASD are not enjoying the sessions with the robot and parents and practitioners do not consider a robot as desirable agent to conduct an intervention.

Nevertheless, this study demonstrates that both an intervention conducted by a robot and an intervention conducted by a human trainer is effective in promoting self-initiated questions in children with ASD. The findings in this study provide support for using robots in interventions for children with ASD. However, a human trainer is still needed to control the robot. The clinical application of a robot, in e.g. practicing skills with the robot outside the training, is therefore still limited. Future research should reveal whether robot interventions are more or less effective than interventions conducted by human trainers and whether these interventions are socially valid.

Acknowledgments

The authors gratefully acknowledge the support of the Innovation-Oriented Research Program "Integral Product Creation and Realization (IOP IPCR)" of the Dutch Ministry of Economic Affairs, Agriculture and Innovation, The Hague. The authors thank the children, their parents, and the staff of the day treatment group and the clinical treatment facility of the Dr Leo Kannerhuis for their participation. We also would like to thank Margreet Weide and Anke van der Linden for their assistance during preparation and data collection. Finally, we would like to thank Iris Soute for her assistance with the robot in the sessions and Tino Lourens, TiViPE, for his assistance in programming the robot.

Declaration of interest: The authors have no declaration of interest to report. The authors alone are responsible for the content and writing of this article.

References

1. Koegel LK, Koegel RL, Harrower JK, Carter CM. Pivotal response intervention I: Overview of approach. *Journal of the Association for Persons with Severe Handicaps* 1999;24:174–185.
2. Koegel LK, Koegel RL, Shoshan Y, McNeerney EK. Pivotal response intervention II: Preliminary long-term outcome data. *Journal of the Association for Persons with Severe Handicaps* 1999;24:186–198.
3. Koegel RL, Koegel LK, McNeerney EK. Pivotal areas in intervention for autism. *Journal of Clinical Child & Adolescent Psychology* 2001;30:19–32.
4. Koegel RL, Koegel LK. Pivotal response treatments for autism: Communication, social, & academic development. Baltimore: Paul H. Brookes Publishing; 2006.
5. Freedden RM, Koegel RL. The pivotal role of initiations in habilitation. In: Koegel RL, Koegel LK, editors. *Pivotal response treatments for autism: Communication, social, and academic development*. Baltimore: Paul H. Brookes Publishing; 2006. pp 165–186.
6. Koegel LK, Camarata SM, Valdez-Menchaca M, Koegel RL. Setting generalization of question-asking by children with autism. *American Journal on Mental Retardation* 1998;102:346–357.
7. Stone WL, Ousley OY, Yoder PJ, Hogan KL, Hepburn SL. Non-verbal communication in two- and three-year old children with autism. *Journal of Autism and Developmental Disorders* 1997;2:677–696.
8. Koegel LK, Carter CM, Koegel RL. Teaching children with autism self-initiations as a pivotal response. *Topics in Language Disorders* 2003;23:134–145.
9. Bono MA, Daley T, Sigman M. Relations among joint attention, amount of intervention and language gain in autism. *Journal of Autism and Developmental Disorders* 2004;34:495–505.
10. Koegel LK, Koegel RL, Green-Hopkins I, Carter Barnes C. Brief report: Question-asking and collateral language acquisition in children with autism. *Journal of Autism and Developmental Disorders* 2010;40:509–515.
11. Barakova EI, Gillesen JCC, Huskens BEBM, Lourens T. End-user programming architecture facilitates the uptake of robots in social therapies. *Robotics and Autonomous Systems* 2012; <http://dx.doi.org/10.1016/j.robot.2012.08.001>. Accessed 1 October 2012.
12. Gillesen JCC, Barakova EI, Huskens BEBM, Feijs LMG. From training to robot behavior: Towards custom scenarios for robotics in training programs for ASD. 2011 IEEE International Conference on Rehabilitation Robotics (ICORR) 2011;5975381:1–7.
13. Werry I, Dautenhahn K, Harwin W. Investigating a robot as a therapy partner for children with autism. 6th European Conference for the Advancement of Assistive Technology; Berlin/Heidelberg: Springer Verlag; 2011. pp 57–74.
14. Van der Meer LAJ, Rispoli M. Communication interventions involving speech-generating devices for children with autism: A review of literature. *Developmental Neurorehabilitation* 2010;13:294–306.
15. Piper AM, O'Brien E, Morris MR, Winograd T. SIDES: A cooperative tabletop computer game for social skills development. In: Hinds P, Martin D, editors. 20th Anniversary conference on computer supported cooperative work; New York, NY: Association for Computing Machinery; 2006. pp 1–10.
16. Bellani M, Fornasari L, Chittaro L, Brambilla B. Virtual reality in autism: State of art. *Epidemiology and Psychiatric Sciences* 2011;20:235–238.
17. Dautenhahn K, Billard A. Games children with autism can play with Robota, a humanoid robotic doll. In: Keates S, Langdon PM, Clarkson PJ, Robinson P, editors. *Universal access and assistive technology*. London: Springer Verlag; 2002. pp 179–190.
18. Scassellati B. How social robots will help us to diagnose, treat, and understand autism. *Robotics Research* 2007;28:552–563.
19. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders* 2012;6:249–262.
20. Dautenhahn K, Werry I. Towards interactive robots in autism therapy. *Pragmatics & Cognition* 2004;12:1–35.
21. Flannery KB, Horner RH. The relationship between predictability and problem behavior for students with severe disabilities. *Journal of Behavioral Education* 1994;4:157–176.
22. Feil-Seifer D, Mataric MJ. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. *Experimental Robotics* 2009;54:201–210.
23. Duquette A, Michaud F, Mercier H. Exploring the use of a mobile robot as an imitations agent with children with low-functioning autism. *Autonomous Robots* 2008;24:147–157.
24. Robins B, Dautenhahn K, te Boekhorst R, Billard A. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot encourage social interaction skills? *Universal Access in the Information Society* 2005;4:115–120.
25. Stanton CM, Kahn PH Jr, Severson RL, Ruckert JH, Gill BT. Robotic animals might aid in the social development of children with autism. 3rd ACM/IEEE International Conference on Human-Robot Interaction; 2008. New York: ACM Press; 2008. pp 271–80.
26. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author; 2000.
27. Rutter M, Bailey A, Lord C. *Social communication questionnaire*. Los Angeles: Western Psychological Services; 2003.
28. Warreyn P, Raymaekers R, Roeyers H. SCQ: Handleiding vragenlijst sociale communicatie [SCQ: manual social communication questionnaire]. Destelbergen: SIG vzw; 2004.
29. Mataric MJ, Eriksson J, Feil-Seifer DJ, Winstein CJ. Socially assistive robots for post-stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation* 2007;4:5, <http://www.jneuroengrehab.com/content/4/1/5>. Accessed 1 October 2012.
30. Lourens T, Barakova EI. User-friendly robot environment for creation of social scenarios. In: Ferrandez JM, Alvarez JR, de la Paz F, Toledo FJ, editors. *New challenges on bioinspired applications – Proceedings of the 4th international work-conference on the interplay between natural and artificial computation, Part II, IWINAC 2011, Lecture Notes in Computer Science* (Vol. 6687); La Palma, Canary Islands, Spain, May 30–June 3, 2011; pp 213–221.
31. Kennedy CH. *Single-case designs for educational research*. Boston: Pearson Education; 2005.
32. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; 20:37–46.

33. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 1993;46:423–429.
34. Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, Tyrer P. Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *The Journal of Nervous and Mental Disease* 2006;194:557–564.
35. Parker RI, Vannest KJ, Davis JL. Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification* 2011;35:303–322.
36. Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy* 2011;42:284–299.
37. Abrahamson HH. Winpepi updated: Computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations* 2011;8:1–9.
38. Buitelaar JK, Van der Wees M, Swaab-Barneveld H, Van der Gaag RJ. Verbal memory and IQ predict theory of mind and emotion recognition ability in children with autism spectrum disorders and in psychiatric control children. *Journal of Child Psychology and Psychiatry* 1999;6:869–881.
39. Howlin P, Magiati I, Charman T. Systematic review of early intensive behavioural interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities* 2009;114:23–41.
40. Matson JL, Smith KRM. Current status of intensive behavioral interventions for young children with autism and PDD-NOS. *Research in Autism Spectrum Disorders* 2008;2:60–74.
41. McDuffie A, Turner L, Stone W, Yoder P, Wolery M, Ulman T. Developmental correlates of different types of motor imitation in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2007;37:401–412.
42. Palmen A, Didden R, Lang R. A systematic review of behavioral intervention research on adaptive skill building in high-functioning young adults with autism spectrum disorder. *Research in Autism Spectrum Disorders* 2012;6:602–617.
43. Ramdoss S, Lang R, Mulloy A, Franco J, O'Reilly M, Didden R, Lancioni G. Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: A systematic review. *Journal of Behavioral Education* 2011;20:55–76.
44. Ramdoss S, Machalicek W, Rispoli M, Mulloy A, Lang R, O'Reilly M. Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: A systematic review. *Developmental Neurorehabilitation* 2012;15:119–135.
45. Rispoli M, Neely L, Lang R, Ganz J. Training paraprofessionals to implement interventions for people with autism spectrum disorders: A systematic review. *Developmental Neurorehabilitation* 2011;6:378–388.
46. Callahan K, Henson RK, Cowan AK. Social validation of evidence-based practices in autism by parents, teachers and administrators. *Journal of Autism and Developmental Disorders* 2008;38:678–692.
47. Schwartz IS, Baer DM. Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis* 1991;24:189–204.