UC Santa Barbara

Title

Social conversation skill improvements associated with the Social Tools And Rules for Teens program for adolescents with autism spectrum disorder: Results of a randomized controlled trial

Permalink

Journal

ISSN

Authors

Ko, Jordan A
Miller, Amber R
Vernon, Ty W

Publication Date

DOI

Original Article

Social conversation skill improvements associated with the Social Tools And Rules for Teens program for adolescents with autism spectrum disorder: Results of a randomized controlled trial

Jordan A Ko, Amber R Miller and Ty W Vernon

Abstract

There has been a significant increase in the development of interventions to improve the social competence and success of adolescents with autism spectrum disorder. The current investigation used direct observation and coding of social conversations as a rigorous method to further assess the efficacy of the Social Tools And Rules for Teens socialization intervention for adolescents with autism spectrum disorder in the context of a randomized controlled trial. A total of 35 adolescents with high-functioning autism spectrum disorder were randomized to either a treatment or waitlist control group. The 20-week group intervention took place once a week for 90 min per session. Brief video-recorded conversations between participants and unfamiliar, untrained peers were recorded at pre- and post-time points and coded for selected social behaviors (i.e. questions asked, positive facial expressions, and mutual engagement). Results revealed a significant Group × Time treatment effect for both questions asked and positive facial expressions. The findings support that the Social Tools And Rules for Teens intervention can positively impact specific, observable social behaviors through systematic coding of live social conversations within the context of a randomized controlled trial. This investigation is one of the first randomized controlled trials of a group socialization intervention to use systematic coding of live social conversations to assess social competence improvements.

The unique needs of youth with autism spectrum disorder (ASD) have prompted a surge of research interest in the development and evaluation of socialization interventions for this population (see Gates et al., 2017, for a recent review). One such model, the Social Tools And Rules for Teens (START) program (Vernon et al., 2018), is a socialization intervention approach that combines experiential and didactic learning components into a single multi-component treatment model. Spanning 20 weeks, this group intervention uses a social club-like format, same-aged peer models, structured and unstructured socialization opportunities, and both individualized and group skill components to target core socialization challenges of adolescents with ASD.

There is emerging evidence that the START program is an efficacious approach for improving the social competencies and motivation of individuals diagnosed with ASD. An initial pilot study investigation of six participants yielded encouraging improvements across a multitude of outcome measures, including parent-report survey measures, adolescent self-report survey measures, and coded observational data derived from video-recorded social conversations (Vernon et al., 2016). A randomized controlled trial (RCT) of the START program revealed significant improvements on parent and self-report survey measures unique to the treatment participants (Vernon et al., 2018). Specifically, participants receiving

University of California, Santa Barbara, USA

Corresponding author:
Jordan A Ko, Department of Counseling, Clinical, & School Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106-9490, USA.
Email: jordanalexisko@gmail.com

the 20-week START program demonstrated significant gains in survey measures associated with autism symptom severity, social competency, and social motivation. The START program joins a variety of other promising models that have primarily utilized parent and self-report survey measures to assess the efficacy of treatment programs (see Gates et al. (2017) for a meta-analysis and see McMahon et al. (2013a), Miller et al. (2014), and White et al. (2007) for reviews). However, these studies advocate for using additional outcome measures in future social skill efficacy studies to account for the complexity and multidimensionality of social competence.

While survey measures are easy to administer and are a relatively inexpensive and efficient means for collecting pre-intervention and outcome data, they have also notable drawbacks. Specifically, there is risk of their report being influenced by social desirability bias, random responding, demand characteristics, or other reporting biases (Furr and Funder, 2007; McMahon et al., 2013a). Another drawback of survey measures is that the size of intervention effects often varies greatly depending on the reporter, with parents generally endorsing smaller effects than the actual participants (Gates et al., 2017). In addition, parent-report measures presume that parents possess an adequate and accurate knowledge of their adolescent's socialization efforts and success, which may not always be the case. Adolescents often spend a limited amount of time with their parents and sometimes employ different interaction strategies with family members than they do with their peer group. These factors may also explain the discrepancy in parent-report versus adolescent self-report (McMahon and Solomon, 2015), which may reflect some level of reporting biases in both parties.

Given the various limitations associated with survey measures, various research groups have called for augmenting efficacy trials with additional, potentially more rigorous and objective measures (Cunningham, 2012; Lord et al., 2005; Matson and Rieske, 2014; McMahon et al., 2013a). It is crucial to identify reliable and socially valid measures both within the immediate context of treatment and within naturalistic contexts (White et al., 2007). Maximizing the objectivity of the utilized measurement strategy is another core consideration (Kasari, 2002). Direct observation and systematic coding of social behavior within a naturalistic context offer these advantages. Such observational methods have a long history in the autism research literature (see Matson and Rieske, 2014). They are commonly used in single-case experimental designs within behavioral intervention paradigms (see Vismara and Rogers, 2010) and increasingly used in larger intervention trials of children and adolescents with ASD (e.g. Dolan et al., 2016; Kasari et al., 2012; Ratto et al., 2011). However, according to a review of measurement tools in ASD intervention studies with large sample sizes (20 or more participants), videotaped observation data were only used in 1.9% of studies (Bolte and Diehl, 2013). The cost and time of systematic behavioral coding using trained and reliable observers may be factors that prevent more widespread use of these outcome measures (Furr and Funder, 2007).

Despite requiring additional time and resources, using behavioral observation of socialization efforts as an outcome measure offers several strengths. This method provides a direct, quantifiable measurement that is not subject to the same biases inherent in survey measures (Bakeman, 2000; Johnson and Turner, 2003; Moskowitz, 1986; Ratto et al., 2011; Suen and Ary, 1989). Focusing on specific, operationally defined target behaviors eliminates the subjectivity associated with general impression ratings (Matson and Rieske, 2014). Another benefit is that behavioral observation can be conducted within the context of a natural social encounter and is therefore a direct, socially valid sampling of social performance (McMahon et al., 2013a, 2013b). Specifically, observation allows researchers to analyze multiple dimensions of interpersonal interaction, including the occurrence of key social skills. Behavioral observations are applicable to a variety of populations, behaviors, and settings, thus providing a wealth of information critical for assessing generalization. Furthermore, insight into individuals' general social functioning often emerges from observing samplings of their behavior. Specifically, this method measures behaviors that individuals and parents may not be able to accurately describe or are disinclined to report (Hartmann et al., 2004).

The current investigation aimed to expand the evidence base of the START socialization program for adolescents with ASD. Specifically, this study examined pre- to post-intervention changes in key verbal and nonverbal social behaviors during naturalistic conversation probes between participants and unfamiliar, untrained peers to further assess the efficacy of the START program in situations that replicate real-world social encounters. It was hypothesized that there would be significant differences on all examined target social behaviors (questions asked, positive facial expressions, and mutual engagement) between treatment and waitlist groups after the conclusion of the 20-week trial period.

Method

Participants

A total of 44 potential participants were originally recruited for this study through communications with local high schools, referrals from community organizations, responses from online announcements, and word of mouth. Participants were required to be between 12 and 17 years of age with a diagnosis of ASD meeting current *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5) criteria, possess a verbal IQ of 70 or

above, and demonstrate the ability to communicate and comprehend full sentence phrases. Interested families were required to provide documentation (e.g. psychological evaluation report) of an ASD diagnosis. Parents were administered the Social Responsiveness Scale, Second Edition (SRS-2; Constantino and Gruber, 2005) and participants were required to fall between the mild and severe range for ASD symptoms. A verbal IQ of 70 or above was confirmed through being administered the Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman and Kaufman, 2004). Of the 44 individuals recruited, a total of 40 adolescents were deemed eligible to participate (three did not have an ASD diagnosis and one did not meet language and verbal IQ requirements). One parent/caregiver per adolescent was also required to participate in the program. Parent participation involved taking part in a weekly checkout meeting (i.e. the last 5–10 min of each START session) and completing parent survey measures at pre- and post-intervention. This study received institutional review board approval and all families completed an informed consent and assent process prior to participating. During the intake meeting, oral and written descriptions of the project components were described and families were given the opportunity to ask questions or express concerns.

Recruitment occurred in two separate yearly cycles. Approximately half of the project participants were recruited and randomized in the first year and the other half were recruited in the second year. Of the 40 participants who met eligibility criteria and were randomized, five participants did not complete the project. Four participants in the treatment group withdrew for the following reasons: participating in a highly desired social activity that conflicted with the group time ($n = 2$), expressing dissatisfaction with assigned group members being much younger in age ($n = 1$), and not being able to continue driving the 2-h round trip to group ($n = 1$). One waitlist participant was no longer able to participate in the study and therefore did not complete post-waitlist assessments. Therefore, 35 participants (16 participants in the treatment group and 19 participants in the waitlist group) and their parents were included in the final analyses. The full sample of 35 participants had a mean age of 13.51 years ($SD = 1.50$) and grade of 8.06 ($SD = 1.43$). In total, 69% of participants were male ($n = 24$) and 31% were female ($n = 11$). The ethnic background of the sample was 60% European American/White ($n = 21$), 14% Hispanic American/Latino(a) ($n = 5$), 6% Asian American ($n = 2$), 6% Middle Eastern ($n = 2$), and 14% Multiracial/Multiethnic (White/Hispanic $n = 4$; White/Hispanic/Filipino $n = 1$). There were no significant group differences on any demographic variables at baseline (all $p > 0.05$). Figure 1 depicts the project CONSORT diagram (using guidelines provided by Schulz et al. (2010)).

Research design

An RCT design was employed to examine group differences in social behavior changes. The 40 participants were randomly assigned to either the START treatment group or waitlist group (20 in each group). Pairs of eligible participants were randomized using a simple randomization procedure (a coin flip) in which each participant had an equal chance of being assigned to the treatment or waitlist group. Adolescents who were randomized to the treatment group immediately began the 20-week intervention and were re-evaluated at the end of the 20 weeks. Adolescents who were randomized to the waitlist group went through a 20-week waiting period while continuing any existing therapy services. They were re-evaluated at the end of the 20-week waiting period (post-waitlist assessments) and then received the 20-week intervention themselves shortly after completing the post-waitlist assessments.

Procedure

Pre-intervention sessions. Before participating in the trial, adolescents completed an intake session. During these intake sessions, adolescent participants and their parents completed the consent/assent process, filled out a number of survey measures, and participants engaged in two 5-min videotaped social conversations with unfamiliar peers (described in detail below).

Target skill selection and self-management. After completing the intake session, the participant, parent, and intake clinician were asked to independently rank the three most prominent social vulnerabilities of the adolescent from a list of possible skills (with the option to write-in other challenges not listed). Based on these endorsements, consensus was reached on an individualized target behavior that would be the participant's primary focus for the first 5 weeks of the intervention program. Selected target behaviors varied between participants and often included: making on-topic comments, asking questions, sharing personal information, limiting verbal contributions (keeping it brief), making positive statements, and showing interest.

Participants were taught to use a self-management strategy to track and increase the use of their target behavior during the weekly START sessions. This method has been used in previous studies of the START intervention (Vernon et al., 2016), as well as previous studies that have specifically targeted social skills (e.g. Doggett et al., 2013; Koegel et al., 1992; Newman and Ten Eyck, 2005). When a new target behavior was taught, the behavior was first clearly defined for the participant. Then participants were taught to use a small, discrete digital tally counter that fit in the palm of their hand to track their use of their target
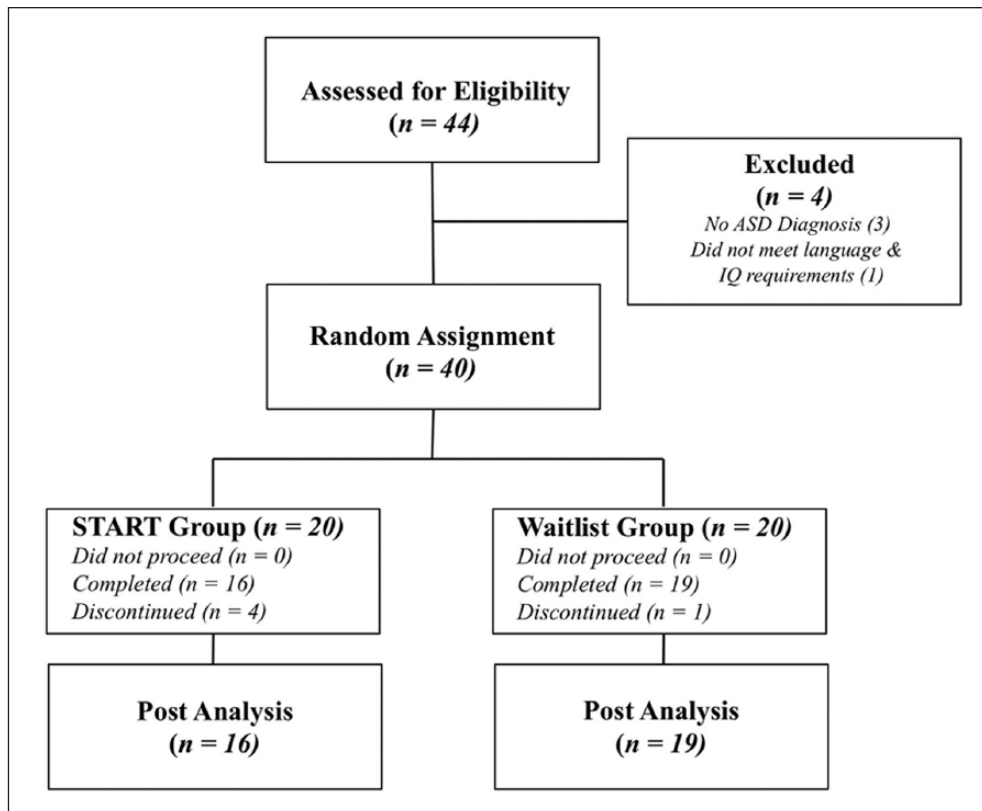
Figure 1. START RCT CONSORT diagram.

skill. After the facilitator modeled how to use the digital tally counter to track the target skill by providing examples, the facilitator checked for the participant's understanding of the behavior and correct use of self-management using the digital tally counter through engaging in a brief practice conversation. At the check-in phase of each weekly group, participants practiced tracking for their target skill during a brief conversation and received feedback. Participants were asked to track their use of the target skill during each weekly START group session (throughout the entire group) and report on their experience at the end of each session during the checkout phase. During this time, the facilitator also collected the digital tally counter and recorded the number of times the participant used the target skill each week.

Every 5 weeks, participants and facilitators jointly made the decision to continue with the current target skill for another 5 weeks or to focus on another skill. This decision was made through examination of currently selected skill improvements and noted challenges in other social domains that might be impeding the participant's social success. Improvement was determined by examining the recorded number of times the participant was using the skill in the past 5 weeks and also the facilitator's clinical judgment of whether it would be more beneficial to continue practicing the same skill or if another target skill was more pressing. Please see Vernon et al. (2016) and Vernon et al. (2018) for more details on target skills and self-management procedures.

START program overview. The START program is a 20-week socialization intervention that blends experiential and instructional components with an emphasis on motivational elements and active involvement from participants and peer models. Trained undergraduate research assistants serve as primary group facilitators. In addition, typically developing same-aged peers, recruited from local high schools, serve as social partners and function as crucial peer models to improve participants' social knowledge, motivation, and skills. The program allows for individually tailored treatment components through opportunities to practice the previously described individualized target social skill while immersed in a group intervention setting. This structure allows a balance between addressing the global social needs of adolescents with ASD while also attending to each participant's individual areas of difficulty. The experiential component immerses participants in a dynamic and shared learning environment where participants actively test out newly acquired skills and strategies in a socially supportive atmosphere, receive personalized feedback, and reflect on their experiences with other group members. It is through active exploration and
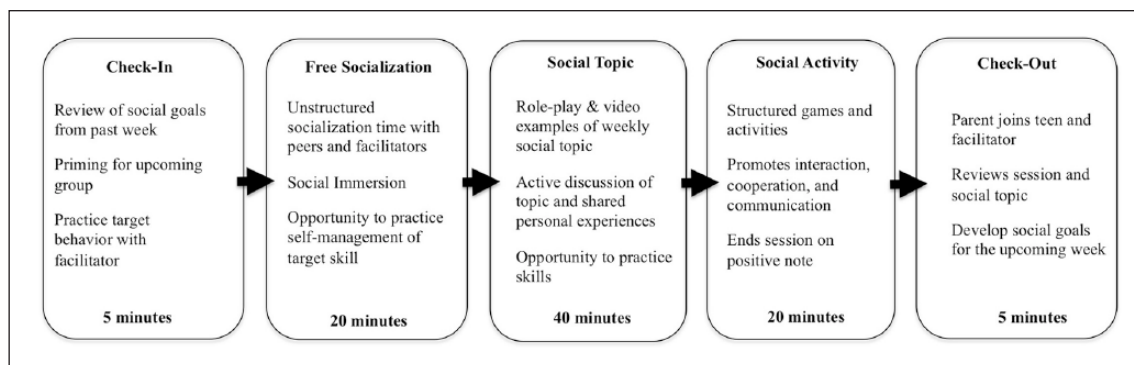
Figure 2. START program session format.

live testing of social strategies during group that participants not only gain crucial social practice but also have successful social experiences in which to draw on when practicing outside of group. The social topic for each group session also changes weekly, beginning with basic conversation skills (e.g. greetings and initial introductions) and building to more complex skills (e.g. working in a group) as the program progresses. Parents are involved at the end of each group, during a checkout session, to become oriented to the current social topic and help set weekly social goals to promote generalization of newly learned skills.

The START program consisted of weekly 90-min group sessions (20 sessions that occurred once a week). Each group typically consisted of four to six participants, two high school peers, and four undergraduate facilitators. Intervention was conducted in a spacious lounge-like room with couches, chairs, and a flat screen television. START group facilitators completed a comprehensive training that lasted a total of approximately 10 h. Training consisted of a comprehensive overview of the START program and procedures, an introduction to ASD, group facilitation techniques, basic therapy/helping skills, and behavioral strategies. Each component of the weekly sessions was introduced (described in the next section), and then sequentially discussed, modeled, and role-played. Common challenges and effective strategies for handling these scenarios were also discussed. Facilitators received continued weekly supervision in a group format from an advanced doctoral student and a licensed psychologist throughout the study duration. During these supervision meetings, facilitators provided an update on their assigned participants' social progress, discussed group dynamics, and sought support with current challenges. The supervisors provided feedback on specific strategies to increase participant engagement, address group concerns, and teach core program content. High-school peers were volunteers recruited through announcements at local high schools and through word of mouth. Since their involvement during group sessions was similar to that of the participants, they received an orientation to the group program but did not receive intensive clinical training.

Their main roles were to model appropriate social skills, serve as additional social partners for participants, and contribute to a positive peer environment. Peers received volunteer hours for their participation and time.

Each weekly intervention session included the following components: an individual check-in session, a group session (consisting of an unstructured socialization time, a group activity, and a discussion and practice of a weekly social topic), and an individual checkout session with the parent. See Figure 2 for a visual depiction of a weekly session.

Individual check-in session. The 5-min individual check-in session primarily consisted of a review of the participant's social goals from the previous week, priming for the upcoming group's discussion topic and activity, and practicing the target skill with a facilitator. The participants and facilitators were then brought together for the group session.

Free socialization phase. Each group session began with 20 min of unstructured socialization time, which provided an opportunity for participants to engage in natural interactions with each other and group facilitators. Topics were not pre-determined and often included weekend plans, favorite movies and video games, and hobbies. During these interactions, everyone (i.e. participants, facilitators, and peers) discretely tracked their use of their individualized target skill using the previously described self-management strategy. Facilitators and peers were encouraged to track the use of a skill themselves in order to normalize the self-management process for participants. This casual, unstructured socialization phase allowed participants to freely experiment and engage in experiential social learning in a safe, supportive environment.

Social topic discussion and practice. A 40-min discussion of a weekly social topic followed the free socialization phase (see Table 1 for a list of weekly social topics covered throughout the 20 weeks). The topic was introduced with role-play demonstrations and video clips from popular

Table 1. START curriculum weekly session topics.

Session	START topic
1	First impressions/greeting others/making initial introductions
2	Using questions in conversation
3	Using comments in conversation
4	Showing interest—attention, eye contact, and facial expressions
5	Choosing appropriate topics for conversations
6	Making and keeping friends
7	Changing topics/ending conversations/saying goodbye
8	Reducing anxiety/being comfortable during social exchanges
9	Expressing empathy
10	Complimenting others
11	Giving social feedback
12	Receiving social feedback
13	Respectfully disagreeing with others
14	Demonstrating good sportsmanship/being a good winner and loser
15	Working in a group/being a good team member/ leader
16	Understanding/using appropriate humor and sarcasm
17	Having social courage/joining a new group of peers
18	Using cell phones and social media
19	Hosting others at your home/being a good guest at someone's home
20	Summary of group topics and conclusion

START: Social Tools And Rules for Teens.

television shows and movies illustrating "good" and "bad" examples of a social skill. Next, the adolescents and group facilitators engaged in an active discussion in which participants were encouraged to share their experiences and suggestions. Facilitators used engaging techniques, such as posing scenarios and sharing personal stories of social successes and failures related to the topic, which brought the concepts to life and elicited interest and involvement from participants. Participants were also provided the opportunity to practice the skills related to the weekly social topic with facilitators and peers.

Social activity phase. The group continued with 20 min of structured social icebreakers, adolescent-oriented party games, and other activities designed to create opportunities to work on communication skills, cooperation, and group cohesion among members. Examples of social activities included the following: Get to know you BINGO, Apples to Apples, Human Knot, and Charades. This phase allowed another opportunity for participants and facilitators to socialize in a more structured context, which reduced the emphasis on face-to-face interaction by focusing attention on a shared group task.

Checkout session. Finally, participants completed a 5-min checkout session with their parent and a facilitator, where they reviewed the weekly social topic and developed social goals for the upcoming week.

Fidelity of implementation

In order to ensure that all described components were consistently implemented each group, a research assistant completed a weekly fidelity checklist during each session. Phases and content were checked off as they were covered and any deviations from the specified time frame were noted. A copy of this fidelity form can be obtained as a digital appendix in Vernon et al. (2018). An analysis of these fidelity checklists indicated that procedures were correctly implemented 97.7% of the time.

Social conversation data collection

Social conversation data were collected during the pre-intervention intake session and after program completion (i.e. after 20 weeks of the START program or waitlist participation). At these time points, participants engaged in 5-min conversations with unfamiliar, peers whom they had never met. In contrast to existing social conversation measures using trained confederates (e.g. Ratto et al., 2011), these peer conversational partners were not trained or coached prior to their interactions with the participants. They also possessed no knowledge of the purpose of the research study and did not know the diagnostic status of the participants. Peers were recruited through advertisements and word of mouth and were compensated with US$5 gift cards for their time. They were randomly assigned to participants and never completed more than one conversation with the same participant.

Conversations took place in a private room not used for the START group intervention and were recorded by a camera in the corner of the room. After verbally confirming that the conversation partners did not know one another, participants and peer conversation partners were provided with the instruction: "You will be having a conversation with another person that you have never met before. You will have five minutes to get to know each other." Participants were not provided with any further instructions or encouraged to use specific social skills. Our social conversation probes were also unique in that participants completed two of these video-recorded conversations at each time point, one with a female conversation partner and another with a male conversation partner. Conversations took place with both a male and a female in order to take into account participants' performance with peers of the same and opposite gender. The order of the conversations (female versus male) was altered each time so that participants did not always speak to the same gender first.

Dependent measures

Dependent measures were selected to provide representative data associated with improvements to the core social deficits of ASD (American Psychiatric Association (APA), 2013). Video-recorded conversations pre- and post-intervention (140 total 5-min conversation videos) were systematically coded for each dependent measure. The three dependent measures coded were questions asked, positive facial expressions, and mutual engagement.

Questions asked. Social inquiries have been identified as a crucial interpersonal strategy, but compared to typically developing adolescents, adolescents with ASD are known to make far fewer social initiations to their peers (Doggett et al., 2013; Palmen et al., 2008; Weiss and Harris, 2001). Social initiations were measured through the participant's attempts to learn more about their social partner by asking questions. A question was defined as a verbal query that is intended to elicit a response from the conversational partner. Trained coders, who were blind to the treatment status of the participants, counted each question asked in the video clips by both the participant and the conversational partner. In order to take into account the number of questions asked by the conversational partner, questions asked was defined as the percentage of questions asked by the participant during conversation.

Positive facial expressions. Individuals with ASD commonly have deficits in appropriate affective expression (Bieberich and Morgan, 1998; Hobson and Lee, 1998; Stagg et al., 2014). Smiling and laughter have long been established as important nonverbal indicators of social attunement in conversation and are generally associated with friendliness and positive impressions (Lau, 1982). Nonverbal social engagement was measured through observed efforts to convey interest in a social partner's conversation (positive facial expressions). A 5-s partial interval coding scheme was implemented to code for the presence or absence of a positive facial expression (defined as visible smiling or laughing). A percentage was calculated to determine the percentage of time the participant was displaying positive facial expressions during each 5-min clip.

Mutual engagement. Challenges with social reciprocity are a hallmark characteristic of ASD (APA, 2013; White et al., 2007). To measure changes in reciprocity, mutual engagement, or the extent to which both social partners were jointly engaging in conversation, was examined to explore the balance of conversational contributions. This measure specifically examined the contributions of a participant relative to their conversational partners. Coders examined 5-s intervals and focused on whether (1) the participant was the primary speaker, (2) the conversational partner was the primary speaker, (3) both partners contributed equally to the conversational volley, or (4) no one spoke. Mutual engagement was defined as the percentage of intervals in which both individuals contributed equally in a back-and-forth manner.

Inter-observer reliability

There were three video coders who coded all 5-min videos. Coders were trained in using Noldus Observer XT 10.5 software and were blind to the treatment status of participants and time point of the videos. From the three coders, teams of two coders were responsible for coding each behavior. Inter-observer reliability was established between the two coders prior to coding videos through establishing a clear operational definition for each behavior and practicing using a minimum of 20 practice videos. In the process of establishing reliability, coders attended a weekly supervision meeting to discuss disagreements, ask questions, and refine the operational definition if needed. Once reliability was achieved, research assistants coded videos, which were assigned in a random order to control for potential observer drift. Inter-observer reliability was calculated for videos coded each week, and coders continued attending a weekly supervision meeting throughout the duration of coding in order to ensure adherence to the definition and consistency in coding. Inter-observer reliability was calculated for approximately 33% of the coded video recordings. Reliability was estimated using both percent agreement and kappa statistic. Using percent agreement, inter-observer reliability averaged 0.88 ($SD = 0.06$) for questions asked, 0.93 ($SD = 0.05$) for positive facial expressions, and 0.87 ($SD = 0.05$) for mutual engagement. Using kappa statistic, inter-observer reliability averaged 0.61 ($SD = 0.24$) for questions asked, 0.65 ($SD = 0.32$) for positive facial expressions, and 0.79 ($SD = 0.09$) for mutual engagement.

Data analysis

Chi-square tests for independence and independent-samples t tests were used to examine pre-intervention differences between the treatment and waitlist groups. To assess for differences in social interaction skills between treatment and waitlist groups from pre- to post-intervention, two-way mixed analysis of variances (ANOVAs) were conducted with group (treatment or waitlist) as the between subjects factor and time (pre- or post-intervention) as the within-subjects factor. The significance level for all comparisons was set at $p = .05$. Group × Time differences on each of the three measures were examined. Partial η^2 effect size was calculated for all significant effects.

Results

Table 2 summarizes the mean demographic and baseline variables for the 35 adolescents who participated in the

Table 2. Participant demographics.

Variable	M (SD), n (%)		p value
	Treatment (n = 16)	Waitlist (n = 19)	
Age (years)	13.25 (1.48)	13.74 (1.52)	ns
Grade	7.75 (1.57)	8.32 (1.29)	ns
Gender			
Male	11 (69%)	13 (68%)	ns
Female	5 (31%)	6 (32%)	ns
Ethnicity			
European American/White	10 (62%)	11 (58%)	ns
Hispanic American/Latino(a)	2 (13%)	3 (16%)	ns
Asian American	1 (6%)	1 (6%)	ns
Middle Eastern	0 (0%)	2 (10%)	ns
Multiracial/Multiethnic	3 (19%)	2 (10%)	ns
IQ (KBIT-2)			
Overall IQ	99.06 (16.57)	94.05 (19.34)	ns
Verbal IQ	98.56 (15.59)	91.89 (20.26)	ns
Baseline measures			
Questions asked	17.72 (21.03)	11.06 (7.47)	ns
Positive facial expressions	18.26 (18.63)	17.69 (14.83)	ns
Mutual engagement	27.76 (8.53)	29.32 (11.66)	ns

SD: standard deviation; M: mean; KBIT-2: Kaufman Brief Intelligence Test—Second Edition.

Table 3. Dependent measures for both groups at pre and post.

Variable	Group M (SD)				Group × Time	
	Treatment (n = 16)		Waitlist (n = 19)		p value	Partial eta squared
	Pre	Post	Pre	Post		
Questions asked	17.72 (21.03)	24.98 (21.86)	11.06 (7.47)	11.43 (8.92)	0.035*	0.128
Positive facial expressions	18.26 (18.63)	32.39 (27.90)	17.69 (14.83)	18.91 (13.72)	0.009**	0.190
Mutual engagement	27.76 (8.53)	32.34 (9.99)	29.34 (11.66)	32.06 (11.00)	0.576	0.010

M: mean; SD: standard deviation.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

study. Chi-square analyses for gender and ethnicity were not significant (all $p > 0.780$). Independent samples t tests for group differences on age, grade, verbal IQ, overall IQ, and behavioral measures for questions asked, positive facial expressions, and mutual engagement were also not significantly different at baseline (all $p > 0.244$).

Results revealed a significant Group × Time interaction effect for questions asked, $F(1, 33) = 4.86$, $p = .035$. Participants in the treatment group significantly increased the percentage of questions asked from pre- to post-intervention, whereas the percentage of questions asked by the waitlist group remained relatively unchanged. Effect size was medium-large (partial $\eta^2 = .128$). In addition, analyses revealed the presence of a significant Group × Time effect on the percentage of positive facial expressions, $F(1, 33) = 7.73$, $p = .009$. The treatment group demonstrated significant improvements in the percentage of directed positive facial expressions during social conversations

compared to the waitlist group. Effect size was large (partial $\eta^2 = .190$). Group × Time interaction did not reach statistical significance for mutual engagement, $F(1, 33) = 0.32$, $p = .576$, as both groups demonstrated minimal change. The results are presented in Table 3.

Discussion

The data from this RCT revealed significant Group × Time differences across two out of three examined social behaviors. A significant effect was found for the treatment group for the amount of questions asked from pre- to post-intervention, suggesting that the focus on the importance of social initiations during START sessions increased participant use of social inquiries during live interactions. Question asking initiations are largely considered a crucial social skill that conveys empathy and a desire for connection (Doggett et al., 2013; Palmen et al., 2008; Weiss and Harris, 2001).

There were also significant Group × Time differences in the use of positive facial expressions. The frequency of positive facial expressions directed to the social partner during conversations significantly improved for participants in the treatment group. An increase in this behavior can be construed as greater nonverbal engagement and attunement in interpersonal interactions, which may lead to more favorable social impressions (Lau, 1982). Therefore, the considerable increase in adolescents' facial positivity noted in this study has important implications. These behaviors are crucial elements in impression management, as they clearly communicate interest, engagement, and empathy to a social partner, which in turn are key elements in adolescent friendship formation and maintenance (Chow et al., 2013).

There were no Group × Time differences in mutual engagement, indicating that observed changes were minimal and did not rise to the level of statistical significance. The frequency of intervals in which conversational participants contributed equally remained relatively constant between groups and across time, suggesting that perhaps the START intervention did not yield improvements to mutual engagement as defined in our investigation. It is likely that this definition needs revising. One suggestion worth exploring would be to code this behavior using a time interval that is larger than 5 s. Given that one individual may talk for more than 5 s during a single turn, a larger time interval, such as 15 s, may be able to better quantify reciprocity and turn-taking in conversation. Another thought is that data on the number of speaking intervals alone may not accurately capture the essence of social reciprocity. For instance, two social partners could still be mutually engaged in a conversation even when one individual is speaking the majority of the time. Future studies could explore larger time intervals for coding mutual engagement and also investigate other definitions of the behavior that more adequately capture the nuances and complexities of social reciprocity.

The configuration of the social conversation probes not only allowed for an examination of social skill use with novel social partners but it also provided natural opportunities for generalization of these acquired social competencies. Participants were intentionally paired with unfamiliar, naïve peers at the end of the investigation to ensure that any observed improvements could not simply be attributed to increased familiarity with a prior social partner. As participants conversed with similarly aged individuals unaffiliated with the investigation, the data indicate evidence of generalization of these skills to other people. These findings are highly encouraging, as the generalization of skills beyond other group members and facilitators is a critical benchmark for the success of a socialization intervention (Cunningham, 2012; Dekker et al., 2014; Koegel et al., 1998; Schmidt and Stichter, 2012).

In our social conversation probes, there may be unique advantages to using peer social partners who are not trained or coached in any way. Perhaps most importantly, the resulting conversations are truly authentic interactions that closely mirror actual real-world encounters. Such interactions yield unique social data that would not be possible using trained conversational partners. By definition, trained confederates are taught to interact in ways that deviate from their default interpersonal approach. In addition to the obvious behavioral changes associated with adhering to specific guidelines and/or scripts, there may also be subtle changes in their behavior due to knowledge of participant social vulnerabilities.

For example, a trained confederate may be much more accommodating and tolerant of participant idiosyncrasies given their knowledge of the social quirks associated with ASD. They may intentionally or inadvertently repair or rescue a derailed conversation, which could influence participant social behavior and cast them a more favorable light. In contrast, an untrained peer is much more likely to provide an authentic reaction to the same awkward exchange and allow it to unfold naturally. This approach enables a more accurate read of the participant's true social aptitude. If the ultimate goal of a social skills outcome measure is to obtain the closest possible approximation to real-world social performance, then it makes sense to use live social encounters with untrained peers to obtain these data.

Limitations and future directions

Participant inclusionary criteria required past documentation of an ASD diagnosis and scores below the autism cutoff on the SRS-2, which has been found to strongly correlate with the Autism Diagnostic Interview-Revised (Constantino et al., 2003). However, a limitation of our investigation is that participant ASD diagnoses were not verified by our research team using comprehensive gold-standard diagnostic measures, such as the Autism Diagnostic Observation Schedule—Second Edition (ADOS-2; Lord et al., 2012) and the Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003).

Although systematic coding of conversational behaviors provides a more direct, objective outcome measure, there are a number of drawbacks to conducting this type of measure that may limit more widespread implementation. This method can be resource and time-intensive to obtain recording equipment and a team of coders (Furr and Funder, 2007). In addition, it is time-consuming to train coders, establish and maintain reliability, and code videos (Bakeman, 2000), which can restrict the use of these techniques in other research labs. In the future, the use of shorter conversation probes and a standardized coding paradigm could be implemented to make observational assessment into a more efficient, standardized tool for widespread use.

Another limitation when using behavioral observations is that participant social performance may have been influenced as a result of knowing that they were being observed and other possible demand characteristics (Kazdin, 1981).

While attempts were made to minimize the intrusiveness of the video (using small, unobtrusive recording devices in a casual social setting), there is a possibility that participants opted to interact differently due to the fact that they were being recorded. Future methods for unobtrusively gathering social performance data from participants in everyday social settings (perhaps in home, school, and community environments) would be essential to thoroughly evaluate generalized improvements. Fortunately, there is some existing evidence of convergent social competence improvements in parent and self-report data from the START program, which offers some support for the generalization of improvements to other settings (Vernon et al., 2018).

An additional limitation of the current investigation is the focus on three observational measures. Because of the time-intensive nature of coding the 140 5-min videos for each measure (approximately 35 h of total coding), representative measures were selected that were believed to encapsulate core social deficits associated with ASD. Specifically, deficits in social initiations, affective expression, and social reciprocity were studied by examining total questions asked, positive facial expressions, and mutual engagement, respectively. However, there are a wide range of clinically relevant social skills and associated behaviors that could have also been examined. In addition, while quantitative data were collected on the selected measures, we did not collect more detailed qualitative information. As such, we cannot make claims as to the contextual appropriateness of the observed participant social behavior. For example, problems with off-topic questions or inappropriate smiling at a conversation partner would not be captured in the data. Future studies may wish to explore additional behavioral dimensions that may provide a more comprehensive profile of social functioning.

Finally, although our results indicated that participants made a measurable improvement in the use of some social behaviors, there is a need to determine if these social skill improvements actually cause participants to be perceived as more socially desirable conversation partners. In the future, we plan to collect subjective evaluations of social competence from naïve observers unaffiliated with the current project. After watching the video-recorded conversations, these peers will be able to provide social competence ratings. These data would not only help determine if everyday observers are able to discern improvements in social aptitude after completing the START program but also yield insight into which behaviors are most associated with desirable social competence ratings.

Conclusion

This investigation constitutes one of the first RCTs to use coding of live peer conversations as social skills outcome data (joining Dolan et al., 2016). These findings provide evidence that the START program may have the ability to make a positive and meaningful impact on a variety of social behaviors. Furthermore, the observational measures used are potentially a more rigorous, objective, and unbiased assessment tool, which can augment other methods of measuring outcomes and contribute to a more complete picture of social functioning.

Overall, the results of this study further support the social benefit of the START socialization intervention for adolescents with ASD. The results of this study augment the findings of Vernon et al. (2018), which previously examined results from the parent and self-report measures from this sample of participants. The START program appears to impact the use of key verbal and nonverbal conversational strategies that may be important for both social skill development and positive social impressions. Considered collectively, converging sources of evidence provide strong support for the utility of the START program for improving the social functioning of adolescents with ASD.

References

American Psychiatric Association (APA) (2013) *Diagnostic and Statistic Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Press.

Bakeman R (2000) Behavioral observation and coding. In: Reis HT and Judd CM (eds) *Handbook of Research Methods in Social and Personality Psychology*. New York: Cambridge University Press, pp.138–159.

Bieberich AA and Morgan SB (1998) Brief report: affective expression in children with autism or down syndrome. *Journal of Autism and Developmental Disorders* 28(4): 333–338.

Bolte EE and Diehl JJ (2013) Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 43(11): 2491–2501.

Chow CM, Ruhl H and Buhrmester D (2013) The mediating role of interpersonal competence between adolescents' empathy and friendship quality: a dyadic approach. *Journal of Adolescence* 36(1): 191–200.

Constantino JN and Gruber CP (2005) *Social Responsiveness Scale*. Torrance, CA: Western Psychological Services.

Constantino JN, Davis SA, Todd RD, et al. (2003) Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders* 33(4): 427–433.

Cunningham AB (2012) Measuring change in social interaction skills of young children with autism. *Journal of Autism and Developmental Disorders* 42(4): 593–605.

Dekker V, Nauta MH, Mulder EJ, et al. (2014) A randomized controlled study of a social skills training for preadolescent children with autism spectrum disorders: generalization of skills by training parents and teachers? *BMC Psychiatry* 14(1): 189–203.

Doggett RA, Krasno AM, Koegel LK, et al. (2013) Acquisition of multiple questions in the context of social conversation in children with autism. *Journal of Autism and Developmental Disorders* 43(9): 2015–2025.

Dolan BK, Van Hecke AV, Carson AM, et al. (2016) Brief report: assessment of intervention effects on in vivo peer interactions in adolescents with autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders* 46(6): 2251–2259.

Furr R and Funder D (2007) Behavioral observation. In: Krueger RF and Fraley RC (eds) *Handbook of Research Methods in Personality Psychology*. New York: Guilford Press, pp.273–291.

Gates JA, Kang E and Lerner MD (2017) Efficacy of group social skills interventions for youth with autism spectrum disorder: a systematic review and meta-analysis. *Clinical Psychology Review* 52: 164–181.

Hartmann DP, Barrios BA and Wood DD (2004) Principles of behavioral observation. In: Thomas JC and Hersen M (eds) *Comprehensive Handbook of Psychological Assessment*, vol. 3. Hoboken, NJ: John Wiley & Sons, pp.108–127.

Hobson RP and Lee A (1998) Hello and goodbye: a study of social engagement in autism. *Journal of Autism and Developmental Disorders* 28(2): 117–127.

Johnson B and Turner LA (2003) Data collection strategies in mixed methods research. In: Tashakkori A, Teddlie C and Teddlie CB (eds) *Handbook of Mixed Methods in Social & Behavioral Research*. London: SAGE, pp.297–319.

Kasari C (2002) Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders* 32(5): 447–461.

Kasari C, Rotheram-Fuller E, Locke J, et al. (2012) Making the connection: randomized controlled trial of social skills at school for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry* 53(4): 431–439.

Kaufman AS and Kaufman NL (2004) *Kaufman Brief Intelligence Test*. 2nd ed. San Antonio, TX: Pearson.

Kazdin AE (1981) Behavioral observation. In: Bellack AS and Hersen M (eds) *Behavioral Assessment: A Practical Handbook*. London: Pearson, pp.101–124.

Koegel LK, Camarata SM, Valdez-Menchaca M, et al. (1998) Setting generalization of question-asking by children with autism. *American Journal on Mental Retardation* 102(4): 346–357.

Koegel LK, Koegel RL, Hurley C, et al. (1992) Improving social skills and disruptive behavior in children with autism through self-management. *Journal of Applied Behavior Analysis* 25(2): 341–353.

Lau S (1982) The effect of smiling on person perception. *The Journal of Social Psychology* 117(1): 63–67.

Lord C, Rutter M, DiLavore PC, et al. (2012) *Autism Diagnostic Observation Schedule*. 2nd ed. Torrance, CA: Western Psychological Services.

Lord C, Wagner A, Rogers S, et al. (2005) Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders* 35(6): 695–708.

McMahon CM and Solomon M (2015) Brief report: parent–adolescent informant discrepancies of social skill importance and social skill engagement for higher-functioning adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 45(10): 3396–3403.

McMahon CM, Lerner MD and Britton N (2013a) Group-based social skills interventions for adolescents with higher-functioning autism spectrum disorder: a review and looking to the future. *Adolescent Health, Medicine and Therapeutics* 4: 23–28.

McMahon CM, Vismara LA and Solomon M (2013b) Measuring changes in social behavior during a social skills intervention for higher-functioning children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 43(8): 1843–1856.

Matson JL and Rieske RD (2014) Are outcome measures for early intensive treatment of autism improving? *Research in Autism Spectrum Disorders* 8(3): 178–185.

Miller A, Vernon T, Wu V, et al. (2014) Social skill group interventions for adolescents with autism spectrum disorders: a systematic review. *Review Journal of Autism and Developmental Disorders* 1(4): 254–265.

Moskowitz DS (1986) Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *Journal of Personality* 54(1): 294–317.

Newman B and Ten Eyck P (2005) Self-management of initiations by students diagnosed with autism. *Analysis of Verbal Behavior* 21(1): 117–122.

Palmen A, Didden R and Arts M (2008) Improving question asking in high-functioning adolescents with autism spectrum disorders: effectiveness of small-group training. *Autism* 12(1): 83–98.

Ratto AB, Turner-Brown L, Rupp BM, et al. (2011) Development of the contextual assessment of social skills (CASS): a role play measure of social skill for individuals with high-functioning autism. *Journal of Autism and Developmental Disorders* 41(9): 1277–1286.

Rutter M, Le Couteur A and Lord C (2003) *Autism Diagnostic Interview, Revised*. Torrance, CA: Western Psychological Services.

Schmidt C and Stichter JP (2012) The use of peer-mediated interventions to promote the generalization of social competence for adolescents with high-functioning autism and Asperger's syndrome. *Exceptionality* 20(2): 94–113.

Schulz KF, Altman DG and Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine* 8(1): 18.

Stagg SD, Slavny R, Hand C, et al. (2014) Does facial expressivity count? How typically developing children respond initially to children with autism. *Autism* 18(6): 704–711.

Suen HK and Ary D (1989) *Analyzing Quantitative Behavioral Observation Data*. London: Psychology Press.

Vernon TW, Miller AR, Ko JA, et al. (2016) Social Tools And Rules for Teens (the START program): program description

and preliminary outcomes of an experiential socialization intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 46(5): 1806–1823.

Vernon TW, Miller AR, Ko JA, et al. (2018) A randomized controlled trial of the Social Tools And Rules for Teens (START) program: an immersive socialization intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 48(3): 892–904.

Vismara LA and Rogers SJ (2010) Behavioral treatments in autism spectrum disorder: what do we know? *Annual Review of Clinical Psychology* 6: 447–468.

Weiss MJ and Harris SL (2001) Teaching social skills to people with autism. *Behavior Modification* 25(5): 785–802.

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